Instructional Research Group

A Meta-analysis of Mathematics Instructional Interventions for Students with Learning Disabilities: Technical Report

2009

Russell Gersten

Instructional Research Group

David J. Chard

Southern Methodist University

Madhavi Jayanthi

Instructional Research Group

Scott K. Baker

Pacific Institutes for Research and University of Oregon

Paul Morphy

Vanderbilt University

Jonathan Flojo

University of California at Irvine

This research project was funded by the Center on Instruction under the Cooperative agreement S283B050034 with the U.S. Department of Education. The Center on Instruction is operated by RMC Research Corporation in partnership with the Florida Center for Reading Research at Florida State University; Instructional Research Group; the Texas Institute for Measurement, Evaluation, and Statistics at the University of Houston; and the Meadows Center for Preventing Educational Risk at the University of Texas at Austin.

Preferred citation:

Gersten, R., Chard, D., Jayanthi, M., Baker, S., Morphy, P., & Flojo, J. (2009). *A Meta-analysis of Mathematics Instructional Interventions for Students with Learning Disabilities: A Technical Report.* Los Alamitos, CA: Instructional Research Group.

To download a copy of this document, visit www.inresg.org.

Table of Contents

Abstract

The purpose of this study was to use meta-analysis to synthesize findings from randomized control trials and quasi-experimental research on instructional approaches that enhance the mathematics proficiency of students with learning disabilities. A search of the literature from January 1971 to August 2007 resulted in a total of 42 interventions (41 studies) that met the criteria for inclusion in the study. We examined the impact of four categories of instructional components: (a) approaches to instruction and/or curriculum design, (b) providing formative assessment data and feedback to teachers on students' mathematics performance, (c) providing formative data and feedback to students with LD on their mathematics performance, and (d) peer-assisted mathematics instruction. We first examined the effectiveness of each instructional component in isolation by determining unconditional stratified mean effects and heterogeneity. All instructional components except for student feedback with goal-setting and peer-assisted learning within a class resulted in significant mean effects ranging from 0.21 to 1.56. We then examined the effectiveness of these same components conditionally, using hierarchical multiple regressions. We created a model to understand instructional variables that explain significant amounts of unique variance in outcomes. Two instructional components were associated with practically and statistically important increases in effect size – teaching students to use heuristics and explicit instruction. Limitations of the study, suggestions for future research, and applications for improvement of current practice are discussed.

Introduction

Current prevalence estimates for students with learning disabilities and deficits in mathematics competencies typically range from 5% to 7% of the school age population (Fuchs et al., 2005; Geary, 2003; Gross-Tsur, Manor, & Shalev, 1996; Ostad, 1998). When juxtaposed with the well-documented inadequate mathematics performance of students with learning disabilities (Bryant, Bryant, & Hammill, 2000; Cawley, Parmar, Yan, & Miller, 1998; Geary, 2003), these estimates highlight the need for effective mathematics instruction based on empirically validated strategies and techniques.

Until recently, mathematics instruction was often treated as an afterthought in the field of instructional research on students with LD. A recent review of the ERIC literature base (Gersten, Clarke, & Mazzocco, 2007) found that the ratio of studies on reading disabilities to mathematics disabilities and difficulties was 5:1 for the decade 1996-2005. This was a dramatic improvement over the ratio of 16:1 in the prior decade.

During the past five years, two important bodies of research have emerged and helped crystallize mathematics instruction for students with learning disabilities. The first, which is descriptive, focuses on student characteristics that appear to underlie learning disabilities in mathematics. The second, which is experimental and the focus of this meta-analysis addresses instructional interventions for students with learning disabilities.

We chose to conduct a meta-analysis on interventions with students with learning disabilities and to sort studies by major types of instructional variables rather than conduct a historical, narrative review of the various intervention studies. Although three recent research syntheses (Kroesbergen & Van Luit, 2003; Swanson & Hoskyn, 1998; Xin & Jitendra, 1999) involving meta-analytic procedures target aspects of instruction for students experiencing mathematics difficulties, major questions remain unanswered.

Swanson and Hoskyn (1998) investigated the effects of a vast array of interventions on the performance of adolescents with LD in areas related to academics, social skills, or cognitive functioning. They conducted a meta-analysis of experimental intervention research on students with LD. Their results highlight the beneficial impact of cognitive strategies and direct instruction models in many academic domains, including mathematics.

Swanson and Hoskyn (1998) organized studies based on whether there was a measurable outcome in a target area and whether some type of treatment was used to influence performance. Swanson and Hoskyn were able to calculate the effectiveness of interventions on mathematics achievement for students with LD, but did not address whether the treatment was an explicit focus of the study. A study investigating a behavior modification intervention, for example, might examine impacts on both reading and mathematics performance. The link between the intervention and math achievement would be made even though the focus of the intervention was not to improve mathematics achievement per se. Thus, the Swanson and Hoskyn meta-analysis only indirectly investigated the effectiveness of mathematics interventions for students with LD.

The other two relevant syntheses conducted so far investigated math interventions directly (i.e., math intervention was the independent variable) but focused on a broader sample of subjects experiencing difficulties in mathematics. Xin and Jitendra (1999) conducted a meta-analysis on word problem solving for students with high incidence disabilities (i.e., students with learning disabilities, mild mental retardation, and emotional disturbance), as well as students without disabilities who were at-risk for mathematics difficulties. Xin and Jitendra examined the impacts associated with four instructional techniques - representation techniques (diagramming), computer-assisted instruction, strategy training and "other" (i.e., no instruction like attention only, use of calculators, or instruction not included on other categories like key word or problem sequence). They included both group design and single subject studies in their meta-analysis; the former were analyzed using standard mean change, while the later were analyzed using percentage of nonoverlapping data (PDN). For group design studies, they found computer-assisted instruction to be most effective, and representation techniques and strategy training superior to "other".

Kroesbergen and Van Luit (2003) conducted a meta-analysis of mathematics interventions for elementary students with special needs (students at-risk, students with learning disabilities, and low-achieving students). They examined interventions in the areas of preparatory mathematics, basic skills, and problem solving strategies. They found interventions in the area of basic skills to be most effective. In terms of method of instruction for each intervention, direct instruction and self-instruction were found to be more effective than mediated instruction. Like Xin and Jitendra (1999) Kroesbergen and Van Luit included both single subject and group design studies in their meta-analysis, however they did not analyze data from these studies separately. We have reservations about the findings as the data analytic procedures used led to inflated effect sizes in single subject studies (Busse, Kratochwill, & Elliot, 1995).

Neither of the two meta-analyses (i.e., Kroesbergen & Van Luit, 2003; Xin & Jitendra, 1999) focused specifically on students with learning disabilities. We believe there is relevant empirical support for a research synthesis that focuses on mathematical interventions conducted for students with learning disabilities. Our reasoning was most strongly supported by a study by Fuchs, Fuchs, Mathes, and Lipsey (2000), who conducted a meta-analysis in reading to explore whether students with LD could be reliably distinguished from students who were struggling in reading but were not identified as having a LD. Fuchs et al. found that low-achieving students with LD performed significantly lower than students without LD. The average effect size differentiating these groups was 0.61 standard deviation units (Cohen's d), indicating that the achievement gap between the two groups was substantial. Given this evidence of differentiated performance between students with LD and low-achieving students without LD, we felt it was necessary to synthesize mathematical interventions conducted with students with LD specifically.

Our intent was to analyze and synthesize research using parametric statistical procedures (i.e., calculating effect sizes using Hedges *g*). Calculating the effect sizes (Hedges *g*) for studies with single-subject designs would result in extremely inflated effect size scores (Busse, Kratochwill, & Elliot, 1995) and any mean effect size calculations would be impossible. Since there is no known statistical procedure for valid combination of single subject and group design studies, we limited our meta-analysis (as do most researchers) to those utilizing randomized control trials (RCTs) or high quality quasi-experimental designs.

Purpose of the Meta-Analysis

The purpose of this study was to synthesize RCTs and quasi-experimental research on instructional approaches that enhance the mathematics performance of school-age students with learning disabilities. We only included RCTs and quasi-experimental designs (QEDs) in which there was at least one treatment and one comparison group, evidence of pretest comparability for QEDs, and sufficient data with which to calculate effect sizes.

Method

Selection of Studies: Literature Review

In this study, we defined mathematical interventions as instructional practices and activities designed to enhance the mathematics achievement of students with LD. We reviewed all studies published from January 1971 to August 2007 that focused on mathematics interventions to improve the mathematics proficiency of school-age students with LD. Two searches for relevant studies were undertaken. The first search was from 1971 to 1999. The second search extended the time period to August 2007.

The 1971-1999 search began with a literature review using the ERIC and PSYCHINFO databases. The following combinations of descriptors were used in the search: mathematics achievement, mathematics education, mathematics research, elementary education, secondary education, learning disabilities, and learning problems. We also conducted a systematic search of Dissertation Abstracts

International and examined the bibliographies of research reviews on various aspects of instructional intervention research for students with learning disabilities (i.e., Kroesbergen & Van Luit, 2003; Maccini & Hughes, 1997; Mastropieri, Scruggs, & Shiah, 1991; Miller, Butler, & Lee, 1998; Swanson & Hoskyn, 1998; Swanson, Hoskyn, & Lee, 1999) for studies that may not have been retrieved from the computerized searches. Finally, we conducted a manual search of major journals in special, remedial, and elementary education (Journal of Special Education, Exceptional Children, Journal of Educational Psychology, Journal of Learning Disabilities, Learning Disability Quarterly, Remedial & Special Education, Learning Disabilities Research & Practice) to locate relevant studies.

These search procedures for the period between 1971 and 1999 resulted in the identification of 579 studies. Of this total, 194 studies were selected for further review based on analysis of the title, key words, and abstracts. Of these 194 studies located in the first search, 30 (15%) met our criteria for inclusion in the meta-analysis.

We conducted the 1999 to August 2007 search using a similar, but streamlined procedure. For this literature search, we used the terms mathematics and LD or arithmetic and LD. We also excluded dissertations from the search. The second search resulted in a pool of an additional 494 potential studies. We narrowed this set of studies to 38 by reviewing the title, keyword, and abstracts. Finally, 14 of the 38 studies (37%) met the criteria for inclusion in the meta-analysis. Thus, the two searches resulted in a total of 44 research studies.

During the first search, two of the authors determined if a study met the pre-established criteria for inclusion using a consensus model; any disagreements were reconciled. During the second round, this determination was made by the senior author. Another author independently examined 13 of the 38 studies (approximately one-third). Inter-rater reliability based on whether to include the study or not was 84.6%. The authors initially disagreed on the inclusion of two of the 13 studies; however, after discussion they reached a consensus. To ensure that studies from both searches were included using the same criteria, we randomly selected 20% of the studies (N=9) and conducted inter-rater reliability. A research assistant who was not involved with this project examined each study and determined if the study could be included or not based on the inclusion criteria. All nine studies met the criteria for inclusion; inter-rater reliability was 100% (which was calculated using the formula: number of agreements divided by number of agreements plus disagreements multiplied by 100).

Criteria for Inclusion

Three criteria were used to determine whether to include a study in this meta-analysis.

Focus of the Study

To be included in this meta-analysis, the study had to focus on an evaluation of the effectiveness of a well-defined method (or methods) for improving mathematics proficiency. This could be done in the following ways: (a) Specific curricula or teaching approaches were used to improve mathematics instruction (e.g., teacher use of "think-aloud" learning strategies; use of real world examples); (b) various classroom organizational or activity structures were used (e.g., peer-assisted learning); or (c) formative student assessment data were used to enhance instruction (e.g., curriculum-based measurement data; goal-setting with students using formative data). Studies that only examined the effect of test-taking strategies on math test scores, taught students computer-programming logic, or focused on computer-assisted instruction (i.e., technology) were not included. We felt that computer-assisted instruction would be more appropriate for a meta-analysis in the area of technology. Studies that assessed the achievement impact of changes in structural or organizational elements in schools, such as co-teaching or inclusion – but did not address a specific instructional approach – were excluded, even though they may have included a mathematics achievement measure (in this decision, we differed from Swanson & Hoskyn, 1998).

Design of the Study

We included studies that could lead to strong claims of causal inference, i.e., randomized controlled trials (RCTs) or quasi-experimental designs (QEDs). We noted if the study was a RCT[1] or a QED based on the presence of random assignment to the intervention conditions. Studies with single-case designs were not included, as they cannot be integrated into a meta-analysis. Quasi-experiments were included if students were pre-tested on a relevant mathematics measure and one of the following three conditions were met: (a) researchers in the original study adjusted post-test performance using appropriate analysis of covariance (ANCOVA) techniques; (b) authors provided pretest data so that effect sizes could be calculated using the Wortman and Bryant (1985) procedure or (c) if post-test scores could not be adjusted statistically for pretest differences in performance, there was documentation showing that no significant differences (<.25 SD units) existed between groups at pretest on relevant measures of mathematics achievement.

Participants in the Study

The participants had to be students with an identified LD. A study that also included students without LD was included if it met one of the following criteria: (a) separate outcome data were presented for the different participant groups so that effect sizes could be computed separately for students with LD; or (b) if separate outcome data were not presented for students with LD, then over 50% of the study participants needed to be students with LD. All studies provided operational definitions of LD or mathematical disabilities (MD). Definitions of LD often pertained either to state regulations regarding LD (e.g., Fuchs, Fuchs, Hamlett, Phillips, & Bentz, 1994) or district regulations (e.g., Marzola, 1987). For studies conducted outside the United States (Bar-Eli & Raviv, 1982; Manalo, Bunnell, & Stillman, 2000) we depended on authors' descriptions that focused on the I.Q. – achievement discrepancy, and for a more recent study we used contemporary language that reflects IDEA 2004 (Fuchs, Fuchs, & Prentice, 2004).

Coding of Studies

Phase I Coding: Quality of Research Design

We coded the studies that met the final eligibility criteria in three phases. In Phase I, two of the authors examined each study in terms of the strength and quality of the design. We examined the control groups to determine if the content covered in those groups was consistently relevant or minimally relevant to the purpose of the study. We also determined if the unit of analysis was class or student, and whether the unit of analysis and unit of assignment were the same. This information is important for statistical analyses as a mismatch can lead to spurious inferences since it fails to account for clustering at the classroom level (Bryk & Raudenbush, 1992; Donner & Klar, 2000; Gersten & Hitchcock, 2008). We also checked the studies to determine if only one teacher/school was assigned per condition, as this is a major confound (What Works Clearinghouse, 2006). This confound was not present in any of the studies.

Phase II Coding: Describing the Studies

In Phase II, all studies were coded on the following variables: (a) mathematical domain, (b) sample size, (c) grade level, (d) length of the intervention, and (e) dependent measures. We also determined who implemented the intervention (i.e., classroom teacher, other school personnel, or researchers), if fidelity of treatment was assessed, and whether scoring procedures for relevant mathematics performance scores included inter-rater agreement procedures.

Operational definition of mathematical domain. We used the work of the National Research Council (NRC) (2001), Fuchs and colleagues (e.g., Calhoon & Fuchs, 2003; Fuchs, Fuchs, Hamlett, & Appleton, 2002; Owen & Fuchs, 2002), and the National Mathematics Panel (2008) to identify and operationalize five mathematical domains of math achievement. These domains were (a) operations, (b) word problems, (c) fractions, (d) algebra, and (e) general math proficiency.

The domain of *operations* includes basic operations (i.e., addition, subtraction, multiplication, and/or division) of whole numbers, fractions, or decimals. *Word problems* includes all types of word problems – those requiring only a single step, those requiring multiple steps, those with irrelevant information, and those considered by the authors to be “real world” problems. The domain of *fractions* includes items that assess understanding of key concepts involving fractions such as equivalence, converting visual representations into fractions (and vice versa) or magnitude comparisons. *Algebra* was defined as simple algebraic procedures. *General math proficiency* covers a range of mathematical domains.

Dependent measures. We determined if a measure was researcher-developed or a commercially available norm-referenced test. We also categorized the measures in terms of the skills and knowledge involved in solving the problems. For example, did a measure test a range of skills like the Wide Range Achievement Test-Revised (WRAT-R) (Jastak & Wilkinson, 1984), focus on a narrow skill area such as the Math Operations Test-Revised (Fuchs, Fuchs, & Hamlett, 1989), or address general mathematics such as the Test of Mathematical Abilities (Brown, Cronin, & McEntire, 1994)? We examined the alignment between the focus of the intervention and the skills and knowledge being assessed by each measure. Finally, we gathered data on the technical adequacy of outcomes measures.

We needed uniform operational definitions for post-tests, transfer tests, and maintenance tests so that we could synthesize findings across disparate studies. Authors varied in terms of how they defined immediate post-test versus maintenance test, and what they considered a transfer test. Some considered a test given two days after a unit was completed a maintenance test. Some authors were extremely liberal in what they considered a transfer item (e.g., a word problem with a similar structure to what had been taught, but using slightly different words than those in the curriculum).

Consequently, we defined post-tests, maintenance tests, and transfer tests in the following manner:

A post-test had to measure skills covered by the instructional intervention. If the post-test measured new skills not covered during instruction, we made a note of it for subsequent use in interpreting the findings. In addition, most post-tests were given within three weeks of the end of the instructional intervention. If a post-test administration extended past the 3-week period we made a note of it.

A maintenance test is a parallel form of the post-test given 3 or more weeks after the end of the instructional intervention to assess maintenance of effects (i.e., retention of learning). If a maintenance test was given earlier than 3 weeks, we designated it as a post-test, and used it in our outcome calculations.

A transfer test measures the students' ability to solve problems that they were not exposed to during instruction. We used the definition of far transfer used in the work of Fuchs et al. (2002), and Van Luit and Naglieri (1999). Transfer tests include tasks that are different (sometimes substantially) from the tasks students were exposed to during the instructional intervention. For example, if the instruction covered single-digit addition problems, the transfer test could include two-digit addition or three-digit addition problems. Likewise, if the instruction was on mathematical operations (addition, subtraction, division, and multiplication), the transfer test could include problems requiring application of these skills (e.g., money, measurement, word problems, interpretation of charts or graphs etc). If the word problems included additional steps or asked the student to discern which information was irrelevant, these were considered transfer problems as well. Such far transfer measures were included in only nine studies and were used in calculating transfer effect sizes.

We included all near transfer tests in our outcome (post-test) calculations since near transfer measures require students to solve problems that closely resemble the ones used during instruction. Thus, the problems on the transfer measure differ from the post-test tasks in minor ways: for example, new numbers/quantities (change 23+45 to 13+34; six candies to 4 candies), different cover stories (buy pencils instead of erasers), and different keyword (how many boxes versus how many sacks).

Finally, measures that were parallel forms of post-tests (so clearly stated in the manuscripts) were not considered transfer tests, but were coded as either second post-tests or maintenance tests (depending on when they were administered).

Exclusion of studies during phase II coding. During Phase II coding, we excluded three studies from the meta-analysis. Friedman (1992) was excluded as the dependent measure Wide Range Achievement Test (WRAT) was poorly aligned with the intervention because the WRAT only assesses computation and the intervention focused on word problems. Greene (1999) was excluded from the meta-analysis because of a confounded design. Jenkins (1992) was excluded as the differential attrition in this study exceeded 30% and there was no attempt to conduct an intent-to-treat analysis. Shadish, Cook, and Campbell (2002) define an intent-to-treat analysis as: "In an intent-to-treat analysis, participants are analyzed as if they received the treatment to which they were assigned …This analysis preserves the benefits of random assignment for causal inference but yields an unbiased estimate only about the effects of being assigned to treatment, not of actually receiving treatment" (p. 320).

See Appendix A for a list of the 41 studies included in the meta-analysis and their characteristics. (Note: total number of experiments/quasi-experiments was 42, as one of the articles included two different experiments).

Phase III Coding: Determining the Nature of the Independent Variable(s)

The primary purpose of Phase III coding was to determine a set of research issues that could be explored in this set of studies. Two of the authors developed a coding scheme for the selected set of studies through an iterative process that spanned several months. During the first reading of the article, we coded according to a broad category (e.g., curriculum design, providing feedback to teachers and students on an ongoing basis). We then reviewed these initial codes, reviewed our notes, and reread relevant sections of each article to pinpoint the precise research questions addressed. This involved rereading of all the studies by at least two of the authors. The authors completed all the coding at this level, although we often involved research assistants in discussions.

In our final analysis, we settled on four major categories for the studies. These categories include (a) approaches to instruction and/or curriculum design, (b) providing ongoing formative assessment data and feedback to teachers on students' mathematics performance, (c) providing data and feedback to students with LD on their mathematics performance, and (d) peer-assisted mathematics instruction. These four broad categories were further dissected in terms of instructional components. The process of identifying these specifics was iterative and involved two authors and spanned several months.

Note that several studies included three or more intervention conditions, and thus addressed several research questions. These studies were therefore coded into more than one category whenever applicable. We used orthogonal contrasts to capture the unique research questions posed. Some research studies had complex instructional interventions that were based on fusion of instructional variables (e.g., use of visuals and explicit instruction). These studies were also coded into more than one category whenever applicable. However, two categories with the same complex intervention were never compared with each other.

We calculated inter-rater reliability on our coding of studies. We randomly picked 20% of the studies (N=9) and had a research assistant (a doctoral student), who was not involved in the meta-analysis, code these nine studies according to the definitions we had established for the instructional components (described in the next section). Inter-rater agreement was calculated by using the formula agreements divided by the number of agreements plus disagreements, multiplied by 100. The inter-rater agreement for 20% of the total studies included in the meta-analysis was 88%.

In the next section, we describe and present the operational definitions of the four major categories. They are as follows:

Approaches to instruction and/or curriculum design. Under this category we list six instructional components.

1*. Explicit instruction*. A good deal of the special education literature in mathematics has called for instruction to be explicit and systematic (e.g., Fuchs & Fuchs, 2003; Gersten, Baker, Pugach, Scanlon, & Chard, 2001; Swanson & Hoskyn, 1998). However, the term is used to describe a wide array of instructional approaches. We found a reasonable amount of variance in the way explicit instruction was defined in the 42 interventions reviewed. In order to operationalize the construct, we only coded examples of systematic, explicit instruction if they possessed the following three specific components: (a) the teacher demonstrated a step-by-step plan (strategy) for solving the problem, (b) this step-by-step plan needed to be specific for a set of problems (as opposed to a general problem solving heuristic strategy) and (c) students were asked to use the same procedure/steps demonstrated by the teacher to solve the problem. Thus in the studies we coded as including explicit instruction, students were not only taught explicitly a strategy that provided a solution to a given problem type but were also required to solve the problem using the same strategy.

2. *Use of heuristics*. To be included under this instructional component, the intervention had to address the use of one or more heuristics for solving a given problem. We have defined a heuristic as a method or strategy that exemplifies a generic approach for solving a problem. For example, a heuristic strategy can include steps such as “Read

the problem. Highlight the key words. Solve the problems. Check your work". Thus instruction in heuristics, unlike explicit instruction (as defined in this manuscript), is not problem specific. Heuristics can be used in organizing information and solving a range of math problems. Instruction in multiple heuristics exposes students to multiple ways of solving a problem and usually includes student discourse and reflection on evaluating the alternate solutions and finally selecting a solution for solving the problem.

3. *Student verbalizations of their mathematical reasoning*. To be coded here, the instructional intervention had to include some aspect of student verbalizations (e.g., verbalizing solution steps, self-instruction, etc). Student verbalization or encouragement of students' thinking-aloud about their approach to solving a problem is often a critical component in scaffolded instruction (e.g., Palincsar, 1986). Approaches that encourage and prompt this type of verbalization have been found to be effective for students with LD in a wide array of curricula areas, including content area subjects such as history and science, as well as foundational areas in reading and math (Baker, Gersten, & Scanlon, 2002).

Most discussions of mathematics teaching note that a key component of effectiveness is "manag(ing) the discourse around the mathematical tasks in which teachers and students engage. … [Teachers] must make judgments about when to tell, when to question, and when to correct. They must decide when to guide with prompting and when to let students grapple with a mathematical issue" (NRC, 2001; p. 345). The process of verbalizing how to solve problems should encourage students to select an appropriate representation and, in discussion with peers and/or their teacher, evaluate its relevance. It can also lead to discussions of which strategies apply to particular situations (Van Luit & Naglieri, 1999).

4. *Using visual representations while solving problems*. Visual representations of mathematical relationships are consistently recommended in the literature on mathematics instruction (e.g., Griffin, Case, & Siegler, 1994; NRC, 2001; Witzel, Mercer, & Miller, 2003). The NRC Report notes that "mathematical ideas are essentially metaphorical (p. 95) ... Mathematics requires representations ... Representations serve as tools for mathematical communication, thought, and calculation, allowing personal mathematical ideas to be externalized, shared and preserved. They help clarify ideas in ways that support reasoning and building understanding" (p. 94).

In order for a study to be coded as having a visual representation, either (a) the students had to use the visual representation while solving the problem, or (b) the teacher had to use the visual representation during the initial teaching and/or a demonstration of how to solve the target problem. If the study focused on student use of the visual, we required that student use be mandatory and not an optional step for students working to solve the problems.

5. *Sequence and/or range of examples*. The literature on effective mathematics instruction stresses the importance of example selection in teaching concepts to students (e.g., Ma, 1999; Silbert, Carnine, & Stein, 1989; Witzel, Mercer, & Miller, 2003). To be coded as having this instructional component, studies needed to assess the effectiveness of either (a) a specified sequence/pattern of examples (concrete to abstract, easy to hard, etc), or (b) represent systematic variation in the range of

examples (e.g., teaching only proper fractions versus initially teaching proper and improper fractions).

6. *Other instructional and curricular variables*. A study was coded as such, if it included instructional and curricular components other than the five previously listed.

***Providing ongoing formative assessment data and feedback to teachers on students' mathematics performance*.** Ongoing assessment and evaluations of students' progress in mathematics can help teachers measure the pulse and rhythm of their students' growth in mathematics, and also help them in fine-tuning their instruction to meet the needs of their students. We were interested in determining the effects of teacher monitoring of student performance on students' growth in mathematics, that is, an indirect impact of the use of assessments. To be included in this category, the teachers had to be provided with information on student progress. The information that was provided to the teachers could be (a) just feedback on student progress or (b) feedback plus options for addressing instructional needs (e.g., skill analysis, instructional recommendations, etc).

Providing formative assessment data and feedback to students with LD on their mathematics performance. Providing students with information regarding their performance or effort is considered by many to be a key aspect of effective instruction. Information about performance or effort may serve to positively reinforce student effort, it may serve as a way to keep students accountable for staying engaged in working as expected on the business of learning mathematics, and it may provide useful information for students in understanding where they have been successful and unsuccessful in their learning. To be included in this category, students had to receive some sort of feedback regarding their performance or effort. The students could have received (a) just feedback or (b) feedback that was tied to a specific performance goal. The feedback could also be from a variety of sources including teachers (e.g., Schunk & Cox, 1986), other peers (e.g., Slavin, Madden, & Leavey, 1984a), and computer software programs (e.g., Bahr & Reith, 1991).

Peer-assisted math instruction. Students with LD are often provided with some type of assistance or one-on-one tutoring in areas for which they need help. Sometimes students' peers provide this assistance or one-on-one tutoring. There are two types of peer tutoring. The more traditional is cross-age, wherein a student in a higher grade functions primarily as the tutor for a student in a lower grade (Robinson, Schofield, & Steers-Wentzell, 2005). In the newer within-classroom approach, two students in the same grade essentially tutor or assist each other. In many cases, a higher performing student is strategically placed with a lower performing student but typically both students work in the role of the tutor (who provides the tutoring) and the tutee (who receives the tutoring) (Fuchs, Fuchs, Yazdian, & Powell, 2002). For example, in working on a set of problems, the higher performing child will work on the problems first and the lower performing child will provide feedback. Then roles will be reversed and the lower performing child will work on problems for which he/she just had a model for how to solve them. Or in providing explanations for math solution, the higher performing child will provide the explanation first and the lower performing child will have had a model for a strong explanation (Fuchs, Fuchs, Phillips, Hamlett, & Karns, 1995).

Generally, students use their time in peer-assisted instruction practicing math problems for which they have received previous instruction from their teacher. To be included in this category, the studies had to include a peer-assistance instructional component as their independent variable.

Data Analysis

Effect Size Computation

Effect sizes for each contrast were calculated as Hedges *g* in the following manner: first, the difference between the experimental and comparison condition means was divided by the pooled standard deviation (Cooper & Hedges, 1994).

Then, for studies that reported both pretest and post-test scores, we calculated post-test effect sizes adjusting for pretest performance (i.e., $d_{adjusted} = d_{post\text{-}test} - d_{pretest}$) (Wortman & Bryant, 1985). This technique is especially useful for quasi-experimental studies or any study reporting initial non-equivalence of groups on a pretest measure. Our prior work has indicated that this adjustment provides more accurate gauges of effect size than simple unadjusted post-test effects (Baker, Gersten, & Lee, 2002; Gersten & Baker, 2001). Finally, the estimate was corrected for small sample bias using the Hedges correction (Hedges, 1981; What Works Clearinghouse, 2007). The effect sizes for each study are presented in Appendix B.

In this meta-analysis we encountered several unusual issues while computing effect sizes. They are as follows:

Effect size computation for studies with three or four experimental conditions. Many of the studies in our sample reported outcomes from two or three experimental conditions, each involving different combinations of instructional components. We could not compare each condition to the control group because the set of contrasts would not be orthogonal. We therefore developed either two or three orthogonal contrasts based on the research questions posed by the study's authors. We thus were able to compute two or three *g*s that were orthogonal and also addressed a specific research question (Hedges, personal communication, 2003).

Effect size computation for studies with classroom as the unit of analysis. Four research studies assigned classes to treatment conditions and assessed all of the students with LD in the class on pretest and outcome measures, but then entered the mean score of 1 to 4 selected LD students into the analysis of variance. While appropriately analyzing treatment effects at the level of assignment for the F-ratios and p values present in the study, the variance reported in the studies is problematic for meta-analysis. That is because effect sizes at the classroom level will tend to be somewhat inflated. Had the authors reported the ratio of between-class to within-class variance (ICC) we could have adjusted the Level-2 variance reported to the total variance (Level-2 + Level-1) required. Without the *ICC* report, an alternative for estimation was found in unmoderated *ICC* values reported by Hedges and Hedberg (2007, p. 72). These *ICC*s were further adjusted based on the differential ratios of Level-2 to Level-1 units in data sets from which they were drawn to sample sizes in studies analyzed here. Adjustment of *g* from these studies was then calculated:

$$g_{ICC} = g\sqrt{ICC_{adj}} \qquad \text{Where:} \quad ICC_{adj} = ICC\left((n_{level2} / n_{level1})_{study} / (n_{level2} / n_{level1})_{dataset}\right)$$

Aggregation and comparison across factors. Typical studies reported effects for multiple factors other than treatment group (e.g., gender, grade-level, measurement-type, or measurement-time-point). In addition, treatments themselves range in complexity from single component (e.g., same-grade peer tutoring) to multiple component interventions (e.g., peer tutoring + student feedback + goal setting). Considered separately these factors divide into *instructional components (*e.g., use of peer-assisted learning, example sequencing, use of think aloud procedures) *participant factors* (e.g., gender or grade-level), and *end-point factors* (e.g., measures or time-points), each of which was aggregated differently depending on their importance to the study and their estimability (Seely & Birkes, 1980).

For the present analysis, stratified analyses of treatment components allowed consideration of complex intervention effects in multiple treatment categories. Participant and endpoint factors, however, were aggregated to one effect size estimate per study. For participant factors (e.g., gender, grade-level) where each participant may be considered an independent unit for analysis, summary statistics (i.e., *mean, sd, and N*) were aggregated to single values using a procedure attributed to Nouri and Greenberg (Cortina & Nouri, 2000). For studies that reported multiple endpoints (i.e., multiple post-test measures or parallel versions of a test administered within 3-weeks after intervention concluded) different procedures were employed. Both of these endpoint off-factors may be expected to have correlated error that required different approaches.

In cases of parallel forms of a post-test being administered at multiple time points within 3-weeks of the end of the intervention, we treated these as a larger measure at a single time-point (i.e., Total score = $N_{\text{items}} \times k_{\text{time-points}}$). To aggregate multiple time-points a modification of the Nouri and Greenberg formula was used (Cortina & Nouri, 2000). For studies reporting outcomes with multiple measures a different approach was used: An effect size for each measure was first computed and effects so-computed were then combined into a simple average effect (i.e., $g_{average} = g_1 + g_2 \ldots + g_k / k$).

Although simple averaging implies the risk of overestimating the aggregated effect by underestimating the variance among measures, and remedies for this problem do exist (e.g., Gleser & Olkin, 1994; Rosenthal & Rubin, 1986), these remedies require correlational information which may neither be reported nor be directly estimable for meta-analysis (excepting cases where raw data are available). Also, while statistically meaningful, the difference of averaging effects and computing a latent effect from multiple measures may be small. For this study we judged such averaging to permit a reasonable approximation of the true score effect, capitalizing on the unit-free nature of the standardized mean difference statistic (i.e., *g*).

In some studies a combination of factors were presented for aggregation (e.g., multiple, groups, time-points, and measures), which required systematic application of the aggregation strategies described. Once computed, a sensitivity analysis of aggregated effects was conducted by regressing effect size onto the number of groups, time-points, and measures aggregated in a fixed-weighted analysis. This analysis revealed no systematic biasing (r_{max} = .10). Thus, having applied these selection and estimation procedures systematically, we calculated a pool of independent effect sizes (N = 51) for meta-analysis.

Q Statistic. For each instructional component (e.g., explicit instruction, feedback to students) we determined if the *g*s were consistent across the studies (i.e., shared a common effect size) by calculating a homogeneity statistic Q (Hedges & Olkin, 1985). The Q-statistic is distributed as chi-square with $k - 1$ degrees of freedom, where k is the number of effect sizes (Lipsey & Wilson, 2001) and is: $Q = \sum w_i \times \left(ES_i - \mu_{\cdot\cdot ES}\right)^2$. A significant chi-square indicates excessive variation in *g*, suggesting a set of effects to come from more than one population, and justifies further analysis in order to identify the study, population, and treatment characteristics which moderate this variation. As the Q in the current study was significant (i.e., $Q > df$; $p < .05$), a mixed-weight regression analyses was conducted to estimate the moderating influence of participant, intervention, and method characteristics on mathematics outcomes. (Raudenbush & Bryk, 2002).

Regression Analysis. While 51 effects were available for stratified analysis, we rarely saw a treatment in a pure form, i.e. a treatment with only one instructional component. More commonly each treatment differed from other treatment conditions by either one or two instructional components. Similarly, treatments varied in other ways arguably influential on outcomes (e.g., amount of instruction in mathematics content relevant to the outcome measure in control condition, use of commercially available norm-referenced or researcher-developed measures, and grade level of students). To evaluate mathematics intervention effects with consideration of these complexities required a hierarchical regression analysis of effects.

Since the goal of regression analysis is to consider both the correlations of treatments with effect size while controlling for the intercorrelations of these components with each other and with systematic differences among studies, stratified analyses were discarded in favor of simultaneous analysis of independent effects from each study. To achieve independence 41 effects were selected from the larger pool of 51 effects in the stratified analysis.

Since the effect size is by definition an index of treatment effect, post-hoc correlational investigations are limited to considering potential moderators of treatments. The development of a moderator model was undertaken to incrementally identify method differences (e.g., use of a meaningful control, inclusion of norm-referenced tests), participant characteristics (e.g., grade-level) and finally specific instructional component differences in explaining the variance among studies within mixed-weight simultaneous analyses using hierarchical linear models (HLM) (e.g., Raudenbush & Bryk, 2002).

Results

A total of 42 intervention studies were examined in this meta-analysis. Of these 42 studies, not all reported participant information such as SES, race/ethnicity, and gender. The SES of the participants was reported in 12 studies. On average 59.3% of the participants were low SES/free or reduced lunch (range 7% to 100%). Ethnic background was provided in 20 studies with non-minority (Caucasian) participants averaging 50.4% and ranging from less than 10% to 87.5%. Thirty-two studies report the ratio of male to female participants, with male participants averaging 59.8% (range of 5% to 100%).

Many of the studies in our meta-analysis received multiple codes because they contain two or more instructional components. For this reason, we first present data on the effectiveness of each instructional component when examined in isolation and then on the relative strengths and weaknesses of each instructional component when compared with each other (i.e., findings from the hierarchical multiple regressions).

Effectiveness of Instructional Components in Isolation

In Table 1 we present the mean effect sizes *(Hedges g)* using a random effects model and a measure of heterogeneity of outcomes *(Q)* for each of the instructional components. Statistical significance levels are also presented. All instructional components save *Student Feedback with Goal Setting* and *Peer-Assisted learning within a class* resulted in significant mean effects. One-third of the instructional components produced *Q* ratios that were statistically significant, indicating that the impact of that component did not lead to a coherent pattern of findings. We attempt to explain to the reader likely sources for some of this heterogeneity. In some cases, extremely large effects seemed to be caused, in part, by a control condition with a minimal amount of relevant instruction. We took this issue into account in the multiple regression analysis. In other cases, we simply speculate as to sources of variance. We also try to provide the reader with a sense of some of the interventions that are included in this meta-analysis.

Approaches to Instruction and/or Curriculum Design

Explicit instruction. In 11 studies, explicit instruction was used to teach a variety of strategies and a vast array of topics. For example, in Jitendra, Griffin, McGoey, and Gardill (1998) and Xin, Jitendra, and Deatline-Buchman (2005), students were taught explicitly how to use specific visual representations, and in Marzola (1987), Ross and Braden (1991), and Tournaki (1993, 2003) students were taught a verbalization strategy. The studies also varied in their instructional focus. In half the studies, the focus was quite narrow—for example, teaching students to find half of a given quantity (Owen & Fuchs, 2002) or solving one-step addition and subtraction word problems (Lee, 1992). In the remaining half, the focus was much broader. Nonetheless, the common thread among the studies was the use of systematic, explicit strategy instruction.

The mean effect size for the explicit instruction category was 1.22 ($p < .001$; range = 0.08 to 2.15) and significant. Substantial variation in the scope and the mathematical sophistication of the strategies taught explicitly might have accounted for the variation in effect sizes. As the effect size range suggests, the Q statistic (41.68) for this category was significant ($p < .001$), indicating that the outcomes were heterogeneous. Fuchs, Fuchs, Hamlett, and Appleton (2002) taught students to solve different types of word problems (e.g., determining half of a given quantity; determining money needed to buy a list of items). Steps of the problem specific solution were prominently displayed. Students were shown the application of the solution steps using fully and partially worked examples. Students were required to apply the steps of the solution as they worked through the problems. The effect size for this study was 1.78. In the Ross and Braden (1991) intervention (g = 0.08), students work through reasonable steps to solve the problem, but are not explicitly shown how to do the calculations.

Xin et al. (2005) (g = 2.15) incorporated explicit instruction in their instructional intervention, but in this case the strategy is derived from research on how experts solve mathematical problems (e.g., Fuson & Willis, 1989). In Xin et al. students were taught that there are several distinct problems involving multiplication and division. When given a problem, students first identify what type of problem it is (i.e., “proportion” or “multiplicative compare”) and then use a diagram linked to that specific problem type in order to create a visual representation of the critical information in that problem and the mathematical procedure(s) necessary to find the unknown. Students next translate the diagram into a math sentence and solve it. Unlike the Ross and Braden study, the Xin et al. intervention incorporates other instructional components such as sequencing instructional examples to obtain proficiency in each type of problem before systematically introducing contrasting examples. Although the control condition in this study did include use of visual representations, Xin et al. provide a much higher degree of structure and specificity associated with the visual representations in the experimental condition. It may be that a combination of these factors associated with effective instruction resulted in the observed impact of 2.15.

Use of heuristics. The mean effect size for this category (4 studies) was 1.56 ($p < .001$; range = 0.54 to 2.45) and significant. The Q for this category (9.10) was significant (p = .03), indicating that the outcomes were heterogeneous. Woodward, Monroe, and Baxter (2001) exposed students in their study to multiple ways of solving a problem. In the Woodward et al. intervention, "as different students suggested a strategy for solving the problem, the tutor probed the other students to see if they agreed and encouraged different individuals to work the next step in the problem… Explicit suggestions were made only after a substantial period of inactivity and after the tutor had determined that the students could not make any further progress without assistance" (p. 37). This intervention resulted in an effect size of 2.00.

Similarly, Van Luit and Naglieri (1991) also emphasized multiple heuristics (g = 2.45). They trained teachers to first model several different approaches for solving a computational problem. For most of the lesson, however, the teacher's task was to lead the discussion in the direction of using strategies and to facilitate the discussion of the solutions provided by the students. Each student was free to select a strategy for use, but the teacher assisted the children in discussion and reflection about the choices made.

Student verbalizations of their mathematical reasoning. Eight studies examined the impact of student verbalizations to improve mathematics performance. The mean effect size for this category was 1.04 ($p < .001$). The Q for this category (53.39) was significant ($p < .001$), indicating heterogeneity of outcomes across the studies. The studies reveal differences in the amount of student verbalization encouraged, the specificity of the verbalizations, and the type of verbalizations. Some studies gave students very specific questions to ask themselves. Others were based on cognitive behavior modification and provided students with very general guidance. For example, Hutchinson (1993) taught students to ask themselves: “Have I written an equation?” “Have I expanded the terms?” (p.39). Schunk and Cox (1986) provided even broader guidance to students - instructing them to verbalize what they were thinking as they solved the problems. The Schunk and Cox study resulted in an effect size of 0.07. The Marzola (1987) intervention had an effect size of 2.01, which is likely to be due to an artifact of the study; the control group received no instruction at all, just feedback on the accuracy of their independent work.

Using visual representations while solving problems. Twelve studies were included in this set. In seven of the studies, teacher use of the visual representation was followed by *mandatory student use of the same visual* while solving problems. These were sub-classified as *Visuals for Teacher and Student.* In the remaining five studies, only the teacher used the visual representations. We hypothesized that the first sub-category would produce different effects than the second. However, the mean effect sizes of 0.46 ($p < .001$) and 0.41 ($p = .02$) were similar, and consequently we discuss them as one set. The Q (14.13) for the 12 studies in this category was not statistically significant ($p = .23$), indicating relatively homogeneous outcomes.

The studies in the visual category used diverse, complex intervention approaches. In Owen and Fuchs (2002), students solved the problems in a graphic manner (e.g., draw a rectangle and divide it into half to make two boxes; distribute circles evenly into the two boxes; determine the number of circles in one box to determine the answer) without having to translate their problem into mathematical notation. This intervention resulted in as effect size of 1.39. The impact may be due to several factors. One may have been the specificity of the visual. The second may have been that the mathematical problems addressed in the study had a narrow focus: calculating half for a given numerical quantity. Task demands were also the lowest among the set of 12 studies. In contrast, in D. Baker (1992) (*g* = 0.31), students were exposed to the concept of visually representing the information presented in the problem using a variety of examples, but were not told to use a specific visual. In Kelly, Gersten, and Carnine (1990) visual representations were used only by the teachers as they initially explained the mathematical concepts and problems. The intervention resulted in an effect size of 0.88, which we think is attributable not just to the use of visuals, but also to the overall instructional package that was designed and controlled for using effective instruction principles.

Sequence and/or range of examples. Nine studies were included in this category. The mean effect size was 0.82 ($p < .001$; range = 0.12 to 2.15). The *Q* statistic (19.78) was also significant ($p = .01$) indicating heterogeneity of outcomes for this set of studies. The researchers utilized different frameworks for sequencing and selecting examples, thus heterogeneity in impacts is not surprising.

One approach was to build sequences to highlight distinctive features of a given problem type. This approach appears to be effective (*g* = 2.15) as illustrated by the research of Xin and colleagues (Xin et al., 2005). Another effective principle for sequencing exemplars was utilized by Wilson and Sindelar (1991) (*g* = 1.55) and Woodward (2006) (*g* = 0.54). Here, instructional examples progressed from easy to more complex and difficult examples in a systematic fashion. Butler, Miller, Crehan, Babbitt, and Pierce (2003) (g = 0.29) and Witzel et al. (2003) (g = 0.50) used a CRA (concrete-representational-abstract) instructional sequence to ensure that students actually understood the visual representations before using them as a means to illustrate mathematical concepts. The authors hypothesized that students with LD, even in the secondary grades, still need brief periods of time devoted to using concrete objects to help them understand the meaning of visual representations of fractions, proportions, and similar abstract concepts. Concepts and procedures involving fractions and algebraic equations were taught first with concrete examples, then with pictorial representations, and finally with abstract mathematical notation.

Furthermore, three studies (Fuchs et al., 2004; Kelly et al., 1990; Owen & Fuchs, 2002) addressed the issue of range of examples in their instructional sequencing. Fuchs et al. (g = 1.14) exposed students to a range of problems that encompassed four superficial problem features (i.e., novel context, unfamiliar keyword, different question, larger problem solving context) but used the same mathematical structure. As highlighted by the various studies in this category, the potential role of careful selection and sequencing of instructional examples to illustrate contrasts, build in-depth knowledge of mathematical processes, and highlight common features to seemingly disparate word problems seems to be quite important in helping students with LD learn mathematics.

Other curriculum and instruction variables. One study, by Bottge, Heinrichs, Mehta, and Hung (2002), did not fit into any of our coding categories. This study explored the impact of *enhanced anchored instruction (EAI).* The intent of EAI is to provide students with opportunities (for applications of mathematical principles and processes) that would focus on engaging real world problems in a systematic, abstract fashion. The underlying concept was that if students were asked to solve engaging real world problems (e.g., build a skateboard ramp) involving use of previously taught concepts like fractions and other computation skills, then the resulting enhanced motivation would dramatically increase their engagement in the learning task. Another unique feature of this intervention was that students were taught foundational knowledge using paper and pencil tasks and traditional texts, but application problems were typically presented via video or CD, rather than by traditional print. The effect size was 0.80, indicating some promise for this technique.

Providing Ongoing Data and Feedback to Teachers on Students' Mathematics Performance: The Role of Formative Assessment

Seven studies met the criteria for inclusion in this category. All but two studies included three experimental conditions and a control condition, enabling us to identify 3 orthogonal contrasts per study. By using orthogonal contrasts, the assumption of statistical independence was maintained, which is critical for meta-analysis (Hedges, 2003, personal communication). Overall, the seven studies resulted in a total of 10 contrasts. Consequently, the orthogonal contrasts were classified into two sub-categories: a) teachers were provided with feedback on student progress (formative assessment data), and b) teachers were provided with feedback plus options for addressing instructional needs (e.g., skills analysis, instructional recommendations, etc).

Providing teachers with feedback on student progress. For seven contrasts, teachers were provided with ongoing student performance data. Five of these studies involved special educators and two involved general education teachers, but only data from the special education students in the classroom were included in the statistical analysis. The mean effect size for this set of studies was 0.21 (*p* = .04; range = 0.14 to 0.40). The *Q* statistic was not significant (0.32, *p* = 1.0), indicating effects were relatively consistent. Feedback was provided to the teachers periodically (in most cases bimonthly) over periods of time ranging from 15 weeks to 2 school years.

***Providing teachers with feedback plus* options for addressing instructional needs *(e.g., skill analysis, instructional recommendations)*.** Three studies included an orthogonal contrast that allowed us to test the "value added" by an option for addressing instructional needs. The mean effect size for this set of studies was 0.34 (*p* = .10; range = -0.06 to 0.48) and approached significance. In other words, the guidance options made the formative assessments significantly more effective.

The options for addressing instructional needs provided in these three studies helped teachers in planning and fine tuning their instruction. For e.g., Allinder, Bolling, Oats, and Gagnon (2000) provided teachers with a set of prompts (written questions) to help them use the formative assessment data for adaptation of instruction. These prompts included the following: "On what skill(s) has the student improved compared to the previous 2-week period?" "How will I attempt to improve student performance on the targeted skill(s)?" Teachers detailed their responses on a one-page form. Then they repeated the process two weeks later using both the new assessment data and the previous form to assist in decisions. In another study, Fuchs et al. (1994) provided teachers with a set of specific recommendations to accompany the student performance data. Recommendations included: a) topics requiring additional instructional time for the entire class, b) students requiring additional help via some sort of small group instruction or tutoring, and c) topics to include in small group instruction, peer tutoring, and computer assisted practice for each student experiencing difficulty.

In summary, when we analyze all the 10 contrasts in the teacher feedback category, we note that the set of studies is coherent and has a mean effect size of 0.23 ($p = .01$). Thus, providing feedback to teachers with or without additional guidance appears to be beneficial to students with LD.

Providing Formative Assessment Data and Feedback to Students with LD on their Mathematics Performance

Studies were categorized into two subcategories: (a) studies that provided data and feedback to students on their performance or effort (seven studies); and (b) studies that provided feedback that was also linked to some type of goal (five studies). Providing students with information on their progress in graphic form was statistically significant (0.23, $p = .01$). However, the mean effect size for involving students in the goal-setting process (wherein they take part in setting goals or are made aware of pre-set goals) and using formative assessment data to assess progress towards that goal was not statistically significant, 0.17 ($p = .29$). Thus, the key finding from this set of studies is that providing feedback to students enhances achievement. However, the evidence does not suggest that involving students in setting instructional goals is beneficial.

Providing feedback only to students. In all studies except Schunk and Cox (1986), students were given feedback on their mathematical performance. This performance feedback ranged from a simple communication of the number of problems answered correctly to more extensive and in-depth communication systems that presented graphs of scores, skill profiles, and mastery status information (skills learned and not learned). In Schunk and Cox, feedback was given on effort expended (e.g., "You've been working hard."). Interestingly, this study had an effect size of 0.60. Effect sizes for other studies in this category ranged from -0.17 to 0.24. We also noted there were variations in the sources of feedback — adults, peers, and software programs.

Providing feedback to students with goals. In three studies (Bahr & Reith, 1991; Fuchs et al., 1997; Fuchs, Fuchs, Hamlett, & Whinnery, 1991), goal setting was examined in terms of its value-added function (i.e., feedback with goal setting versus feedback only). Effect sizes in the range of -0.34 to 0.07 were associated with these three studies, which make sense given that the control condition was also involved in providing feedback to students. However, when the research question did not examine the value-added aspect of goal setting, as in Fuchs et al. (2004) and Reisz (1984), impacts appeared stronger. In Fuchs et al. ($g = 1.14$), a combination of instructional variables including feedback with goal setting were contrasted with regular classroom instruction.

It appears that goal setting does not add additional value over providing feedback to students with LD. Given the problems many of these students have with self-regulation (Graham & Harris, 2003; Wong, Harris, Graham, & Butler, 2003; Zimmerman, 2001), providing feedback on progress by a teacher or peer may be more effective than actually asking students to take part in the goal setting process and then adjust their learning based on the performance data.

Peer-Assisted Mathematics Instruction

Eight studies met the criteria for inclusion in this category. Two studies used cross-age tutoring, and six studies focused on peer-assisted learning within a class.

Cross-age tutoring. The two studies that investigated cross-age tutoring yielded impressive effect sizes (1.15 & 0.75). The tutees in both studies were elementary students, and the tutors were well-trained upper elementary students. The main difference between the two interventions is that Bar-Eli and Raviv (1982) trained the tutors to actually teach lessons to the student with LD, whereas Beirne-Smith (1991) provided tutors with a detailed protocol which specified the types of feedback to provide students when they experienced difficulty or made mistakes. Beirne-Smith also provided tutors with ideas on how to explain problem solving strategies. In both studies a good deal of training was provided to the tutors.

Peer-assisted learning within a class. In sharp contrast to cross-age tutoring, the mean effect size for this category was 0.14, which was not statistically significant ($p = .27$). The Q statistic (2.66) also was not significant ($p = .75$), indicating that the outcomes were relatively homogeneous. Based on the evidence to date, peer-assisted learning within a class does not appear to result in beneficial impacts for students with LD.

A critical feature in most of the studies we reviewed was the amount and extensiveness of the training provided to students who assumed the role of tutor. There was also extensive variation in the roles and responsibilities of the members of the team or group. In earlier studies highly constricted roles were given to the peer tutor. For example, Slavin et al. (1984a, 1984b) limited the role of the partner or tutor to providing feedback on accuracy of a student's responses. More recent studies involved more complex roles for the tutor. One other interesting factor to consider in the evolution of this research is that the early research of Fuchs, Fuchs, Phillips, Hamlett, and Karns (1995) used the conventional tutor-tutee model where the tutor was the relative "expert" and the tutee the relative novice. In the latter studies by Fuchs and colleagues, tutoring is reciprocal in nature. In other words, students alternate between assuming the role of tutor and tutee.

This is one of the few areas where the mean effect size is not significantly different from zero, and effect sizes are consistently more modest than they are in other categories. It seems reasonable to conclude that more research needs to be done to examine whether peer-assisted learning within a class is an effective practice for students with LD.

Findings from Regression Analysis[2]

As noted earlier, many of the instructional variables (e.g., explicit instruction, verbalization) were often a part or component of a complex instructional intervention. In the previous section we presented data on the effectiveness of each instructional component in isolation. We now present the data from multiple regressions that essentially answer questions regarding the relative contribution of each instructional component in the overall effectiveness of the complex multi-component instructional interventions.

Mean Effects of Mathematics Instruction Programs

Regression analysis was conducted on 41 independent effects. The overall or unmoderated random effects mean for this subset was 0.63 ($p < .001$), indicating that mathematics interventions were generally effective across students, settings, and measures. The effect sizes ranged widely from - 0.29 to 2.45.

As expected there was considerable variation in treatment effects. A test of the heterogeneity of these effects across all instructional components was significant ($Q_{(40)} = 149.70$). This variation may be attributed to method differences and/or general and specific treatment characteristics. To better estimate categorical treatment effects, we developed an incremental regression model to assess the net outcomes of method and general characteristics assumed to be influential. These variables were assessed for their correlation with effect size (r or $\beta > .10$) and negligible correlations with each other. Size of correlation rather than statistical significance was used as the criterion for continued inclusion in the model since the latter criterion posed an undue initial burden on the single moderator in the initial stages of model development. Statistical inference testing was reserved for the later fuller model.

The Relation of Methodological Factors to Effect Size

We began by examining the correlations of effects with two research design characteristics: (a) quasi-experimental versus experimental design and (b) whether or not the control group received any mathematics instruction relevant to the outcome measures. We tested moderators using random weighting (i.e., assuming that effect size variance unexplained by moderators is randomly distributed). While studies using an experimental design had generally smaller effect sizes than quasi experiments ($\triangle g$ = -0.40), this difference was not statistically different from zero (p = .33). However, studies that provided a meaningful treatment in the control condition tended to have significantly smaller effect sizes when compared to studies that did not ($\triangle g$ = -0.99; p = .01).

The Relation of Study Characteristics to Effect Size

Substantive study characteristics were analyzed while adjusting for the presence of a meaningful control. Characteristics considered included publication characteristics (i.e., year and type), measurement characteristics (i.e., researcher developed measures, computation measures, word problems), student grade level, and treatment characteristics (i.e., number of sessions, treatment components, and interventionist characteristics). Each of these was tested individually to avoid the confounding influence of other study characteristics using a mixed-weight regression analysis.

Year of publication was positively associated with effect size (β = .21) indicating recent studies had somewhat larger effects than earlier studies ($\triangle g$ = 0.02), and peer reviewed studies reported larger effects than dissertations ($\triangle g$ = 0 .27; β = .14).

Although interesting, these distinctions were not considered theoretically useful as controls and thus were not carried forward in model development. We did consider measurement characteristics, student grade level, and treatment characteristics as potentially relevant for model development. Effects from norm-referenced measures of mathematics proficiency were generally smaller than those from researcher-developed measures ($\triangle g$ = -0 .35; β = -0.23). This replicates an earlier finding by Swanson & Hoskyn (1998). Effects for measures of word problems were associated with much larger effects than those from other domains ($\triangle g$ = 0.42; β = 0.28). Studies addressing older students had generally smaller effects than those for younger students, with effect sizes decreasing .07 standard deviations per grade level increase ($\triangle g$ = - 0.07; β = -0.23).

Treatments having a longer duration yielded generally smaller effects ($\triangle g$ = -0.003; β = -0.19) than brief duration studies. The number of treatment sessions correlated negatively with effect size (β = -0.19), indicating longer treatments were generally less effective. However, the choice of the interventionist (whether a member of a research team or a classroom teacher) appeared to have minimal impact on treatment outcomes (β = 0.16).

A subset of these method moderators correlating with treatment effects was then tested simultaneously by regression analysis. This preliminary set of moderators was then reduced even further to include only those that remained statistically significant when considered jointly in this regression analysis. This final set of control moderators thus selected included (a) whether or not the control group received any mathematics instruction relevant to the outcome measures (△*g* = -0.63; *β* = -0.22), (b) use of norm referenced measures as an outcome (△*g* = -0.21; *β* = -0.13), (c) use of word problems as an outcome (△*g* = 0.37; *β* = 0.25) and (d) grade level of students (△*g* = -0 .07; *β* = -0.23). Before considering treatment intervention components, these method moderators accounted for 27% of total between-study variance, but left a substantial amount of variance unexplained ($Q_{residual(36)}$ = 108.64; $p < .01$).

Thus a final comparison of treatment intervention components was tested while controlling for these three factors. Among these moderators only the presence of a norm referenced mathematics outcome measure remained significant for explaining effect size variation after entering treatment components (see Table 2).

The Relation of Treatment Interventions to Effect Size

Previously we considered each instructional component separately. However, many of the studies in the set included multiple instructional components, that is, they were overlapping or non-exclusive components. For example, a study might have included both explicit instruction and teacher use of visual representations. In order to analyze treatments as components a series of dummy codes were examined representing the 12 instructional components (for e.g., explicit instruction, verbalizations) and tested simultaneously controlling for method and treatment characteristics. While this moderated treatment model accounted for the majority of between-study variance in *g* ($Q_{model\ (16)}$ = 100.06; p <.001; R^2 = .69), the unexplained residual variance in *g* was also large suggesting unconsidered factors contributed to observed effects ($Q_{\text{residual (24)}}$ = 49.63, p = .002) necessitating a mixed-weight analysis.

As indicated by B-weights and confidence intervals presented in Table 2, majority of contrasts did not deviate significantly from the intercept. In other words, they were not significantly different from the mean effect size of 0.51. However, there were several exceptions. Use of heuristics was associated with an effect increase of 1.21 above the average adjusted effect of 0.51 ($p < .001$). Studies incorporating explicit instruction had larger treatment effects as well ($\triangle g = 0.53$; $p < .05$). At the other extreme, use of visuals by teachers only or by teachers and students together was associated with negligible predicted impacts (- 0.17 or -0.15, respectively) that were smaller than other treatments ($p = .06$). It appears that this component was not effective unless combined with other instructional components (e.g., explicit instruction, careful sequencing of examples). The impact of the instructional component cross-age tutoring approached significance, $p < .10$. It is noteworthy that treatment components varied widely in the number of associated effects and observations contributing to the standard error associated with each (see Table 2). The large effects associated with cross-age tutors despite poor precision indicate the potential of this treatment component.

Discussion

The major focus of this meta-analysis was on analyzing instructional components in mathematics intervention studies conducted with students with learning disabilities. Each intervention study was coded for a series of instructional components. We operationalized instruction broadly, using common dimensions from contemporary curriculum analysis (e.g., think alouds, explicit instruction, teaching of multiple heuristics, sequencing of examples) as well as other key aspects of instruction that transcend specific curricula (e.g., peer-assisted learning, formative assessment). Most interventions contained two, three, or even four of these components.

We analyzed specific instructional components because we saw little benefit in analyzing interventions based on specific researcher developed programs or practices. Our interest was in the detailed curriculum design and teaching practices that resulted in enhanced mathematics proficiency. In this way, our work resembled the seminal meta-analysis of intervention research for students with LD conducted by Swanson and Hoskyn (1998). However, a major difference between our analysis and the analysis conducted by Swanson and Hoskyn is that we limited the domain to instructional interventions in mathematics allowing us to focus on essential attributes of effective practice.

We examined the effectiveness of each instructional component at first in isolation. As each intervention can be conceived as a unique set of instructional components, we built a model using a series of hierarchical multiple regressions to discern the relative impact of each component. When examined individually, results indicated that only two instructional components did not yield a mean effect size significantly greater than zero: a) asking students to set a goal and measure attainment of that goal and b) peer-assisted learning within a class. All other instructional components that appear in Table 1 produced significant positive impacts on mathematics proficiency. The instructional components did however vary greatly in their effects, ranging from mean effect sizes of 0.14 to 1.56.

The small non-significant findings for goal setting may indicate that students with LD–who struggle with organizing abstract information–are simply overwhelmed by the abstractness of setting a reasonable goal and measuring attainment of that goal. Perhaps they become frustrated and demoralized by their low rate of progress or even one data point that happens to be particularly low on a given day. Although peer-assisted learning in a classroom did not harm students with LD, the average benefit was meager (*0.14)* and not significantly different than zero. Interestingly, within classroom peer-assisted learning produced a statistically significant impact with low achieving students (Baker et al., 2002). This apparent discrepancy may be influenced by the fact that students with LD are simply so far below the average performance of their classmates that feedback and prompting from a peer is insufficient to help them understand concepts that are difficult for them. In contrast, tutoring by a well-trained older student or adult appears to accelerate mathematics proficiency significantly.

When the instructional components were contrasted with each other in the regression analysis (Table 2), we found a majority of the instructional components to be non-significant. This does not imply that the instructional components were ineffective, but rather that they offered obvious advantage or disadvantage compared to other instructional components and were associated with average effects for the population of effects sampled in this study. Two instructional components provided significant, unique contributions– teaching students use of heuristics to solve problems and explicit instruction (which teaches one approach to a given problem type but also addresses distinguishing features of that problem type). The unique contribution of expert cross-age tutors was marginally significant.

Some of the findings highlighted in this meta-analysis are consonant with recommendations made in the practice guide on “Organizing Instruction and Study to Improve Student Learning” (Pashler, et al., 2007) developed for Institute of Education Sciences, based on recent findings from cognitive research. For example, the authors concluded that use of graphic presentations to illustrate new processes and procedures, and encouraging students to think aloud in speaking or writing their explanations tend to be effective across disciplines. They also suggest teaching both abstract and concrete representations of concepts, as the former facilitated initial learning while the later enhanced performance in new contexts. Similar outcomes were also observed in Swanson and Hoskyn’s (1998) meta-analysis of instructional research for students with LD. They found that direct instruction and cognitive strategy instruction tended to produce positive outcomes across all instructional disciplines. Also, Xin and Jitendra (1999) found beneficial impacts for representation techniques and strategy training, as was the case in this meta-analysis.

Role of Methodological and Study Characteristics

The role of methodological and study characteristics (e.g., relevant control group, type of design, type of measures used, student grade level, treatment characteristics) was assessed independently and simultaneously in our regression analysis. The estimated influence of many of these variables was statistically significant when tested individually; but when tested simultaneously, only the presence of norm-referenced measures approached significance for explaining effect size variation ($p < .10$). We found, as did Swanson and Hoskyn (1998), that use of norm-referenced achievement tests led to significant decreases in effect size. Typically, the norm referenced measures were less closely aligned to the content taught, and resulted in a significant negative regression coefficient of -0.44, meaning that overall, impacts on norm-referenced measures were lower than on researcher-developed measures. Effects for measures of word problems were associated with much larger effects than those from other domains. However, when all outcomes were tested simultaneously, the difference was found to be non-significant ($p = .19$). We speculate on several possible reasons for the larger impact. One reason could be that 12 out of the13 word problem measures were also researcher-developed measures, and, as previously discussed, the researcher-developed measures were typically more closely aligned with the content taught and resulted in larger effect sizes ($p = .06$). On the other hand, it is possible that the interventions involving instruction in word problem tended to be among the most effective in the set of studies. Many were quite contemporary and reflected insights from cognitive psychology in innovative ways.

When tested in isolation, effect sizes decreased .07 standard deviations per grade level increase ($p < .05$). However when evaluated conditionally with other moderators, grade level no longer accounted for the effect size above chance variation ($p = .76$). In coding the studies, we noted that in a small number of cases the mathematics content that was taught systematically and thoroughly to the experimental students was only covered in a cursory fashion with the control students. We determined if the content covered in the control group was consistently relevant or minimally relevant to the purpose of the study. Regression coefficients were not significantly different than zero in both instances.

Implications of the Instructional Components Analysis for Improving Practice

We would like to draw attention to five instructional components in order of importance. All these components had significant effect sizes.

Explicit instruction. Explicit instruction, a mainstay feature in many special education programs, once again was a key feature of many studies included in this meta-analysis. To create a common basis for comparisons, we defined explicit instruction in the following way: (a) the teacher demonstrated a step-by-step plan (strategy) for solving the problem; (b) the plan was problem-specific and not a generic, heuristic guide for solving problems; and (c) students were actively encouraged to use the same procedure/steps demonstrated by the teacher. Explicit instruction was often implemented in conjunction with other instructional components (for e.g., visual representations, student verbalizations) in many of the studies we reviewed.

Overall, the studies that used explicit instruction as an instructional delivery tool resulted in significant effects and produced a mean effect size of 1.22. Data from the multiple regression analysis strongly suggest that explicit instruction consistently contributed to the magnitude of effects regardless of whether it was paired with other instructional components. These findings confirm that explicit instruction is an important tool for teaching mathematics to students with learning disabilities.

Given its potential to impact student math performance, it is important to explore the evolution of the term explicit instruction. The meaning of explicit instruction has shifted over the years from behavioral and cognitive-behavioral interventions that were in essence content free to principles of Direct Instruction (e.g., Engelmann & Carnine, 1982) in which explicit teaching is content driven (e.g., mathematics) through the optimal sequencing of examples to help students understand critical features in the discipline. However, like the behavioral models, this approach is not rooted in the research from child development or the work of mathematics educators. Several recent studies (e.g., Owen & Fuchs, 2002; Woodward, 2006; Xin et al., 2005) artfully integrate research from child development and mathematics education with the direct instruction tradition, a tradition that continues to play a major role in special education research.

While these findings confirm that explicit instruction is an important tool for teaching mathematics to students with learning disabilities, it is important to note that there is no evidence supporting explicit instruction as the only mode of instruction for these students. There is good reason to believe that the construct of *explicit instruction* will continue to evolve in both research and practice, and the breakdown between explicit instruction and use of heuristics will continue to blur future research studies and practice.

Visual representations. Teachers have used visual representations of problems to illustrate solution strategies for mathematical problems intuitively for many years. Our findings from the meta-analysis do support the use of visual representations by teachers and students. When used in isolation, use of visuals during instruction led to consistent significant effects (mean $g = 0.47$). However, the multiple regression analysis suggests that better effects were obtained when visuals were used in combination with other instructional components than when used alone. For example, studies in which visuals were not paired with other instructional components (D. Baker, 1992; Lambert, 1996; Manalo et al., 2000) resulted in lesser impacts than studies in which visuals were paired with other instructional components.

Results also suggest that the specificity of the visuals plays a major role in how they affect learning. For example, Xin et al. (2005) had two conditions that used visuals; however, the experimental group was exposed to a visual that was more specific and based on our understanding of how experts solve mathematical problems. Also in the D. Baker (1992) study, students were given multiple visuals but not directed on which ones to use. This less specific approach resulted in a smaller impact, supporting the hypothesis regarding visual specificity. (Future researchers may want to examine the role of visual specificity.) In general, visual diagrams resulted in bigger positive effects when visuals were part of a multi-component approach to instruction.

The use of visual representations is also being emphasized in the field of mathematics education; for example, there is an increased emphasis by mathematicians on the importance of the number line, which attempts to provide early grounding in the concept that mathematics problems have a visual foundation. Also, Witzel et al. (2003) cite the work of Bruner, who argued that mathematical principles are best understood by having students work with concrete objects, and then transferring this knowledge systematically to graphic representations and, finally, to abstract arithmetic symbols. Thus, the results of the present meta-analysis confirm what teachers have sensed for many years; using graphic representations and teaching students how to understand them can help students with LD.

Sequence and/or range of examples. Thoughtfully planning instruction in mathematics, by carefully selecting and sequencing instructional examples appears to impact mathematics performance. The mean effect size for this group of studies was 0.82. The regression analysis indicated that this instructional component produced a regression weight of 0.42.

We believe that the sequence of examples may be most important during early acquisition of new skills when scaffolding is critical for student success. The range of examples taught is probably most critical to support transfer of newly learned skills. In other words, if the teacher teaches a wide range of examples, it will result in the learner being able to apply a skill to a wider range of problem types. Given the nature of students' concerns about their ability to be successful, early success with new concepts and problems can be supported by sequencing examples and problems with increasing complexity and ensuring that students have an opportunity to apply their knowledge to the widest range of problems to promote transfer of their knowledge to unfamiliar examples.

Consideration of sequence and range of examples presented is also highlighted in the seminal works on effective instruction by Engelmann and Carnine (1982) and also serves as a major priority area in the lesson study professional development model discussed by Lewis, Perry, and Murata (2006). Both of these planning devices, sequence of examples and range of examples, should be considered carefully when teaching students with LD.

Student verbalizations. Many students with learning disabilities are impulsive behaviorally and when faced with multi-step problems frequently attempt to solve the problems by randomly combining numbers rather than implementing a solution strategy step-by-step (Fuchs et al., 2003). One very promising finding is that the process of encouraging students to verbalize their thinking or their strategies, or even the explicit strategies modeled by the teacher, was always effective (mean $g = 1.04$). This included generic problem solving strategies that were derived from cognitive psychology as well as the more "classic" direct/explicit instruction strategies where students were taught one specific way to solve a problem. Although the meta-analysis did not suggest that thinking aloud made a unique contribution to effectiveness, we need to keep in mind that both explicit instruction and use of heuristics almost invariably involve encouragement of student verbalization.

Verbalization may help to anchor skills and strategies both behaviorally and mathematically. The consistently positive effects suggest that verbalizing steps in problem solving may be addressing students' impulsivity directly; suggesting that verbalization may serve to facilitate students' self-regulation during problem solving. Unfortunately, it is not common to see teachers encouraging verbalization in special education. Our findings would suggest that it is important to teach students to use language to guide their learning.

Providing ongoing feedback. One clear finding was that providing teachers with specific information on how each student was performing enhanced student math achievement (mean $g = 0.23$). Furthermore, providing specific information to special educators produced even stronger effects. Regarding the added benefit with special educators, it may that because special education teachers are prepared to use detailed student performance data to set individual goals for students, their familiarity with using information on performance is particularly useful for this group.

Providing general education teachers with detailed information on progress for the students with disabilities in their class had, on average, an extremely small impact on student performance. In addition to general education teachers being less familiar with data than special education teachers, there are several additional reasons for the smaller effect. It may be that the content of the math curricula is too difficult for the students with learning disabilities. A series of observational studies of mathematics instruction with students in the intermediate grades (Williams & Baxter, 1996; Woodward & Baxter, 1997) suggests that there is often misalignment between students' instructional level and their knowledge and skills. Another factor is that the few studies in this category were large-scale field experiments, which tend to produce smaller effects. Variation in implementation quality may have dampened the impact.

In summary, findings converge regarding the practice of providing teachers with precise information on student progress and specific areas of students' strengths and weaknesses in mathematics for enhancing achievement for this population. This is likely to be particularly true if the information as to which topics or concepts require additional practice or re-teaching is precise. It appears that teachers and students also benefit if the teachers are given specific guidance on addressing instructional needs or curricula so that they can immediately provide relevant instructional material to their students. As schools or districts begin developing and implementing progress monitoring systems in mathematics, it might be beneficial if they include not only graphs of student performance, but specific instructional guidelines and curricular materials for teachers or other relevant personnel (special educators who may co-teach or provide support services, peer tutors, cross-age tutors, adults providing extra support) to use with particular students.

Likewise, providing students with LD with similar feedback about their performance produced small impacts. It is interesting to note though that the largest effect in this category was related to non-specific feedback on effort, rather than on specific performance. One possible benefit of effort related feedback could be that it encourages and motivates students with LD to stay on tasks that they find frustrating. However, given that only one study examined the issue of effort related feedback, this approach merits further research attention. Essential to note also is that there seems to be no benefit in providing students with LD specific feedback that is specifically linked to their goal attainment.

Future Research Needs

Meta-analysis is essentially ahistorical because it treats each study as a data source and attempts to impartially locate the impact of various instructional approaches or components. Yet, as we reviewed the findings, several important historical trends emerged that help us interpret the findings.

Use of heuristics. We begin by noting that heuristic strategies provided a mean effect size of 1.56 and make a unique contribution to the effectiveness of an intervention. The heuristics used in these studies addressed a key problem for many students with LD – a weak ability to organize abstract information and to remember organizational frameworks or schema. A distinguishing feature of this set of studies was the accompanying student verbalizations. Students were given opportunities to verbalize their solutions or talk through the decisions they made in choosing a procedure. They were also asked to reflect on their attempts to solve problems and the decisions they made. The underlying concept is that through this process of verbalization and reflection, students with learning disabilities appear to arrive at a higher level of understanding and gain important insights into mathematical principles.

One of the most appealing aspects of this line of research is that it reflects the 2001 report from the National Research Council, Adding it Up, where there is a clear emphasis on teaching students the flexible use of multiple strategies. It also can, and often does, include insights gained from developmental psychology on how students develop mathematical knowledge and the nature of mathematical disabilities (e.g., Geary, 2005). This approach differs from the guided inquiry approach present in several widely used mathematics curricula in that students see many models for solving problems before they are asked to figure out the best procedure to use. However, part of each lesson in the traditional guided inquiry curricula does involve discussion of reasons for the choice the students make.

Given the small number of studies in this set, one should not overgeneralize the beneficial impacts. What remains unclear about the heuristic strategy approach is whether it involves teaching a multi-step strategy or teaching multiple skills that can be employed to derive the solution to a problem. Also, the success of this approach appears, at least on the surface, to be at odds with the notion that students with LD have difficulty with cognitively demanding routines. The flexible use of heuristic strategies would seem to place a cognitive load on students with LD that would make learning difficult. For example, the practical implication may be that learning 7 X 8 as a memorized fact may be less cognitively demanding than learning to decompose 7 X 8 as (7 X 7) + (7 X 1) making memorization a more effective tool. It is not entirely clear why this approach was so successful. This should be explored in future research. Another important step in this line of research is to assess whether students can transfer and generalize to previously untaught problem types, and whether these approaches actually do succeed in building an understanding of the underlying mathematical principles (e.g., an understanding of the distributive property).

Peer-assisted mathematics instruction. For students with LD, within class peer-assisted learning has not been as successful as it has been with other populations of students. However, part of the potential benefit could be that the structure and format of peer-assisted learning provides a natural and obvious way for students to engage in mathematics discourse that does appear to be beneficial for students with learning disabilities. We believe it is the use of mathematical language that may explain why in some cases peer tutoring activities can be successful. This is potentially very important if our hunches about the importance of verbalization we described in the previous section are true.

Our interpretation of significant findings in this meta-analysis is related to other findings regarding the degree of explicitness and scaffolding that appears to support the mathematics development of students with LD. It seems likely that peer tutoring efforts may fall short of the level of explicitness necessary to effectively help students with LD progress. This interpretation is supported by the more positive effects of cross-age peer tutoring wherein the tutor is an older student who has received extensive training in how to provide tutoring. Although there are relatively few studies in this area, cross-age tutoring appears to work more effectively. It may be that this is because the older tutor is better able to engage the learner in meaningful mathematical discourse. Practically speaking, however, cross-age tutoring presents practical logistical difficulties for implementation. Future research should explore the impact of peer-assisted instruction (cross-age and within classroom) when it is linked with a very strong explicit instruction component.

Limitations of Meta-analysis

The meta-analysis included studies identified from two literature searches (1971-1999; 2000-2007). During the second search dissertations were excluded from the search. This differential search procedure (i.e., the exclusion of dissertations in the second more recent search) might have resulted in an upward bias in the effect sizes as peer-reviewed studies reported larger effects than dissertations, or in a downward bias in effect sizes as recent studies had somewhat larger effects than earlier studies. However, a t-test indicated that that the effect sizes for dissertations were not significantly different than the effect sizes for peer-reviewed studies, $t(40) = -.524$, $p = .60$. Another limitation relates to the coding categories in the study. The conceptual framework underlining the coding categories (e.g., explicit instruction, use of visuals, etc) was influenced by three of the authors who have significant experience in effective classroom design and instruction. A behavioral background would have resulted in other coding perspectives (e.g., reinforcement, feedback, and drill repetitions). Finally, the findings of this meta-analysis could be an artifact of the particular sample of studies we used, and because many studies included multiple components, isolating the unique contribution of visual representations during instruction is a significant challenge under the best of circumstances.

Conclusions

Authors of the studies – as do all authors of intervention research – struggle to find the precise language for describing what they attempted to do in the instructional intervention. By coding studies according to these major themes, we attempted to begin to unpack the nature of effective instruction for students with learning disabilities in mathematics. Certainly, we need to do a good deal of additional unpacking to more fully understand the nature of the independent variable(s). As instructional researchers work more closely with mathematicians and cognitive psychologists, we believe this unpacking of major themes will continue.

However, the next major task is, in our view, increased use of the instructional components or techniques to tackle areas that are particularly problematic for students with LD such as word problems, concepts and procedures involving rational numbers, and understanding of the properties of whole numbers such as commutativity. The set of studies included in this meta-analysis indicates that we do have the instructional tools to address these content areas in mathematics. These topics will be critical areas as we move towards response–to-intervention models and three-tiered instruction for student in the area of mathematics.

References

References marked with an asterisk indicate studies included in the meta-analysis.

*Allinder, R. M., Bolling, R., Oats, R., & Gagnon, W. A. (2000). Effects of teacher self-monitoring on implementation of curriculum-based measurement and mathematics computation achievement of students with disabilities. *Remedial and Special Education*, *21*, 219-226.

*Bahr, C. M., & Rieth, H. J. (1991). Effects of cooperative, competitive, and individualistic goals on student achievement using computer-based drill-and-practice. *Journal of Special Education Technology, 11*, 33-48.

*Baker, D. E. (1992). *The effect of self-generated drawings on the ability of students with learning disabilities to solve mathematical word problems*. Unpublished doctoral dissertation, Texas Tech University.

Baker, S., Gersten, R., & Lee, D. (2002). A synthesis of empirical research on teaching mathematics to low-achieving students. *Elementary School Journal, 103*, 51-73.

Baker, S., Gersten, S., & Scanlon, D. (2002). Procedural facilitators and cognitive strategies: Tools for unraveling the mysteries of comprehension and the writing process, and for providing meaningful access to the general curriculum. *Learning Disabilities Research and Practice*, *17*, 65-77.

*Bar-Eli, N., & Raviv, A. (1982). Underachievers as tutors. *Journal of Educational Research, 75*, 139-143.

*Beirne-Smith, M. (1991). Peer tutoring in arithmetic for children with learning disabilities. *Exceptional Children, 57*, 330-337.

*Bottge, B. A., Heinrichs, M., Mehta Z. D., & Hung, Y. (2002). Weighing the benefits of anchored math instruction for students with disabilities in general education classes. *The Journal of Special Education, 35*, 186-200.

Brown, V., Cronin, M., & McEntire, E. (1994). *Test of Mathematical Abilities – Second Edition*. Austin, TX: PRO-ED Inc.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models*. Newbury Park, CA: Sage Publications.

Bryant, D. P., Bryant, B. T., & Hammill, D. D. (2000). Characteristic behaviors of students with LD who have teacher-identified math weaknesses. *Journal of Learning Disabilities, 33*, 168-177, 199.

Busse, R. T., Kratochwill, T. R., & Elliott, S. N. (1995). Meta-analysis for single-case consultation outcomes: Applications to research and practice. *Journal of School Psychology, 33*, 269-286.

*Butler, F. M., Miller, S. P., Crehan, K., Babbitt, B., & Pierce, T. (2003). Fraction instruction for students with mathematics disabilities: Comparing two teaching sequences. *Learning Disabilities Research and Practice*, *18*, 99-111.

*Calhoon, M. B., & Fuchs, L., S. (2003). The effects of peer-assisted learning strategies and curriculum-based measurement on the mathematics performance of secondary students with disabilities. *Remedial and Special Education, 24*, 235-245.

Cawley, J. F., Parmar, R. S., Yan, W., & Miller, J. H. (1998). Arithmetic computation performance of students with learning disabilities: Implications for curriculum. *Learning Disabilities Research & Practice, 13*, 68-74.

Cooper, H., & Hedges, L. V. (Eds.). (1994). *Handbook of research synthesis*. New York: Russell Sage Foundation.

Cortina, J. M., & Nouri, H. (2000). *Effect size for ANOVA designs* (Vol. 129). Thousand Oaks, CA: Sage Publications Inc.

Donner, A., & Klar, N. (2000). *Design and analysis of cluster randomized trials in health research.* London: Arnold Publishing.

Engelmann, S., & Carnine, D. (1982). *Theory of instruction.* New York: Irvington.

Friedman, F. (1992). *An evaluation of the integrated cognitive-behavioral model for improving mathematics performance and attentional behavior of adolescents with learning disabilities and attention-deficit hyperactivity disorders*. Unpublished doctoral dissertation, Columbia University Teachers College.

Fuchs, L.S., Compton, D. L., Fuchs, D., Paulsen, K., Bryant, J.D., & Hamlett, C.L. (2005). The Prevention, Identification, and Cognitive Determinants of Math Difficulty. *Journal of Educational Psychology, 97,* 493-513.

Fuchs, L.S., & Fuchs, D. (2003). Enhancing the mathematical problem solving of students with mathematics disabilities. In H.L. Swanson, K.R. Harris, & S.E. Graham (Eds.), *Handbook on learning disabilities* (pp. 306-322). New York: Guilford.

Fuchs, L. S., Fuchs, D., Hamlett, C. L. (1989). Effects of alternative goal structures within curriculum-based measurement. *Exceptional Children, 55*, 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 mathematics disabilities. *Learning Disabilities Research and Practice*, *17,* 90-106.

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

*Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Stecker, P. M. (1990). The role of skills analysis in curriculum-based measurement in math. *School Psychology Review, 19*, 6-22.

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

*Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Whinnery, K. (1991). Effects of goal line feedback on level, slope, and stability of performance within curriculum-based measurement. *Learning Disabilities Research and Practice, 6*, 65-73.

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

Fuchs, D., Fuchs, L. S., Mathes, P. G., & Lipsey, M. W. (2000). Reading differences between underachievers with and without learning disabilities: A meta-analysis. In E. Schiller & S. Vaughn (Eds.), *Research syntheses in special education*. Mahwah, NJ: Lawrence Erlbaum Associates.

*Fuchs, L. S., Fuchs, D., Phillips, N. B., Hamlett, C. L., & 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*, 604-620.

*Fuchs, L. S., Fuchs, D., Prentice, K. (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*, 293-306.

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

*Fuchs, D., Roberts, P. H., Fuchs, L. S., & Bowers, J. (1996). Reintegrating students with learning disabilities into the mainstream: A two-year study. *Learning Disabilities Research and Practice, 11*, 214-229.

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

Fuson, K. C., & Willis, G. B. (1989). Second graders' use of schematic drawings in solving addition and subtraction word problems. *Journal of Educational Psychology, 81*, 514-520.

Geary, D. C. (2003). Learning disabilities in arithmetic: Problem solving differences and cognitive deficits. In K. Harris & S. Graham (Eds.), *Handbook of learning disabilities* (pp. 199-212). New York: Guilford Publishers.

Geary, D. C. (2005). Role of cognitive theory in study of learning disability in mathematics. *Journal of Learning Disabilities, 38*, 305-307.

Gersten, R., & Baker, S. (2001). Teaching expressive writing to students with learning disabilities: A meta-analysis. *Elementary School Journal*, *101*, 251-272.

Gersten, R., Baker, S. K., Pugach, M., Scanlon, D., & Chard, D. (2001). Contemporary research on special education teaching. In V. Richardson (Ed.), *Handbook for research on teaching* (4th ed., pp. 695-722). Washington, DC: American Educational Research Association.

Gersten, R., Clarke, B., & Mazzocco, M. (2007). Chapter 1: Historical and contemporary perspectives on mathematical learning disabilities. In D.B. Berch & M.M.M. Mazzocco (Eds.), *Why is math so hard for some children? The nature and origins of mathematical learning difficulties and disabilities* (pp. 7-29). Baltimore, MD: Paul H. Brooks Publishing Company.

Gersten, R., & Hitchcock, J. (2008). What is credible evidence in education: The role of What Works Clearinghouse in informing the process. In S. I. Donaldson, C. A.

Christie, & M. M. Mark (Eds.), What counts as credible evidence in applied research and evaluation practice? Thousand Oaks, CA: Sage Press.

Gleser, L. J., & Olkin, I. (1994). Stochastically dependent effect sizes. In H. Cooper & L. V. Hedges (Eds.), *The Handbook of Research Synthesis* (pp. 339-355). New York: Russell Sage Foundation.

Graham, S., & Harris, K. (2003). Students with learning disabilities and the process of writing; A meta-analysis of SRSD studies. In H. L. Swanson, K. R. Harris, & S. Graham (Eds.), *Handbook of learning disabilities* (pp. 323–344). New York: Guilford Press.

Greene, G. (1999). Mnemonic multiplication fact instruction for students with learning disabilities. *Learning Disabilities Research and Practice*, *14*, 141-148.

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. 24-49). Cambridge, MA: MIT Press.

Gross-Tsur V., Manor O., & Shalev, R. S. (1996) Developmental dyscalculia: Prevalence and demographic features. *Developmental Medicine and Child Neurology, 38*, 25–33.

Hedges, L.V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6*, 107-128.

Hedges, L. V., & Hedberg, E. C. (2007). Intraclass correlation values for planning group-randomized trials in education. *Educational Evaluation and Policy Analysis, 29*, 60-87.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. San Diego, CA: Academic Press.

*Hutchinson, N. L. (1993). Effects of cognitive strategy instruction on algebra problem solving of adolescents with learning disabilities. *Learning Disability Quarterly, 16*, 34-63.

Jastak, S., & Wilkinson, G. (1984). *Wide Range Achievement Test-Revised*. Wilmington, DE: Jastak Associates.

Jenkins, A. A. (1992*). Effects of a vocational mathematics curriculum on achievement of secondary learning-disabled students*. Unpublished doctoral dissertation, University of Texas at Austin.

*Jitendra, A. K., Griffin, C. C., McGoey, K., & Gardill, M. G. (1998). Effects of mathematical word problem solving by students at risk or with mild disabilities. *The Journal of Educational Research, 91*, 345-355.

*Kelly, B., Gersten, R., & Carnine, D. (1990). Student error patterns as a function of curriculum design: Teaching fractions to remedial high school students with learning disabilities. *Journal of Learning Disabilities, 23*, 23-29.

Kroesbergen, E. H., & Van Luit, J. E. H. (2003). Mathematics interventions for children with special educational needs: A meta-analysis. *Remedial and Special Education, 24*, 97-114.

*Lambert, M. A. (1996). *Teaching students with learning disabilities to solve word-problems: A comparison of a cognitive strategy and a traditional textbook method*. Unpublished doctoral dissertation, Florida Atlantic University.

*Lee, J. W. (1992). *The effectiveness of a novel direct instructional approach on math word problem solving skills of elementary students with learning disabilities*. Unpublished doctoral dissertation, Ohio State University.

Lewis, C., Perry, R., Murata, A. (2006). How should research contribute to instructional improvement? Case of lesson study. *Educational Researcher, 35*, 3-14.

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

Ma, L. (1999). *Knowing and teaching elementary mathematics.* Mahwah: Lawrence Erlbaum Associates, Inc.

Maccini, P., & Hughes, C. A. (1997). Mathematics interventions for adolescents with learning disabilities. *Learning Disabilities Research and Practice, 12*, 168-176.

*Manalo, E., Bunnell, J., & Stillman, J. (2000). The use of process mnemonics in teaching students with mathematics learning disabilities. *Learning Disability Quarterly*, *23*, 137-156.

*Marzola, E. (1987). *An arithmetic verbal problem solving model for learning disabled students*. New York: Columbia University Teachers College.

Mastropieri, M. A., Scruggs, T. E., & Shiah, S. (1991). Mathematics instruction for learning disabled students: A review of research. *Learning Disabilities Research and Practice, 6*, 89-98.

Miller, S. P., Butler, F. M., & Lee, K. (1998). Validated practices for teaching mathematics to students with learning disabilities: A review of literature. *Focus on Exceptional Children, 31*, 1-24.

National Research Council (NRC) (2001). Adding it up: Helping children learn mathematics. In J. Kilpatrick, J. Swafford & B. Findell (Eds*.), Mathematics Learning Study Committee, Center for Education, Division of Behavioral and Social Sciences and Education*. Washington, D.C.: National Academy Press.

National Mathematics Panel. (n.d.). Retrieved January 29, 2008, from http://www.ed.gov/about/bdscomm/list/mathpanel/index.html

*Omizo, M. M., Cubberly, W. E., & Cubberly, R. D. (1985). Modeling techniques, perceptions of self-efficacy, and arithmetic achievement among learning disabled children. *The Exceptional Child, 32*, 99-105.

Ostad, S. A. (1998). Developmental differences in solving simple arithmetic word problems and simple number-fact problems: A comparison of mathematically disabled children. *Mathematical Cognition, 4*, 1-19.

*Owen, R. L., & Fuchs, L. S. (2002). Mathematical problem-solving strategy instruction for third-grade students with learning disabilities. *Remedial and Special Education*, *23*, 268-278.

Palincsar, A. S. (1986). The role of dialogue in providing scaffolded instruction. *Educational Psychologist*, *21*, 73-98.

Pashler, H., Bain, P., Bottge, B., Graesser, A., Koedinger, K., McDaniel, M., et al. (2007*) Organizing Instruction and Study to Improve Student Learning* (NCER

2007-2004). Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ncer.ed.gov.

*Pavchinski, P. (1988). *The effects of operant procedures and cognitive behavior modification on learning disabled students' math skills*. Unpublished doctoral dissertation, University of Florida.

Raudenbush, S. W., & Bryk, A. S. (2002). Applications in meta-analysis and other cases where level-1 variances are known. In *Hierarchical linear models: applications and data analysis method* (pp. 205-227). Thousand Oaks, CA: Sage Publications.

*Reisz, J. D. (1984). *The effect of goal-setting activities on the locus of control and achievement of learning disabled middle-school students*. Unpublished doctoral dissertation, University of Alabama.

Robinson, D., Schofield, J., & Steers-Wentzell, K. (2005). Peer and cross-age tutoring in math: Outcomes and their design implications. *Educational Psychology Review, 17*, 327-362.

Rosenthal, R., & Rubin, D. B. (1986). Meta-analytic procedures for combining studies with multiple effects sizes. *Psychological bulletin, 99*, 400-406.

*Ross, P. A., & Braden, J. P. (1991). The effects of token reinforcement versus cognitive behavior modification on learning-disabled students' math skills. *Psychology in the Schools, 28*, 247-256.

*Schunk, D. H., & Cox, P. D. (1986). Strategy training and attributional feedback with learning disabled students. *Journal of Educational Psychology, 78*, 201-209.

Seely, J., Birkes, D. (1980). Estimability in partitioned linear models. *The Annals of Statistics, 8*, 399-406.

Shadish, W., R., Cook, T., D., & Campbell, D., T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston: Houghton Mifflin Company.

Silbert, J., Carnine, D., & Stein, M. (1989). *Direct instruction mathematics*. Columbus, OH: Merrill.

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

*Slavin, R. E., Madden, N. A., & Leavey, M. (1984b). Effects of team assisted individualization on the mathematics achievement of academically handicapped and non-handicapped students. *Journal of Educational Psychology, 76*, 813-819.

Swanson, H. L., & Hoskyn, M. (1998). Experimental intervention research on students with learning disabilities: A meta-analysis of treatment outcomes. *Review of Educational Research, 68*, 277-321.

Swanson, H. L., Hoskyn, M., & Lee, C. (1999). *Interventions for students with learning disabilities: A meta-analysis of treatment outcomes*. New York: Guilford Press.

*Tournaki, N. (1993). *Comparison of two methods of teaching addition to learning disabled and regular education students*. Unpublished doctoral dissertation, New York University.

*Tournaki, N. (2003). The differential effects of teaching addition through strategy instruction versus drill and practice to students with and without disabilities. *Journal of Learning Disabilities, 36*, 449-458.

* Van Luit, J. E. H., & Naglieri, J. A. (1999). Effectiveness of the MASTER program for teaching special children multiplication and division. *Journal of Learning Disabilities*, *32*, 98-107.

*Walker, D. W., & Poteet, J. A. (1989/1990). A comparison of two methods of teaching mathematics story problem-solving with learning disabled students. *National Forum of Special Education Journal, 1*, 44-51.

What Works Clearinghouse. (2006, September 13). *WWC teacher-intervention confound*. Retrieved January, 29, 2008, from http://ies.ed.gov/ncee/wwc/pdf/teacher_confound.pdf

What Works Clearinghouse. (2007, April 16). *Technical details of WWC-conducted computations*. Retrieved January, 29, 2008, from http://ies.ed.gov/ncee/wwc/pdf/conducted_computations.pdf

Wilkinson, G. S. (1993). *Wide Range Achievement Test, 1993 Edition*. Wilmington, DE: Wide Range Inc.

Williams, S. R., & Baxter, J. A. (1996). Dilemmas of discourse-oriented teaching in one middle school mathematics classroom. *Elementary School Journal, 97*, 21-38.

*Wilson, C. L., & Sindelar, P. T. (1991). Direct instruction in math word problems: Students with learning disabilities. *Exceptional Children, 57*, 512-518.

*Witzel, B., Mercer, C. D., & Miller, M. D. (2003). Teaching algebra to students with learning difficulties: An investigation of an explicit instruction model. *Learning Disabilities Research and Practice*, *18*, 121-131.

Wong, B., Harris, K., Graham, S., & Butler, D. (2003). Cognitive strategies instruction research in learning disabilities. In H. L. Swanson, K. R. Harris, & S. Graham (Eds.), *Handbook of learning disabilities* (pp. 383– 402). New York: Guilford Press.

Woodward, J., & Baxter, J. (1997). The effects of an innovative approach to mathematics on academically low-achieving students in mainstreamed settings. *Exceptional Children, 63*, 373-388.

*Woodward, J., Monroe, K., & Baxter, J. (2001). Enhancing student achievement on performance assessments in mathematics. *Learning Disabilities Quarterly*, *24*, 33-46.

Woodward, J. (2006). Developing Automaticity in multiplication facts: Integrating strategy instruction with timed practice drills. *Learning Disability Quarterly, 29*, 269-289.

Wortman, P. M., & Bryant, F. B. (1985). School desegregation and black achievement: An integrative review. *Sociological Methods and Research, 13*, 289-324.

Xin, Y. P., & Jitendra, A. K. (1999). The effects of instruction in solving mathematical word problems for students with learning problems: A meta-analysis. *The Journal of Special Education, 32*, 207-225.

*Xin, Y. P., Jitendra, A. K., & Deatline-Buchman, A. (2005). Effects of mathematical word problem-solving instruction on middle school students with learning problems. *The Journal of Special Education, 39*, 181-192.

Zimmerman, B. (2001). Theories of self-regulated learning and academic achievement: An overview and analysis. In B. Zimmerman & D. Schunk (Eds.), *Self-regulated learning and academic achievement: Theoretical perspectives* (2nd ed., pp. 1–37). Mahwah, NJ: Erlbaum.

Footnotes

[1]There were three studies (D. Baker, 1992; Fuchs, Roberts, Fuchs, & Bowers, 1996; Slavin, Madden, & Leavey, 1984b); where randomization was imperfect. In one case, random assignment was conducted in all but 2 of the 5 school sites. In another, all but two teachers were assigned randomly. In the third case, a replacement student was chosen to match a student who could not participate. We considered these studies as RCTs for the purpose of this meta-analysis.

[2]For a list of the studies associated with the various analyses in this section, please contact the author.

Table 1

Simple Comparisons of All Effects (g) and Heterogeneity (Q) for Orthogonal Effects and Effects Stratified by Method

	Random Effects (g)						*Heterogeneity*		
Instructional Component	*Random Effects Mean*	*Median*	*SE*	*p(g)*	*95% CI* Lower	Upper	*Q*	Critical Q	*p(Q)*
Explicit Instruction (n=11)	1.22***	1.39	0.23	0	0.78	1.67	41.68***	18.31	0.00
Use of Heuristics (n=4)	1.56***	1.62	0.46	0.00	0.65	2.47	9.10*	7.82	0.03
Student Verbalizations (n=8)	1.04***	1.42	0.32	0.00	0.42	1.66	53.39***	14.07	0.00
Visuals for Teacher and Student (n=7)	0.46***	0.67	0.12	0.00	0.23	0.69	9.88	12.59	0.13
Visuals for Teacher only (n=5)	0.41*	0.50	0.18	0.02	0.06	0.77	4.33	9.49	0.36
Visuals *Combined* (n=12)	0.47***	0.52	0.12	0.00	0.25	0.70	14.13	19.68	0.23
Sequence and/or Range (n=9)	0.82***	0.54	0.20	0.00	0.42	1.21	19.78**	15.51	0.01
Teacher Feedback (n=7)	0.21*	0.19	0.10	0.04	0.01	0.41	0.32	12.59	1.00
Teacher Feedback plus Options (n=3)	0.34~	0.40	0.21	0.10	-0.07	0.74	1.01	5.99	0.60
Teacher Feedback *Combined* (n=10)	0.23**	0.21	0.09	0.01	0.05	0.41	1.63	16.92	1.00
Student Feedback (n=7)	0.23**	0.17	0.09	0.01	0.05	0.40	3.60	12.59	0.73
Student Feedback with Goal Setting (n=5)	0.17	-0.17	0.30	0.29	-0.15	0.49	12.67**	9.49	0.01
Student Feedback *Combined* (n=12)	0.21*	0.14	0.10	0.04	0.01	0.40	16.37	19.68	0.13
Cross-age Tutoring[a] (n=2)	1.02***	0.95	0.23	0.00	0.57	1.47	0.68	3.48	0.41
Peer-assisted Learning within a Class (n=6)	0.14	0.17	0.11	0.27	-0.09	0.32	2.66	11.07	0.75

Note. SE = standard error; CI = confidence interval. n refers to number of effects.
[a]Fewer observed effects (n =2) reduces confidence Cross-grade estimates
~ p < .10. * p<.05. ** p<.01. *** p<.001.

Table 2

Model Comparison of Treatment Effects (n^a = 41) After Controlling for Select Research Method and Characteristics Moderators

				95% CI		
Moderator	*B weight*	N^b	*Standard Error*	*Lower Limit*	*Upper Limit*	*p*
Method						
Relevant[c] Control (1) vs. Minimally Relevant Control (0)	-0.17[d]	-	.38	-.92	.58	.65
Characteristics						
Word Problems (1) vs. Other (0)	0.45[d]	-	.34	-.22	1.19	.19
Norm-referenced Measure (1) vs. Researcher Developed (0)	-0.44[c]~	-	.24	-.90	.02	.06
Grade Level	0.01[d]	-	.04	-.07	.09	.76
Instructional Components						
Intercept	0.51*	-	.24	.04	.98	.03
Visuals for Teacher only (n = 12)	-0.68~	440	.40	-1.47	.11	.09
Student Feedback with Goals (n=3)	-0.29	110	.51	-1.30	.72	.57
Teacher Feedback (n=7)	-0.18	215	.32	-.80	.45	.58
Peer-assisted Learning within a Class (n=6)	-0.02	283	.38	-.77	.72	.57
Student Feedback (n=8)	0.03	353	.36	-.67	.73	.93
Visuals for Teacher and Student (n=7)	0.02	296	.53	-1.01	1.06	.96
Teacher Feedback plus Options (n=2)	0.16	36	.43	-.68	1.00	.70
Student Verbalizations (n=8)	0.24	389	.28	-.30	.78	.39
Sequence and/or Range (n=6)	0.42	228	.30	-.17	1.01	.16
Explicit Instruction (n=11)	0.53*	436	.24	.05	1.0	.03
Cross-age Tutoring (n=2)	0.80~	90	.43	-.04	1.64	.06
Use of Heuristics (n=4)	1.21***	88	.35	.52	1.90	.000

Note. Mixed-weighted correlations - Random intercept weight added after applying level-2 moderators; Q_{model} (*df* = 16) = 100.06, $p_{(Q)} < .001$; Q_{residual} (*df* = 24) = 49.63, $p_{(Q)} = .002$; $R^2_{\text{model}} = .67$

[a]Lowercase "n" refers to number of effects contributing to each estimate. [b]*N* is the aggregated sample of all studies contributing to each component effect $N = \sum_{j=1}^{n} sample_j$. [c]Relevant - content covered in the control group was consistently relevant to the purpose of the study; Minimally relevant - content covered in control group was minimally relevant to the purpose of the study. [d]Treatment characteristics have been centered so that summing of the intercept with component B weights yields predicted effect of treatment component holding other components constant (e.g., Visuals for teacher and student: g_{pred} = Intercept + Visuals for teacher + visuals for teacher and student = -.15; Visuals for teacher only: g_{pred} = Intercept + Visuals for teacher = -.17).

~ p < .10. * p<.05. ** p<.01. *** p<.001

Appendix A

List of mathematical interventions used in the meta-analysis[a]

#	Study	Coded Under Category	Math Domain	Student N	Grade	Design[b]	Unit of assignment/ analysis	Nature of control group[c]	Length	Instruction Provided by	Fidelity	Maintenance or Transfer Assessed	Reliability of Post Measures[d]	Interscorer Agreement	Type of Dependent Measures
1	Allinder, R. M., Bolling, R., Oats, R., & Gagnon, W. A. (2000).	Feedback to teachers	Operations	54	3-5	RCT[e]	Teachers/ Students	Relevant	36 weeks	Teacher	Yes	_	0.85	99%	Researcher Developed
2	Bahr, C. M. & Reith, H. J. (1991).	Feedback to students; Peer-assisted instruction	Operations	46[f]	7-8	RCT	Students/ Students	Relevant	12 sessions of 10 minutes	Computer	No	--	--	--	Researcher Developed & Norm-referenced
3	Baker, D. E. (1992).	Curriculum/ Instruction	Word problems	46	3–5	RCT	Students/ Students[g]	Relevant	2 sessions of 45 minutes	Researcher	No	--	0.82[h]	--	Researcher Developed
4	Bar-Eli, N., & Raviv, A. (1982).	Peer-assisted instruction	General math proficiency	60	2-6	RCT	Students/ Students	Relevant	33 sessions	Peer tutors	Yes	--	--	--	Norm-referenced
5	Beirne-Smith, M. (1991).	Curriculum/ Instruction; Peer-assisted instruction	Operations	30	1-5	RCT	Students & teachers/ Students & teachers	Relevant	4 sessions of 30 minutes	Peer tutors	Yes	--	--	--	Researcher Developed
6	Bottge, B. A., Heinrichs, M., Mehta, Z. D., & Hung, Y. (2002).	Curriculum/ Instruction	Word problems	8	7	RCT	Students/ Students	Relevant	12 sessions	Teacher	Yes	Maintenance, Transfer	0.73-0.92	98-99%	Researcher Developed
7	Butler, F. M., Miller, S. P., Crehan, K., Babbitt, B., & Pierce T. (2003).	Curriculum/ Instruction	Fractions	50[f]	6-8	RCT	Classes/ Students	Relevant	10 sessions of 45 minutes	Researcher	Yes	--	--	97%	Researcher Developed & Norm-referenced
8	Calhoon, M. B., & Fuchs, L., S. (2003).	Feedback to teachers; Feedback to students; Peer-assisted instruction	Operations; General math proficiency	92	9-12	RCT	Classes/ Students	Relevant	30 sessions of 30 minutes	Teacher	Yes	--	0.87, 0.92	97.2%, 96.4%	Researcher Developed & Norm-referenced
9	Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Appleton, A. C. (2002).	Curriculum/ Instruction	Word problems	38	4	RCT	Students/ Students	Relevant	24 sessions of 33 minutes	Researcher	No	Transfer	0.92, 0.95	96-99%	Researcher Developed

#	Study	Coded Under Category	Math Domain	Student N	Grade	Design[b]	Unit of assignment/ analysis	Nature of control group[c]	Length	Instruction Provided by	Fidelity	Maintenance or Transfer Assessed	Reliability of Post Measures[d]	Interscorer Agreement	Type of Dependent Measures
10	Fuchs, L. S., Fuchs, D., Hamlett, C. L., Phillips, N. B., & Bentz, J. (1994).	Feedback to teachers; Feedback to students	Operations	40	2-5	RCT	Teachers/ Students yoked to teachers	Relevant	25 weeks	Teacher	Yes	--	0.85	99%	Researcher Developed
11	Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Stecker, P. M. (1990).	Feedback to teachers	Operations	91[f]	3-9	RCT	Teachers/ Teachers	Relevant	15 weeks	Teacher	Yes	--	0.85	99%	Researcher Developed
12	Fuchs, L .S., Fuchs, D., Hamlett, C. L., & Stecker, P. M. (1991).	Feedback to teachers	Operations	63	2-8	RCT	Teachers/ Teachers	Relevant	20 weeks	Teacher	Yes	--	0.85	99%	Researcher Developed
13	Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Whinnery, K. (1991).	Feedback to students	Operations	36	2-8	RCT	Students/ Students	Relevant	20 weeks	Teacher	Yes	--	--	--	Researcher Developed
14	Fuchs, L. S., Fuchs, D., Karns, K., Hamlett, C. L., Katzaroff, M., & Dutka, S. (1997).	Feedback to students; Peer-assisted instruction	Operations; General math proficiency	40	2-4	RCT	Classes/ Classes	Relevant	46 sessions	Teacher	Yes	--	0.88	99%	Researcher Developed
15	Fuchs, L. S., Fuchs, D., Phillips, N. B., Hamlett, C. L., & Karns, K. (1995).	Feedback to teachers; Feedback to students; Peer-assisted instruction	Operations	40	2-4	RCT	Teachers/ Classes	Relevant	25 weeks	Teacher	Yes	Transfer	0.86	99%	Researcher Developed
16	Fuchs, L. S., Fuchs, D., Prentice, K. (2004).	Curriculum/ Instruction; Feedback to students	Word problems	45	3	RCT	Teachers/ Students	Relevant	32 sessions of 33 minutes	Teacher, Researcher	Yes	--	0.88-0.97	98%	Researcher Developed
17	Fuchs, D., Roberts, P. H., Fuchs, L. S., & Bowers, J. (1996).	Feedback to teachers	Operations	47	3-7	RCT[i]	Teachers & students/ Students	Relevant	36 weeks	Teacher	No	--	0.85	99%	Researcher Developed
18	Hutchinson, N. L. (1993).	Curriculum/ Instruction	Word problems	20	8-10	RCT	Students/ Students	Insufficient Information	60 sessions of 40 minutes	Researcher	No	--	--	--	Researcher Developed & Norm-referenced
19	Jitendra, A.K., Griffin, C.C.,	Curriculum/ Instruction	Word problems	34[f]	2-5	RCT	Students/ Students	Relevant	19 sessions	Researcher	Yes	--	0.77, 0.88[h]	Yes	Researcher Developed

#	Study	Coded Under Category	Math Domain	Student N	Grade	Design[b]	Unit of assignment/ analysis	Nature of control group[c]	Length	Instruction Provided by	Fidelity	Maintenance or Transfer Assessed	Reliability of Post Measures[d]	Interscorer Agreement	Type of Dependent Measures
	McGoey, K., & Gardill, M.G. (1998).								of 43 minutes						
20	Author. (1990).	Curriculum/ Instruction	Fractions	34[f]	9-11	RCT	Students/ Students	Relevant	10 sessions of 30 minutes	Teacher; Researcher	Yes	--	0.98	--	Researcher Developed
21	Lambert, M. A. (1996).	Curriculum/ Instruction	Word problems	76	9-12	QED	Classes/ Students	Relevant	8 sessions of 55 minutes	Teacher	No	--	0.73, 0.83	--	Researcher Developed
22	Lee, J. W. (1992)	Curriculum/ Instruction	Word problems	33	4-6	RCT	Classes/ Students	Minimally Relevant	9 sessions of 45 minutes	Researcher	No	--	0.59-0.91	95%	Researcher Developed
23	Manalo, E., Bunnell, J., & Stillman, J. (2000).[j] Experiment 1	Curriculum/ Instruction	Operations	29	8	RCT	Students/ Students	Relevant	10 sessions of 25 minutes	Researcher	No	Maintenance	0.71	--	Researcher Developed
24	Manalo, E., Bunnell, J., & Stillman, J. (2000). Experiment 2	Curriculum/ Instruction	Operations	28	8	RCT	Students/ Students	Relevant	10 sessions of 25 minutes	Researcher	No	Maintenance	0.71	--	Researcher Developed
25	Marzola, E. (1987).	Curriculum/ Instruction	Word problems	60	5-6	RCT	Schools/ Students	Minimally Relevant	12 sessions of 30 minutes	Teacher	No	--	--	--	Researcher Developed
26	Omizo, M. M., Cubberly, W. E., & Cubberly, R. D. (1985).	Curriculum/ Instruction	Operations	60	1-3	RCT	Students & teachers/ Students	Relevant	3 sessions of 30 minutes	Teacher, Researcher	No	--	--	--	Researcher Developed
27	Owen, R. L., & Fuchs, L. S. (2002).	Curriculum/ Instruction	Word problems	24[f]	3	RCT	Classes/ Students	Relevant	6 sessions of 30 minutes	Researcher	94.9%	--	0.89	99.5%	Researcher Developed
28	Pavchinski, P. (1988).	Curriculum/ Instruction	Operations	94	1-5	RCT	Teachers/ Students	Relevant	19 sessions of 60 minutes	Teacher	Yes	Maintenance	--	--	Researcher Developed & Norm-referenced
29	Reisz, J. D. (1984).	Feedback to students	General math proficiency	29	7-8	RCT	Students/ Students	Insufficient Information	16 sessions	Researcher	No	Maintenance	0.97	--	Norm-referenced

#	Study	Coded Under Category	Math Domain	Student N	Grade	Design[b]	Unit of assignment/ analysis	Nature of control group[c]	Length	Instruction Provided by	Fidelity	Maintenance or Transfer Assessed	Reliability of Post Measures[d]	Interscorer Agreement	Type of Dependent Measures
30	Ross, P. A., & Braden, J. P. (1991).	Curriculum/ Instruction	Operations	94	1-5	RCT	Teachers/ Students	Relevant	19 sessions of 60 minutes	Teacher	Yes	--	--	--	Researcher Developed & Norm-referenced
31	Schunk, D. H., & Cox, P. D. (1986).	Curriculum/ Instruction; Feedback to students	Operations	90	6-8	RCT	Students/ Students	Relevant	6 sessions of 45 minutes	Researcher	Yes	--	0.82[k]	--	Researcher Developed
32	Slavin, R. E., Madden, N. A., & Leavey, M. (1984a).	Feedback to students; Peer-assisted instruction	Operations; general math proficiency	113	3-5	QED	Schools/ Classes & Students	Relevant	24 weeks	Teacher	Yes	--	--	--	Norm-referenced
33	Slavin, R. E., Madden, N. A., & Leavy, M. (1984b).	Feedback to students; Peer-assisted instruction	Operations	117	3-5	RCT[l]	Schools/ Students	Relevant	10 weeks	Teacher	No	--	--	--	Norm-referenced
34	Tournaki, H. (1993).	Curriculum/ Instruction	Operations	42	3-5	QED	Students/ Students	Relevant	8 sessions of 15 minutes	Researcher	No	Transfer	0.93	98%	Researcher Developed
35	Tournaki, N. (2003).	Curriculum/ Instruction	Operations	42	3-5	RCT	Students/ Students	Relevant	8 sessions of 15 minutes	Researcher	No	Transfer	0.91	98%	Researcher Developed
36	Van Luit, J. E. H., & Naglieri, J. A. (1999).	Curriculum/ Instruction	Operations	42	3-5	RCT	Students/ Students	Insufficient Information	51 sessions of 45 minutes	Teacher	No	Transfer	--	--	Researcher Developed
37	Walker, D. W., & Poteet, J. A. (1989/1990).	Curriculum/ Instruction	Word problems	70	6-8	RCT	Teachers/ Classes & students	Relevant	17 sessions of 30 minutes	Teacher	Yes	Transfer	0.83, 0.91	--	Researcher Developed
38	Wilson, C. L., & Sindelar, P. T. (1991).	Curriculum/ Instruction	Word problems	62	2-5	RCT	Schools/ Students	Relevant	14 sessions of 30 minutes	Researcher	Yes	--	0.88	--	Researcher Developed
39	Witzel, B., Mercer, C. D., & Miller, M. D. (2003).	Curriculum/ Instruction	Algebra	68[f]	6 -7	RCT	Teachers/ Students	Relevant	19 sessions of 50 minutes	Teacher	Yes	--	--	--	Researcher Developed
40	Woodward, J. (2006).	Curriculum/ Instruction	Operations	15	4	RCT	Students/ Students	Relevant	20 sessions of 25 minutes	Teacher	Yes	Transfer	>0.90[j]	--	Researcher Developed

#	Study	Coded Under Category	Math Domain	Student N	Grade	Design[b]	Unit of assignment/ analysis	Nature of control group[c]	Length	Instruction Provided by	Fidelity	Maintenance or Transfer Assessed	Reliability of Post Measures[d]	Interscorer Agreement	Type of Dependent Measures
41	Woodward, J., Monroe, K., & Baxter, J. (2001).	Curriculum/ Instruction	Word problems	11	4	QED	Schools/ Students	Minimally Relevant	69 sessions of 30 minutes	Teacher, Staff	No	--	0.85-0.92	93%	Researcher Developed
42	Xin, Y. P., Jitendra, A. K., & Deatline-Buchman, A. (2005).	Curriculum/ Instruction	Word problems	22[f]	6-8	RCT	Students/ Students	Relevant	12 sessions of 60 minutes	Researcher	Yes	Maintenance, Transfer	0.84[h]	100%	Researcher Developed

Note. Dashes for Maintenance or Transfer assessed indicate the data was not obtained or did not meet our criteria for maintenance and transfer measure. Dashes for Reliability of Post Measures and Interscorer Agreement indicate data was not reported.

[a]Total number of research papers = 41; total number of mathematical interventions = 42. [b]RCT = randomized controlled trial; QED = quasi-experimental design. [c]Relevant = content covered in the control group was consistently relevant to the purpose of the study. Minimally relevant = content covered in control group was minimally relevant to the purpose of the study. Insufficient information = enough information on instruction in control group was not provided to make a decision regarding the relevancy of content to the purpose of the study. [d]Internal consistency coefficients, unless noted otherwise. [e]Random assignment was assumed because participants volunteered and random assignment was used to assign two interventions to treatment group. [f]Sample in this study was primarily LD, but not 100% LD. [g]For approximately 15% of the students, assignment was at the school level. Most students were randomly assigned individually, but two schools were randomly assigned as a whole. [h]Parallel forms reliability, [i]Teachers were randomly assigned with 2 exceptions. Two teachers were assigned based on their previous experience with the interventions. [j]Two mathematical interventions were reported in this research paper. [k]Test-retest reliability. [l]Randomly assigned; attrition of one school; replacement school chosen purposefully. Internal consistency coefficients, unless otherwise specified.

Appendix B

Effect Sizes and Research Questions of Studies Categorically

Category	Study	Research Question	Hedge's *g* (Random Weight-Effect)
Explicit Instruction	Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Appleton, A. C. (2002).	Problem solving tutoring vs. basal instruction only	1.78[a]
	Jitendra, A.K., Griffin, C.C., McGoey, K., Gardill, M.G., Bhat, P., & Riley, T. (1998).	Explicit instruction in diagrammatic representations vs. control (basal curriculum)	0.67[b]
	Kelly, B., Gersten, R., & Carnine, D. (1990).	Instruction incorporating principles of curriculum design vs. control (basal curriculum)	0.88
	Lee, J.W. (1992).	Explicit instruction on using a visual cue vs. control (textbook curriculum)	0.86[c]
	Marzola, E.S. (1987).	Explicit problem solving instruction with verbalizations vs. feedback only (no systematic instruction)	2.01[c]
	Owen, R. L., & Fuchs, L. S. (2002).	Explicit visual strategy instruction vs. control (basal instruction)	1.39
	Ross, P. A., & Braden, J. P. (1991).	Explicit strategy instruction with verbalizations vs. control (typical classroom instruction plus token reinforcement)	0.08[cd]
	Tournaki, H. (1993).	Explicit self-instruction strategy vs. drill and practice	1.74
	Tournaki, H. (2003).	Explicit minimum addend strategy with verbalizations vs. drill and practice	1.61
	Wilson, C. L. & Sindelar, P. T. (1991).	Explicit strategy instructions vs. sequential instruction (simple to more complex problems)	0.91[d]
	Xin, Y. P., Jitendra, A. K., & Deatline-Buchman, A. (2005).	Explicit schema-based strategy instruction vs. general strategy instruction	2.15[d]
Use of Heuristics	Hutchinson, N. L. (1993).	Cognitive strategy instruction vs. control (regular instruction)	1.24[c]
	Van Luit, J. E. H., & Naglieri, J. A. (1999).	MASTER program vs. general instruction program	2.45

Category	Study	Research Question	Hedge's *g* (Random Weight-Effect)
	Woodward, J. (2006).	Strategy instruction plus timed practice drills vs. timed practice drills	0.54[cd]
	Woodward, J., Monroe, K., & Baxter, J. (2001).	Classwide instruction in performance tasks + problem solving instruction in as hoc tutoring vs. regular instruction.	2.00
Student Verbalization of Their Mathematical Reasoning	Hutchinson, N. L. (1993).	Cognitive strategy instruction vs. control (regular instruction)	1.24[c]
	Marzola, E.S. (1987).	Explicit problem solving instruction with verbalizations vs. feedback only (no systematic instruction)	2.01[c]
	Omizo, M. M., Cubberly, W. E., & Cubberly, R. D. (1985).	Modeling by teacher plus student verbalizations vs. modeling by teacher only	1.75
	Pavchinski, P. (1988).	Self-instruction vs. traditional teacher instruction	0.22[cd]
	Ross, P. A., & Braden, J. P. (1991).	Explicit strategy instruction with verbalizations vs. control (typical classroom instruction plus token reinforcement)	0.08[cd]
	Schunk, D. H., & Cox, P. D. (1986).	Continuous student verbalizations vs. no student verbalizations	0.07
	Tournaki, H. (1993).	Self-instruction strategy vs. drill and practice	1.74
	Tournaki, H. (2003).	Explicit minimum addend strategy with verbalizations vs. drill and practice	1.61
Visual Representations: Use by both Teachers and Students	Baker, D. E. (1992).	Strategy with drawing vs. Strategy without drawing	0.31
	Hutchinson, N. L. (1993).	Cognitive strategy instruction vs. control (regular instruction)	1.24[c]
	Jitendra, A.K., Griffin, C.C., McGoey, K., Gardill, M.G., Bhat, P., & Riley, T. (1998).	Explicit instruction in diagrammatic representations vs. control (basal instruction)	0.67[b]
	Lambert, M. A. (1996).	Complex strategy involving visualization vs. control (textbook curriculum strategy)	0.11[c]
	Lee, J.W. (1992).	Explicit instruction on using a visual cue vs. Control (textbook	0.86[c]

Category	Study	Research Question	Hedge's *g* (Random Weight-Effect)
		curriculum)	
	Owen, R. L., & Fuchs, L. S. (2002).	Explicit visual strategy instruction vs. control (basal instruction)	1.39
	Walker, D. W., & Poteet, J. A. (1989/1990).	Diagrammatic representations of problems vs. control (basal key-word strategy)	0.31
Visual Representations: Use by Teachers Only	Kelly, B., Gersten, R., & Carnine, D. (1990).	Instruction incorporating principles of curriculum design vs. control (basal curriculum)	0.88
	Manalo, E., Bunnell, J.K., & Stillman, J.A. (2000). Experiment 1	Strategy instruction plus mnemonics vs. strategy instruction only	-0.01[cd]
	Manalo, E., Bunnell, J.K., & Stillman, J.A. (2000). Experiment 2	Strategy instruction plus mnemonics vs. strategy instruction only	-0.29[cd]
	Witzel, B.S., Mercer, C. D., & Miller, M. D. (2003).	Concrete- representational -abstract sequence of instruction vs. abstract only instruction	0.50[d]
	Woodward, J. (2006).	Strategy instruction plus timed practice drills vs. timed practice drills	0.54[cd]
Sequence and/or Range of Examples	Beirne-Smith, M. (1991).	Sequential presentation of sets of related math facts vs. random presentation of math facts (both within the context of a peer tutoring study)	0.12
	Butler, F. M., Miller, S. P., Crehan, K., Babbitt, B., & Pierce T. (2003).	Concrete-representational-abstract instructional sequence vs. representational-abstract instructional sequence	0.29[c]
	Fuchs, L. S., Fuchs, D., & Prentice, K. (2004).	Transfer training + self-regulation vs. control (regular classroom instruction)	1.14[c]
	Kelly, B., Gersten, R., & Carnine, D. (1990).	Instruction incorporating principles of curriculum design vs. Control (instruction using basal curriculum)	0.88
	Owen, R. L., & Fuchs, L. S. (2002).	Explicit visual strategy instruction vs. Control (basal instruction)	0.26
	Wilson, C. L. & Sindelar, P. T. (1991).	Explicit strategy instruction + sequential instruction (simple to more complex problems) vs. explicit	1.55[d]

Category	Study	Research Question	Hedge's *g* (Random Weight-Effect)
		strategy instruction only	
	Witzel, B.S., Mercer, C. D., & Miller, M. D. (2003).	Concrete-to- representational-to-abstract sequence of instruction vs. abstract only instruction	0.50[d]
	Woodward, J. (2006).	Strategy instruction plus timed practice drills vs. timed practice drills	0.54[cd]
	Xin, Y. P., Jitendra, A. K., & Deatline-Buchman, A. (2005).	Explicit schema-based strategy instruction vs. general strategy	2.15 [d]
Other Instructional and Curricular Variables	Bottge, B. A., Heinrichs, M., Mehta, Z. D., & Hung, Y. (2002).	Anchored instruction vs. Problem solving instruction	0.80[c]
Providing Teachers with Student Performance Data	Allinder, R. M., Bolling, R., Oats, R., & Gagnon, W. A. (2000).	CBM vs. Control (No CBM; basal instruction)	0.27
	Calhoon, M. B., & Fuchs, L. S. (2003).	CBM vs. Control (No CBM; basal instruction)	0.17[c]
	Fuchs, L. S., Fuchs, D., Hamlett, C. L., Phillips, N. B., & Bentz, J. (1994).	CBM vs. Control (No CBM; regular classroom instruction)	0.19
	Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Stecker, P. M. (1990).	CBM vs. Control (No CBM; no systematic performance monitoring)	0.14[c]
	Fuchs, L .S., Fuchs, D., Hamlett, C. L., & Stecker, P. M. (1991).	CBM vs. Control (No CBM-standard monitoring and adjusting teaching)	0.40[c]
	Fuchs, L. S., Fuchs, D., Phillips, N. B., Hamlett, C. L., & Karns, K. (1995).	CBM vs. Control (No CBM)	0.17
	Fuchs, D., Roberts, P. H., Fuchs, L. S., & Bowers, J. (1996).	CBM vs. Control (No CBM)	0.32
Providing Teachers with Student performance Data Plus Options for Addressing Instructional Needs (e.g., instructional recommendations)	Allinder, R. M., Bolling, R.M., Oats, R.G., & Gagnon, W. A. (2000).	CBM with self-monitoring vs. CBM only	0.48
	Fuchs, L. S., Fuchs, D., Hamlett, C. L., Phillips, N. B., & Bentz, J. (1994).	CBM with computer-generated instructional recommendations vs. CBM only	-0.06
	Fuchs, L .S., Fuchs, D., Hamlett, C. L., & Stecker,	Computerized instructional advice vs. CBM only	0.24[c]

Category	Study	Research Question	Hedge's *g* (Random Weight-Effect)
	P. M. (1991).		
Providing Students with Mathematics Performance Feedback	Calhoon, M. B., & Fuchs, L. S. (2003).	CBM + PALS vs. control (basal instruction)	0.17[c]
	Fuchs, L. S., Fuchs, D., Hamlett, C. L., Phillips, N. B., & Bentz, J. (1994).	CBM vs. Control (No CBM; regular classroom instruction)	0.19
	Fuchs, L. S., Fuchs, D., Karns, K., Hamlett, C. L., Katzaroff, M., & Dutka, S. (1997).	CBM vs. control (basal instruction)	-0.17
	Fuchs, L. S., Fuchs, D., Phillips, N. B., Hamlett, C. L., & Karns, K. (1995).	CBM + PALS vs. control (no systematic student performance monitoring)	0.17
	Schunk, D. H., & Cox, P. D. (1986).	Feedback on effort expended vs. No feedback on effort	0.60
	Slavin, R. E., Madden, N. A., & Leavey, M. (1984a).	Working in a cooperative learning group vs. Control (regular instruction)	0.24[c]
	Slavin, R. E., Madden, N. A., & Leavey, M. (1984b).	Working in a cooperative learning group vs. Control (regular instruction)	0.07
Providing Students with Mathematics Performance Feedback and Goal Setting Opportunities	Bahr, C. M. & Rieth, H. J. (1991).	Feedback with goal vs. feedback with no goal	-0.34
	Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Whinnery, K. (1991).	Feedback with goal lines superimposed on graphs vs. feedback with graphs without goal lines	-0.19
	Fuchs, L. S., Fuchs, D., Karns, K., Hamlett, C. L., Katzaroff, M., & Dutka, S. (1997).	Feedback with goal setting vs. feedback only	0.07
	Fuchs, L. S., Fuchs, D., & Prentice, K. (2004).	Transfer training +self-regulation (goal setting) vs. control (regular classroom instruction)	1.14[c]
	Reisz, J. D. (1984).	Feedback with goal setting discussion vs. control (no description)	0.11
Cross-Age Tutoring	Bar-Eli, N., & Raviv, A. (1982).	Cross-age peer tutoring vs. no peer tutoring	1.15
	Beirne-Smith, M. (1991).	Cross-age peer tutoring vs. no peer tutoring	0.75

Category	Study	Research Question	Hedge's *g* (Random Weight-Effect)
Peer-assisted learning within a class	Bahr, C. M. & Rieth, H. J. (1991).	Working in pairs with cooperative goals vs. working individually	0.25
	Calhoon, M. B., & Fuchs, L. S. (2003).	PALS vs. control (basal instruction)	0.17[c]
	Fuchs, L. S., Fuchs, D., Karns, K., Hamlett, C. L., Katzaroff, M., & Dutka, S. (1997).	Peer tutoring vs. control (basal instruction)	-0.17
	Fuchs, L. S., Fuchs, D., Phillips, N. B., Hamlett, C. L., & Karns, K. (1995).	PALS vs. Control (standard procedures)	0.17
	Slavin, R. E., Madden, N. A., & Leavey, M. (1984a).	Working in a cooperative learning group vs. Control (regular instruction)	0.24[c]
	Slavin, R. E., Madden, N. A., & Leavy, M. (1984b).	Working in a cooperative learning group vs. working individually	-0.27

[a]Effect size is based on post-test and near transfer measure. [b]Effect size is based on post-test, short-term retention measure given within a 3-week period, and near-transfer measure. [c]Effect size is based on multiple post-tests. [d]Effect size is based on post-test and short-term retention measure given within a 3-week.