Chapter 6: Report of the Task Group on Instructional Practices

Russell Gersten, Co-Chair
Joan Ferrini-Mundy, Ex Officio, Co-Chair
Camilla Benbow
Douglas H. Clements
Tom Loveless
Vern Williams
Irma Arispe, Ex Officio
Marian Banfield, U.S. Department of Education Staff

CONTENTS

Figures

Tables

Abbreviations

AERA	American Educational Research Association
AGO	Adaptive Instruction and Cooperative Learning/Adaptief Groeps-Onderwijs
ANCOVA	Analysis of Covariance
APA	American Psychological Association
ATB	Active Training with Basals
ATCD	Active Teaching with Empirically Validated Curriculum Design
BASIC	Beginner's All-purpose Symbolic Instruction Code
CAI	Computer-Assisted Instruction
CAT	California Achievement Test
CBI	Computer-Based Instruction
CBL	Computer-Based Laboratories
CGI	Cognitively Guided Instruction
CMI	Computer-managed Instruction
CP	Contextualized Problem
CRA	Concrete-representational-abstract
CTBS	California Test of Basic Skills
CTGV	Cognition and Technology Group at Vanderbilt
EAI	Enhanced Anchored Instruction
EDC	Education Development Center
ERIC	Education Resources Information Center
ETS	Educational Testing Service
FA	Formative Assessment
FIAC	Flanders Interaction Analysis Categories
FT	Project Follow Through
GSI	General Strategy Instruction
ICC	Intra-Class Correlation
IDEA	Individuals with Disabilities Act
ILS	Integrated Learning System
IP	Instructional Practices
ITBS	Iowa Test of Basic Skills
LA	Low Achieving
LD	Learning Disability
MANS	Math Applied to Novel Situations
MASTER	Mathematics Strategy Training for Educational Remediation
MAT	Metropolitan Achievement Test
NAEP	National Assessment of Educational Progress
NALT	Northwest Evaluation Association
NCLB	No Child Left Behind
NCTM	National Council of Teachers of Mathematics
NCME	National Council on Measurement in Education
NICHD	National Institute of Child Health and Human Development
NMP	National Math Panel
NRC	National Research Council

OECD	Organisation for Economic Co-operation and Development
OSEP	Office of Special Education Programs
PA	Performance Assessment
PALS	Peer-Assisted Learning
PDA	Personal Digital Assistant
PIAT	Peabody Individual Achievement Test
PISA	Programme for International Student Assessment
PMI	Peer Mediated Instruction
QED	Quasi-experimental Design
RA	Representative Abstract
RCT	Randomized Control Trials
RME	Realistic Mathematics Education
RPT	Reciprocal Peer Tutoring
SAT	Scholastic Aptitude Test
SAT-M	Scholastic Aptitude Test-Math
SAT-V	Scholastic Aptitude Test-Verbal
SBI	Schema-Broadening Instruction
SBTI	Schema-Based Transfer Instruction
SES	Socioeconomic Status
SESAT	Stanford Early School Achievement Test
SSCI	Social Sciences Citation Index
STAD	Student Teams Achievement Division
STAR	Standardized Testing and Reporting
TAI	Team Assisted Individualization
TEEM	Tucson Early Education Model
USMES	Unified Science and Mathematics for Elementary Schools
WP	Word Problems
WPS	Word Problem Solving
WRAT	Wide Range Achievement Test
WWC	What Works Clearinghouse

Executive Summary

Introduction

Mathematics teaching is an extraordinarily complex activity involving interactions among teachers, students, and the mathematics to be learned in real classrooms (Cohen, Raudenbush, & Ball, 2003). It involves making choices about material and tools to use, planning ways to group and interact with students of differing backgrounds and with differing interests and motivation. It is within this set of areas that some of today's most pressing and debated questions about mathematics instruction are situated.

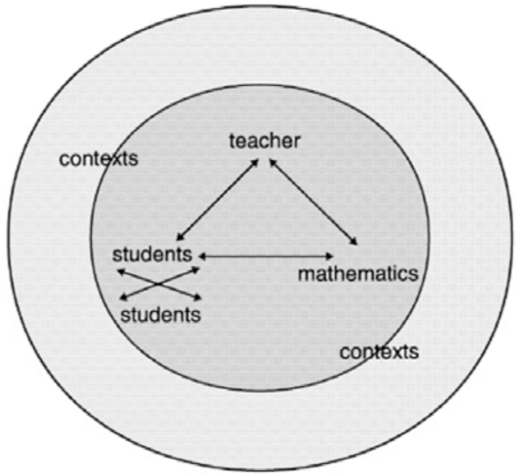
The Instructional Practices (IP) Task Group needed to consider the challenges that this complexity creates while determining what might be learned from research studies on the teaching of mathematics. Not all of the questions that teachers, policymakers, and the public wish to have answered are easily studied or lend themselves to experimental and quasi-experimental research, types of research from which generalizations to practice or for policy can be made. Moreover, many important questions that could be studied using these methods, unfortunately, have not been addressed in these ways. This limits what can validly be said about possible effective practices for the teaching of mathematics. The Task Group's undertaking was to marshal the scientific evidence to make policy recommendations and, thus, only experimental and quasi-experimental studies could be examined.

This situation is hardly unique to mathematics education, or educational research in general. It is—and has been—true in the development of scientific research in any field from engineering to economics to clinical psychology to public health. The accumulation of findings is slow at first, with the expensive experimental designs employed only after a certain amount of knowledge has emerged. Research on teaching and learning is a relatively young field.

With these caveats in mind, the overarching question the Task Group approached is: *What instructional practices enable students to learn mathematics most successfully?* Fortunately, while the knowledge base is not uniformly deep, there has been some progress at assembling evidence about questions of causal impact that has implications for practice and for policy within specific areas of mathematics instructional practice.

Therefore, within this general question, the Task Group identified six questions for investigation, addressing topics that were deemed important by the field, often including issues that have been hotly debated. Questions were identified within all three of the types of interactions comprising teaching as indicated in Figure 1; the Task Group recognizes that most of the questions here engage all three types of interactions specified in the figure, but have classified them according to the types of interactions that seem most salient.

Figure ES-1: Instructional Triangle



Source: *Adding It Up*, National Research Council, 2001, p. 314.

The Task Group realizes that by no means is the list of questions discussed below a comprehensive list of questions about each of these three types of interactions; indeed, it only begins to scratch the surface about what might be learned to inform mathematics teaching practice through research. The Task Group was aware that there are many widely used instructional practices that might have been examined here but that were not included because of limitations of time, resources, and available research. Nonetheless, it is a list of specific issues that will allow the Task Group to draw some conclusions from a small set of rigorous research studies, thereby setting the foundation for a far more expansive program of rigorous research that would fill the gaps in the research on these issues and also take up the many other issues that practitioners face in improving mathematics teaching and learning.

The methodology used in the Instructional Practices Task Group research review process, including an account of how the topics were selected, and the criteria for standards of evidence, are discussed in the full report introduction and in Appendix A.

Interactions Between Teachers and Students

Most contemporary perspectives on instruction argue that finding the best form for those interactions is a complex problem that is dependent on teachers' backgrounds, students' characteristics, school culture, the mathematical topics being addressed, and the instructional materials being used. One advantage of rigorous experimental research is that, over time, the professional community can discern which practices tend to be effective across a broad array of teacher and learning characteristics and a broad array of mathematical topics. One major goal of the Task Group's effort is to critically review the research literature for the small body of rigorous experimental studies and to discern patterns of findings that suggest specific means for improving instructional practice.

It is agreed that there is no single, ideal form in which students and teachers should consistently interact. Nonetheless, there are certain "positions" taken by various organizations and individuals arguing in favor, or in opposition to, such practices as direct instruction, cognitive-strategy instruction, student-centered approaches, cooperative learning, discovery learning, guided inquiry, situated cognition approaches, collaborative learning, and lecture-recitation.

A less polarizing issue, but one that is of great importance to classroom teachers of mathematics, is the challenge of how to best interact with low-achieving students and specifically with students having learning disabilities. A major challenge of mathematics teaching for teachers is to find the combination of instructional approaches and materials that will best meet the needs of the diversity of students in their classrooms.

Research was examined that addresses two basic questions about the forms of teacher and student interactions.

How Effective Is Teacher-Directed Instruction in Mathematics in Comparison to Student-Centered Approaches, Including Cooperative and Collaborative Groups, in Promoting Student Learning?

A controversial issue in the field of mathematics teaching and learning is whether classroom instruction should be more teacher directed or student centered. These terms have come to incorporate a wide array of meanings, with teacher directed ranging from highly scripted direct instruction approaches to interactive lecture styles, and with student centered ranging from students having primary responsibility for their own mathematics learning to highly structured cooperative groups. Schools and districts must make choices about curricular materials or instructional approaches that often seem more aligned with one instructional orientation than another. This leaves teachers wondering about when to organize their instruction one way or the other, whether certain topics are taught more effectively with one approach or another, and whether certain students benefit more from one approach than the other. The review was limited to studies that directly compared these two positions. The studies in the review compare an instructional regime in which teachers do more teaching (and therefore students less) with one in which students do more teaching and teachers less.

Only eight studies were found that met the Task Group's standards for quality that were consistent with this definition. The studies presented a mixed and inconclusive picture of the relative impact of these two forms of instruction. High-quality research does not support the contention that instruction should be either entirely "child centered" or "teacher directed." Research indicates that some forms of particular instructional practices can have a positive impact under specified conditions. All-encompassing recommendations that instruction should be entirely "child centered" or "teacher directed" are not supported by research. The limited research base of rigorous research does not support the exclusive use of either approach.

One of the major shifts in education over the past 25–30 years has been advocacy for the increased use of cooperative learning groups and peer-to-peer learning (e.g., structured activities for students working in pairs) in the teaching and learning of mathematics.

Cooperative learning is used for multiple purposes: for tutoring and remediation, as an occasional substitute for independent seatwork, for intricate extension or collaborative groups has been advocated in various mathematics education reports, policies, and state curricular frameworks and instructional guidelines.

Research has been conducted on a variety of cooperative learning approaches. One such approach, Team Assisted Individualization (TAI) has been shown to significantly improve students' computation skills. This instructional approach involves heterogeneous groups of students helping each other, individualized problems based on student performance on a diagnostic test, and rewards based on both group and individual performance. Effects on conceptual understanding and problem solving were not significant. There is evidence suggesting that working in dyads with a clear structure also improves computation skills in the elementary grades. However, additional research is needed.

What Instructional Strategies for Teaching Mathematics to Students With Learning Disabilities and to Low-Achieving Students Show the Most Promise?

A major challenge of mathematics teaching for teachers is to find the combination of instructional approaches and materials that will best meet the needs of the diversity of students in their classrooms. The Task Group chose to examine research that specifically looks at issues addressing students who bring a range of diversity to mathematics classrooms—those students with learning disabilities and those students who struggle with learning mathematics but who do not have a mathematics learning disability.

Obviously this topic has been of high interest for special educators, but increasingly, surveys of teachers have indicated that, as increasing numbers of students with learning disabilities (LD) receive their mathematics instruction in their regular classroom, strategies for teaching these students have become a high priority for all educators. Fortunately, there is an appreciable body of research on this topic that meets the standards for rigorous scientific research established by this Task Group.

A review of 26 high-quality studies, mostly using randomized control designs, was conducted. These studies provide a great deal of guidance concerning some defining features of effective instructional approaches for students with LD as well as low-achieving (LA) students.

Explicit systematic instruction typically entails teachers explaining and demonstrating specific strategies, and allowing students many opportunities to ask and answer questions and to think aloud about the decisions they make while solving problems. It also entails careful sequencing of problems by the teacher or through instructional materials to highlight critical features. More recent forms of explicit systematic instruction have been developed with applications for these students. These developments reflect the infusion of research findings from cognitive psychology, with particular emphasis on automaticity and enhanced problem representation.

This analysis of the body of research indicated that explicit methods of instruction are consistently and significantly effective with students with learning disabilities in the performance of computations, solving word problems, and solving problems that require the application of mathematics to novel situations.

Only a small number of studies were located that investigated the use of visual representations or student "think alouds." Therefore, no inferences about their effectiveness can be drawn. The research suggests that they are most useful when they are integrated with explicit instruction.

Based on this admittedly small body of research, the Task Group concludes that students with learning disabilities and other students with learning problems should receive some time on a regular basis with some explicit systematic instruction. There is no reason to believe that this type of instruction should comprise all the mathematics instruction these students receive. However, it does seem essential for building proficiency in both computation and the translation of word problems into appropriate mathematical equations and solutions. Some of this time should be dedicated to ensuring that students possess the foundational skills and conceptual knowledge necessary for understanding the mathematics they are learning at their grade level.

Interactions Between Students and the Mathematics They Are Learning

In discussions about effective mathematics instruction, there are multiple questions about the ways the curriculum, instructional materials, and resources for mathematics learning influence student performance in mathematics. The Task Group chose to focus the research review on three controversial areas of this domain: a curricular issue concerning how the mathematics is presented; an issue about the impact of tools as a means of interacting with the mathematics; and a curricular organization issue about the pace and nature of the mathematics for gifted students.

Do 'Real-World' Problem Approaches to Mathematics Teaching, and Efforts to Ensure that Students Can Solve 'Real-World' Problems, Lead to Better Mathematics Performance Than Other Approaches?

The importance of addressing this topic as an especially controversial "hot button" issue in the field was stressed, both by Task Group members, as well as by members of the public in testimony to the Panel. Many textbooks begin each unit with "real-world" problems and consider this a potentially motivating approach. Some instructional materials use "real-world" contexts as a means of introducing mathematical ideas. State and national standards typically include as goals students' ability to apply mathematics to situations that occur in a child's life, or that might occur in future jobs. Consequently, high-stakes assessments such as the National Assessment of Educational Progress (NAEP) and many state tests include "real-world" problems. There are strong perspectives both in support of, and in opposition to, the use of "real-world" problems as a means for students to interact with the mathematics they are to learn. For these reasons, a serious examination of the research on this topic seemed warranted.

The research review focused on two key issues. The first was the extent to which problems that authors call “real-world” problems do, in fact, pique students’ interest and engage them more fully in exploration of mathematical concepts, with a goal of learning mathematics. A related issue is the extent to which use of “real-world” problems in instruction increases students’ ability to transfer the mathematical knowledge they possess to novel situations. Unfortunately, there is no agreed upon definition of “real-world” problems; the terminology is used in very different ways by researchers, teachers, mathematicians, and mathematics educators. And, the matter that what is a “real-world” problem to one student may not be a “real-world” problem to another is an issue. Conducting research in this area is complex; fidelity of the teachers’ implementation of the instructional materials or instructional strategy is difficult to assess. Although not addressed in the studies we examined, teachers’ knowledge and capacity to use such problems effectively varies greatly. Given these caveats, the Task Group addressed the question of whether using “real-world” contexts to introduce and teach mathematical topics and procedures is preferable to more typical instructional approaches.

The body of high-quality studies for this topic is small. Five studies addressed the question of whether the use of “real-world” problems as the instructional approach led to improved performance on outcome measures of ability to solve “real-world” problems, as well as on more traditional assessments. Four of these were similar enough to combine in a meta-analysis. The meta-analysis revealed that if mathematical ideas are taught using “real-world” contexts, then students’ performance on assessments involving similar problems is improved. However, performance on assessments of other aspects of mathematics learning, such as computation, simple word problems, and equation solving, is not improved.

For certain populations (upper elementary and middle grade students and remedial ninth-graders) and for specific domains of mathematics (fraction computation, basic equation solving, and function representation), instruction that features the use of “real-world” contexts can have a positive impact on certain types of problem solving. Additional research is needed to explore the use of “real-world” problems in other mathematical domains, at other grade levels, and with varied definitions of “real-world” problems.

What Is the Relative Impact on Mathematics Learning When Students Use Technology Compared to Instruction That Does Not Use Technology?

There are several types of educational technology that provide opportunities for students to interact with mathematics. The review includes focus on computer software, calculators, and graphing calculators.

Among the many categories of technology, calculators, including graphing calculators, have generated the greatest amount of debate. Some have championed their use in developing problem-solving abilities, by allowing students to perform far more, and more complex, arithmetic operations than would have been possible without technology. Others believe that calculators may reinforce independent skill mastery, or even that they should, along with mental arithmetic, replace some of the paper-and-pencil calculations that dominate elementary school mathematics. On the other hand, some have bemoaned their misuse. One concern is that calculators may have an insidious effect on paper-and-pencil arithmetic and algebraic skills. Some are concerned that reliance on calculators can preclude the development of proficiency with standard calculation algorithms and thus deprive students of an understanding and appreciation of the mathematics that underlies the standard algorithms, as well as ability to quickly retrieve basic arithmetic facts.

A review of 11 studies that met the Task Group's rigorous criteria (only one study was less than 20 years old) found limited to no impact of calculators on calculation skills, problem solving, or conceptual development over periods of up to one year. Unfortunately, these studies cannot be used to judge the advantages or disadvantages of multiyear calculator use beginning in the early years because such long-term use has not been adequately investigated. The Task Group cautions that to the degree that calculators impede the development of automaticity, fluency in computation will be adversely affected.

The Task Group found that computer-assisted instruction (CAI) drill and practice, if of high quality, can improve students' performance compared to conventional instruction. Drill and practice programs **can be** useful tools in developing students' automaticity, or fast, accurate, and effortless performance on computation, freeing working memory so that attention can be directed to the more complicated aspects of complex tasks.

Research has demonstrated that tutorials (CAI programs, often combined with drill and practice) that are well designed and implemented can have a positive impact on mathematics performance. CAI tutorials have been used effectively to introduce and teach new subject-matter content. However, these studies also suggest several important caveats. Care must be taken that there is evidence that the software to be used has been shown to increase learning in the specific domain and with students who are similar to those who are under consideration. Educators should critically inspect individual software packages and studies that evaluate them critically. Furthermore, support conditions to use the software effectively (sufficient hardware and software; technical support; adequate professional development, planning, and curriculum integration), should be in place, especially in large-scale implementations, to achieve optimal results.

Research indicates that computer programming improves students' performance compared to conventional instruction on both mathematics achievement in general and on problem solving. However, computer programming by students can be employed in a wide variety of situations using distinct pedagogies, not all of which may be effective (e.g., integration into the mathematics curriculum may be required for substantial effects). Therefore, the findings are limited to the careful, targeted application of computer programming for learning used in the studies reviewed.

What Instructional Arrangements for Engaging with Mathematics Are Most Promising for Mathematically Gifted Students?

Zimmer, Christina, Hamilton, and Weber Prine (2006) noted that, in a recent survey of teachers implementing the No Child Left Behind Act (NCLB), over half the teachers surveyed felt that implementation of the law resulted in improved learning opportunities for low-performing students, but that teachers and administrators at all levels of schooling worried about high-achieving students receiving adequate instructional challenge in all curricular areas. This review of the research literature explored the immediate and delayed impacts of gifted education approaches aimed at accelerating students' mathematics instruction (e.g., by covering two, or even four years of high school mathematics in 15 months) and those that attempt to provide enrichment or extension activities for mathematically precocious students. This question is addressed in the category of student-mathematics interactions because it is very much about the pace and structure for engaging gifted students with mathematics content.

The Task Group's review of the literature about the kind of mathematics instruction would be most effective for gifted students focused on the impact of programs involving acceleration, enrichment, and the use of homogeneous grouping. The extensive literature searches we conducted yielded few studies that met the Task Group's methodologically rigorous criteria for inclusion. Thus for this topic—and this topic only—we relaxed these criteria in order to fulfill our charge of evaluating the "best available scientific evidence." One randomized control trial study and seven quasi-experimental studies were located. All but one of these studies have limitations.

Despite the flaws in any one study, the set of studies suggests there is value to differentiating the mathematics curriculum for students who are gifted in mathematics and possess sufficient motivation, especially when acceleration is a component (i.e., pace and level of instruction are adjusted). A small number of studies suggest that individualized instruction, where the pace of learning is increased and often managed via computer instruction, produces gains in learning.

Gifted students who are accelerated by other means not only gained time and reached educational milestones earlier (e.g., college entrance) but appear to achieve at levels at least comparable to those of their equally able same-age peers on a variety of indicators even though they were younger when demonstrating their performance on the various achievement benchmarks. One study suggests that gifted students also appear to become more strongly engaged in science, technology, engineering, or mathematical areas of study.

Some support also was found for supplemental enrichment programs. Of the two programs analyzed, one explicitly utilized acceleration as a program component and the other did not. This supports the view in the field of gifted education that acceleration and enrichment combined should be the intervention of choice. We believe it is important for school policies to support appropriately challenging work in mathematics for gifted and talented students.

Interactions Between Teachers and Mathematics

Teachers engage with the mathematical content that they teach in various aspects of teaching practice: in planning and designing lessons, in interpreting and responding to student questions, and in the work of assessing their students' mathematical knowledge. Fortunately, formative assessment is an area of great contemporary interest and is also an area with a rich set of rigorous experimental field studies.

What Is the Impact of Use of Formative Assessment in Mathematics Teaching?

Educators at all levels realize the importance of assessing their students' progress during the year. Formative assessment—the ongoing monitoring of student learning to inform instruction—is generally considered a hallmark of effective instruction in any discipline. Interest in formative assessment has dramatically increased since No Child Left Behind required states to establish accountability systems. Teachers' interpretation and use of the data available to them from instructionally embedded, in-class assessments in the context of teaching, along with high-stakes assessments are critical for improving outcomes for all students. However, many different systems have been established and touted for use as formative assessments. These range from the end-of-unit and mastery tests that accompany major commercial textbook series, to more contingent and informal probes of students' understandings to be used while they solve problems, to weekly tests that sample from the year's instructional objectives in mathematics. The Task Group examined rigorous experimental studies of the impact of teachers' use of formative assessment on students' growth in mathematics proficiency. The Task Group's review of the high-quality studies of this topic produced several conclusions.

Teachers' regular use of formative assessment is marginally significant in improving their students' learning. This is especially true if teachers have additional guidance on using the assessment to design and individualize instruction.

Although the research base is smaller, and less consistent than that on the general effectiveness of formative assessment, the research suggests that several specific tools and strategies can help teachers use formative assessment information more effectively. The first promising strategy is providing formative assessment information to teachers (via technology) on content and concepts that require additional work with the whole class. The second promising strategy involves using technology to specify activities needed by individual students. Both of these aids can be implemented via tutoring, computer-assisted instruction, or help provided by a professional (teacher, mathematics specialist, trained paraprofessional).

The Task Group cautions that only one type of formative assessment has been studied with rigorous experimentation. These are assessments that include random sampling of items that address state standards. These assessments tend to take between 2 and 8 minutes to administer and thus are practical for regular use.

The regular use of formative assessment particularly for students in the elementary grades is recommended. These assessments need to provide information not only on their content validity but also on their reliability and their criterion-related validity (i.e., correlation of these measures with other measures of mathematics proficiency). For struggling students, frequent (e.g., weekly or biweekly) use of these assessments appears optimal, so that instruction can be adapted based on student progress.

Research is needed regarding the content and criterion-related validity and reliability of other types of formative assessments (such as unit mastery tests included with many published mathematics programs, performance assessments, and dynamic assessments involving "think alouds"). This research should include studies of consequential validity (i.e., the impact they have on helping teachers improve the effectiveness of their instruction).

Use of formative assessments in mathematics can lead to increased precision in how instructional time is used in class and can assist teachers in identifying specific instructional needs. Formative measures provide guidance as to the specific topics needed for assistance. Formative assessment should be an integral component of instructional practice in mathematics.

Conclusion

Mathematics instruction is a complex professional practice. The educational research community has made important forays into several of the most controversial and pressing questions about the effectiveness and impact of various types of instructional practice, and in particular have conducted some studies that examine the effects of various interpretations and implementations of practices that have been advocated in the "reform" documents in mathematics education over the past two decades.

The question asked by the Task Group is: *What can be learned from a review of the best available evidence in six important aspects of practice?* These practices included: the use of "real-world" problems in mathematics teaching, the use of technology, the enrichment and acceleration of instruction for mathematically precocious students, the use of cooperative groups and peer instruction, the use of direct instruction with learning disabled students, and the use of formative assessment.

For none of the areas examined did the Task Group find sufficiently strong and comprehensive bodies of research to support all-inclusive policy recommendations of any of the practices addressed. Nor did the Task Group find sufficient evidence to support policy recommendations favoring the status quo in mathematics teaching.

Across all of the areas, the Task Group found that **several instructional practices in mathematics teaching show some promise, in comparison to typical practice, for affecting student learning.** In each case the "promising" practice is clearly specified, somewhat prescriptive, and involves a mix, or combination, of particular distinct practices. Thus, for example, it cannot be said that cooperative learning is a practice whose effectiveness is supported by research—but the Team Assisted Individualization (TAI) approach, with particular students in a particular area of mathematics does appear to be effective. Although formative assessment to inform instruction is useful, it is enhanced when teachers use assessment tools with known validity and reliability. For students performing in the lower third of grade level expectations, explicit instruction using clear models of proficient performance, many opportunities to verbalize their problem-solving strategies, and adequate practice and review should be a part of the mathematics program. It is not surprising that what the Task Group found about effective instructional practice is far more subtle and nuanced than direct answers to the starkly stated questions investigated.

The Task Group found some rather robust findings, but these findings must be accompanied by a caveat. When a practice is demonstrated by high-quality experimental research to have some promise, it is critical to be clear about the promise "for what aspects of mathematics proficiency." Different practices and approaches impact different kinds of outcomes, ranging from computational performance, to "real-world" problem solving, to identifying extraneous problem information, to long-term participation and interest in studying mathematics.

Because researchers and practitioners use different definitions to describe their interventions, it is conceptually problematic to place too much stock in generalizing that a broad category of practice (e.g., using technology or using "real-world" problems) has impact because a set of studies working on the same particular component of this category has impact, which was the case in some of the Task Group's reviews.

The Task Group's process included asking mathematicians and mathematics education reviewers to examine the mathematical content of the research studies—to look at the assessments and interventions, to the extent possible, based on the published reports. They expressed important concerns, including the possibility that an outcome measure item purported to measure computation might not do so because it really measures ability to use the context, for instance. They expressed concern that some topics were underdeveloped (i.e., failed to help students access the underlying mathematics in the topic covered), or that items were mislabeled (e.g., as "problem solving") when the mathematics expert might classify them otherwise. However, they also did note that several of the studies reviewed seemed to help students increase their knowledge of mathematics and how to apply that knowledge to novel situations in a way that is valid from a mathematical perspective.

Seeing how few robust findings emanated from a review of the rigorous research on the topics addressed, it is clear that most practitioners would like more guidance for several areas of instruction. Yet even the inconclusive and limited findings can provide a real service to the profession. If an administrator, a developer or a parent comments, "Research says that lessons must start with 'real-world' problems," or "Students will really learn mathematics only if they are taught using direct instruction," consumers and professionals now know that research is inconclusive on these topics.

This is a necessary step in the evolution of educational research into a more mature science. The paucity of findings and the paucity of high-quality experimental research in the field led the Task Group to realize, early on in the process, that few definitive answers to the research questions posed would be found.

What Would the Instructional Practices Task Group Say to the Practitioner?

There is no one ideal approach to teaching mathematics; the students, the mathematical goals, the teacher's background and strengths, and the instructional context, all matter. The findings here do suggest that it is especially important to:

- monitor what students understand and are able to do mathematically;
- design instruction that responds to students' strengths and weaknesses based on research when it is available; and
- employ instructional approaches and tools that are best suited to the mathematical goals, recognizing that a deliberate and conscious mix of strategies will be needed.

Also, it is important for teachers, school administrators, and the public to understand the importance of helping to formulate research questions and being willing to participate in the types of experimental and quasi-experimental studies that are described here.

What Would the Instructional Practices Task Group Say to the Researcher?

More research that can identify causal claims is needed to guide both policy and practice. Building the mathematics education research portfolio to include this work will involve:

- Formulation of research questions that are of interest to practitioners and policy-makers;
- Collaborations among mathematicians, mathematics education researchers, methodologists, and psychometricians; and
- Motivation to design and undertake rigorous studies.

The work of this Task Group has substantiated understanding of the complexity and challenge of effective mathematics instruction. It is now up to practitioners, policymakers, mathematicians, and mathematics education researchers to take up the challenges of clarifying the definitions of mathematics instructional practices, debunking myths about mathematics instruction, and formulating the types of research studies that can answer the pressing questions that need to be addressed.

In conclusion, instructional practice should be informed by high-quality research, when available, and by the best professional judgment and experience of accomplished classroom teachers.

I. Introduction

A. Instructional Practices

Mathematics teaching is an extraordinarily complex activity involving interactions among teachers, students, and the mathematics to be learned in real classrooms (Cohen, Raudenbush, & Ball, 2003). It involves making choices about material and tools to use, planning ways to group and interact with students of differing backgrounds and with differing interests and motivation. It is within this set of areas that some of today's most pressing and debated questions about mathematics instruction are situated.

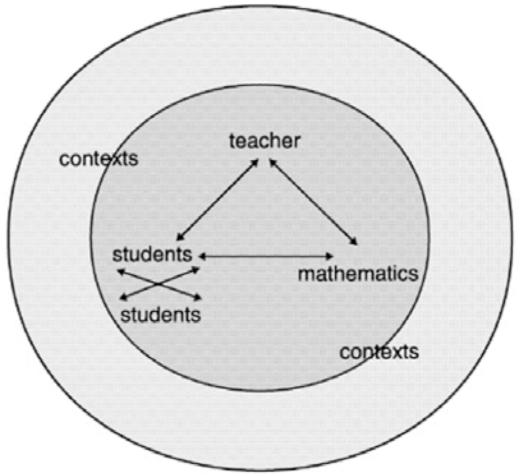
The Instructional Practices Task Group needed to consider the challenges that this complexity creates while determining what might be learned from research studies on the teaching of mathematics. Not all of the questions that teachers, policymakers, and the public wish to have answered are easily studied or lend themselves to experimental and quasi-experimental research, types of research from which generalizations to practice or for policy can be made. Moreover, many important questions that could be studied using these methods, unfortunately, have not been addressed in these ways. This limits what can validly be said about possible effective practices for the teaching of mathematics. The Task Group's undertaking was to marshal the scientific evidence to make policy recommendations and, thus, only experimental and quasi-experimental studies were examined.

This situation is hardly unique to mathematics education or educational research in general. It is—and has been—true in the development of scientific research in any field from engineering to economics to clinical psychology to public health. The accumulation of findings is slow at first, with the expensive experimental designs employed only after a certain amount of knowledge has emerged. Research on teaching and learning is a relatively young field.

With these caveats in mind, the overarching question the Task Group approached is: *What instructional practices enable students to learn mathematics most successfully?* Fortunately, while the knowledge base is not uniformly deep, there has been some progress at assembling evidence about questions of causal impact that has implications for practice and for policy within specific areas of mathematics instructional practice.

Therefore, within this general question, the Task Group identified six questions for investigation, addressing topics that were deemed important by the field often including issues that have been hotly debated. Questions were identified within all three of the types of interactions comprising teaching as indicated in Figure 1; the Task Group recognizes that most of its questions here engage all three types of interactions specified in the figure, but have classified them according to the types of interactions that seem most salient.

Figure 1: Instructional Triangle



Source: *Adding It Up,* National Research Council, 2001, p. 314.

The Task Group realizes that by no means is the list of questions discussed below a comprehensive list of questions about each of these three types of interactions; indeed, it only begins to scratch the surface about what might be learned to inform mathematics teaching practice through research. The Task Group was aware that there are many widely used instructional practices that might have been examined here but that were not included because of limitations of time, resources, and available research. Nonetheless, it is a list of specific issues that will allow the Task Group to draw some conclusions from a small set of rigorous research studies, thereby setting the foundation for a far more expansive program of rigorous research that would fill the gaps in the research on these issues and also take up the many other issues that practitioners face in improving mathematics teaching and learning.

1. Notes About Methodology and Reporting

The methodology used in the Instructional Practices Task Group research review process, including an account of how the topics were selected, and the criteria for standards of evidence, are included in Appendix A. For ease in reading this report key points are summarized here. The studies used in the meta-analyses and syntheses that follow were designated as either Category 1 or 2. Category 1 studies are experimental and quasi-experimental studies that meet or meet with reservations the What Works Clearinghouse (WWC) standards. Studies in this category provide evidence of causal claims and include randomized control trials (RCTs) and strong quasi-experimental studies. Some exceptions to the WWC criteria were allowed; these are described in Appendix A. Category 2 consisted of weak group comparison studies (e.g., failed RCTs and weak nonequivalent comparison designs; other flaws discussed in Appendix A). Category 2 studies are always open to multiple interpretations with regard to causal inferences; however, they are not necessarily weak studies for other purposes such as description. If there were no acceptable experimental studies, sections of the report may include brief discussion of Category 2 studies. If there is a pattern of findings across the studies this may also be mentioned. Panelists were free to use any type of research (descriptive, correlational, qualitative) to set the context for the meta-analyses.

For all studies that met the criteria for inclusion, the What Works Clearinghouse guidelines were used to calculate standardized mean differences in mathematics achievement. Hedges' g standardized mean differences were calculated for each of the studies. In cases in which schools, teachers, or classrooms were assigned (either randomly or nonrandomly) into intervention and comparison groups and the unit of assignment was not the same as the unit of analysis, the effect size and accompanying standard error were adjusted for clustering within schools, teachers, or classrooms. When judged appropriate, effect sizes were pooled across studies meta-analytically using random effects models. Specifically, weighted mean effect sizes were computed using inverse variance weights to reflect the statistical precision of the respective studies stemming from both the subject-level and study-level sampling error.

Multiple contrasts: For each study that included at least three conditions, effect sizes were calculated for all relevant contrasts, provided that they were orthogonal. When pooling the effects using meta-analytic techniques, only independent effect sizes per study were included, i.e., those not based on the same participant samples.

Multiple outcomes: For studies that reported effects on more than one mathematics achievement outcome, either one outcome was chosen, or the results from multiple outcomes were averaged, with decisions made by the authors on a case-by-case basis. Assessments that were overly aligned with an intervention were either not used or noted when used.

Multiple independent samples within a study: In cases in which impacts on independent samples within a study were reported, all independent effect sizes were included separately in the pooled analysis.

Throughout this report, effect sizes are reported as statistically significant only when $p < .01$. Effect sizes where $p < .10$ are described as "bordering on significance". This report conforms with the National Math Advisory Panel (Panel) *Guidelines for Standards of Evidence* in using the following terminology: strong evidence, moderately strong evidence, suggestive evidence, inconsistent evidence, and weak evidence.

2. Interactions Between Teachers and Students

Most contemporary perspectives on instruction argue that finding the best form for those interactions is a complex problem that is dependent on teachers' backgrounds, students' characteristics, school culture, the mathematical topics being addressed, and the instructional materials being used. One advantage of rigorous experimental research is that, over time, the professional community can discern which practices tend to be effective across a broad array of teacher and learning characteristics and a broad array of mathematical topics. One major goal of the Task Group's effort was to critically review the research literature for the small body of rigorous experimental studies and to discern patterns of findings that suggest specific means for improving instructional practice.

It is agreed that there is no single, ideal form in which students and teachers should consistently interact. Nonetheless, there are certain "positions" taken by various organizations and individuals arguing in favor of, or in opposition to, such practices as direct instruction, cognitive-strategy instruction, student-centered approaches, cooperative learning, discovery learning, guided inquiry, situated cognition approaches, collaborative learning, and lecture-recitation.

A less polarizing issue, but one that is of great importance to classroom teachers of mathematics, is the challenge of how to best interact with low-achieving students and specifically with students having learning disabilities. A major challenge of mathematics teaching for teachers is to find the combination of instructional approaches and materials that will best meet the needs of the diversity of students in their classrooms. Research was examined that addresses two basic questions about the forms of teacher and student interactions.

a. How Effective Is Teacher-Directed Instruction in Mathematics in Comparison to Student-Centered Approaches, Including Cooperative and Collaborative Groups, in Promoting Student Learning?

A controversial issue in the field of mathematics teaching and learning is whether classroom instruction should be more teacher-directed or student-centered. These terms have come to incorporate a wide array of meanings, with teacher-directed ranging from highly scripted direct instruction approaches to interactive lecture styles, and with student-centered ranging from students having primary responsibility for their own mathematics learning to highly structured cooperative groups. Schools and districts must make choices about curricular materials or instructional approaches that often seem more aligned with one instructional orientation than another. This leaves teachers wondering about when to organize their instruction one way or the other, whether certain topics are taught more effectively with one approach or another, and whether certain students benefit more from one approach than the other. The review was limited to studies that directly compared these two positions. The studies in the review compare an instructional regime in which teachers do more teaching (and therefore students less) with one in which students do more teaching and teachers less.

One of the major shifts in education over the past 25–30 years has been advocacy for the increased use of cooperative learning groups and peer-to-peer learning (e.g., structured activities for students working in pairs) in the teaching and learning of mathematics. Cooperative learning is used for multiple purposes: for tutoring and remediation, as an occasional substitute for independent seatwork, for intricate extension activities, for initial brainstorming and for numerous other purposes. Use of cooperative or collaborative groups has been advocated in various mathematics education reports, policies, and state curricular frameworks and instructional guidelines.

Provided in a subsequent section of the report is a synthesis of the research that met Task Group criteria on the topic of teacher-directed vs. student-centered learning. The section includes a review of studies that compare general versions of teacher-directed and student-centered mathematics instruction in accordance with the Task Group's definition. There are only a limited number of sufficiently rigorous research studies making this comparison, within this definition. There is also a review of studies that examine various forms of cooperative and collaborative groups, including such specific approaches as Team Assisted Instruction and Peer Assisted Learning, as well as the use of cooperative groups with technology, and other approaches.

b. What Instructional Strategies for Teaching Mathematics to Students with Learning Disabilities and to Low-Achieving Students Show the Most Promise?

A major challenge of mathematics teaching for teachers is to find the combination of instructional approaches and materials that will best meet the needs of the diversity of students in their classrooms. The Task Group chose to examine research that specifically looks at issues addressing students who bring a range of diversity to mathematics classrooms—those students with learning disabilities (LD) and those students who struggle with learning mathematics but who do not have a mathematics learning disability.

Obviously this topic has been of high interest for special educators, but increasingly, surveys of teachers have indicated that, as increasing numbers of students with LD receive their mathematics instruction in their regular classroom, strategies for teaching these students has become a high priority for all educators. Fortunately, there is an appreciable body of research on this topic that meets the standards for rigorous scientific research established by this Task Group.

3. Interactions Between Students and the Mathematics They Are Learning

In discussions about effective mathematics instruction, there are multiple questions about the ways the curriculum, instructional materials, and resources for mathematics learning influence student performance in mathematics. The Task Group chose to focus the research review on three controversial areas of this domain: a curricular issue concerning how the mathematics is presented; an issue about the impact of tools as a means of interacting with the mathematics; and a curricular organization issue about the pace and nature of the mathematics for gifted students.

a. Do 'Real-World' Problem Approaches to Mathematics Teaching and Efforts to Ensure That Students Can Solve 'Real-World' Problems Lead to Better Mathematics Performance Than Other Approaches?

The importance of addressing this topic as an especially controversial "hot button" issue in the field was stressed, in particular, by Task Group members, as well as by members of the public testifying before the Panel. Many textbooks begin each unit with "real-world" problems and consider this a potentially motivating approach. Some instructional materials use "real-world" problems as a means of introducing mathematical ideas. State and national standards typically include as goals students' ability to apply mathematics to situations that occur in a child's life or that might occur in future jobs. Consequently high-stakes assessments such as the National Assessment of Educational Progress (NAEP) and many state tests include "real-world" problems. There are strong perspectives both in support of, and in opposition to, the use of "real-world" problems as a means for students to interact with the mathematics they are to learn. For these reasons, a serious examination of the research on this topic seemed warranted.

The research review focused on two key issues. The first was the extent to which problems that authors call "real-world" problems do, in fact, pique students' interest and engage them more fully in exploring mathematical concepts with a goal of learning mathematics. A related issue is the extent to which "real-world" problems increase students' ability to transfer the mathematical knowledge they possess to novel situations. Unfortunately, there is no agreed upon definition of "real-world" problems; the terminology is used in very different ways by researchers, teachers, mathematicians, and mathematics educators.

b. What Is the Relative Impact on Mathematics Learning When Students Use Technology Compared to Instruction that Does Not Use Technology?

There are several types of educational technology that provide opportunities for students to interact with mathematics. The review includes focus on computer software and calculators, including graphing calculators.

Among the many categories of technology, calculators, including graphing calculators, have generated the greatest amount of debate. Some have championed their use in developing problem-solving ability by allowing students to perform far more, and more complex, arithmetic operations than would have been possible without technology. Others believe that calculators may reinforce independent skill mastery, or even that they should, along with mental arithmetic, replace some of the paper-and-pencil calculations that dominate elementary school mathematics. On the other hand, some have bemoaned their misuse. One concern is that calculators may have an insidious effect on paper-and-pencil arithmetic and algebraic skills. Some are concerned that reliance on calculators can preclude the development of proficiency with standard calculation algorithms and thus deprive students of an understanding and appreciation of the mathematics that underlies the standard algorithms, as well as ability to quickly retrieve basic arithmetic facts.

c. What Instructional Arrangements for Engaging with Mathematics Are Most Promising for Mathematically Gifted Students?

Zimmer, Christina, Hamilton, and Weber Prine (2006) noted that, in a recent survey of teachers implementing the No Child Left Behind Act, more than half the teachers surveyed felt that implementation of the law resulted in improved learning opportunities for low-performing students but that teachers and administrators at all levels of schooling worried about high-achieving students receiving adequate instructional challenge in all curricular areas. This review of the research literature explored the immediate and delayed impacts of gifted education approaches aimed at accelerating students' mathematics instruction (e.g., by covering 2, or even 4 years of high school mathematics in 15 months) and those that attempt to provide enrichment or extension activities for mathematically precocious students. This question is addressed in the category of student-mathematics interactions because it is very much about the pace and structure for engaging gifted students with mathematics content.

4. Interactions Between Teachers and Mathematics

Teachers engage with the mathematical content that they teach in various aspects of teaching practice: in planning and designing lessons, in interpreting and responding to student questions, and in the work of assessing their students' mathematical knowledge. Fortunately, formative assessment is an area of great contemporary interest and is also an area with a rich set of rigorous experimental field studies.

a. What Is the Impact of Use of Formative Assessment in Mathematics Teaching?

Educators at all levels realize the importance of assessing their students' progress during the year (i.e., formative assessment). Interest in formative assessment has dramatically increased since No Child Left Behind required states to establish accountability systems. Teachers' interpretation and use of the data available to them from instructionally embedded, in-class assessment in the context of teaching, along with high-stakes assessments are critical for improving outcomes for all students. However, many different systems have been established and touted for use as formative assessments. These range from the end-of-unit and mastery tests that accompany major commercial textbook series, to more contingent and informal probes of students' understandings to be used while they solve problems, to weekly tests that sample from the year's instructional objectives in mathematics. The Task Group examined rigorous experimental studies of the impact of teachers' use of formative assessment on students' growth in mathematics proficiency.

BIBLIOGRAPHY

Anderson, J.R., Reder, L.M., & Simon, H.A. (1996). Situated learning and education. *Educational Researcher, 25*(4), 5–11.

Cohen, D.K., Raudenbush, S.W., & Ball, D.L. (2003). Resources, instruction, and research. *Educational Evaluation and Policy Analysis, 25(2),* 119–142.

National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (NIH Publication No. 00-4769). Washington, DC: Author.

National Research Council. (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, DC: National Academy Press.

National Research Council. (2002). Scientific research in education. In R.J. Shavelson and l. Towne (Eds.), *Committee on scientific principles for educational research*. Washington, DC, Center for Education: National Academy Press.

National Research Council. (2004). *On evaluating curricular effectiveness: Judging the quality of K–12 mathematics evaluations.* Committee for a Review of the Evaluation Data on the Effectiveness of NSF-Supported and Commercially Generated Mathematics Curriculum Materials. Washington, DC: The National Academies Press.

Woodward, J., & Montague, M. (2002). Meeting the challenge of mathematics reform for students with learning disabilities. *Journal of Special Education, 36*(2), 89.

Zimmer, R., Christina, R., Hamilton, L. S., & Weber Prine, D. (2006). *Evaluation of two out-of-school programs in Pittsburgh public schools: No Child Left Behind's supplemental educational services and state of Pennsylvania's educational assistance program*. (No. WR-451-PPS). Santa Monica, CA: RAND.

II. Teacher-Directed and Student-Centered Instruction in Mathematics

The Task Group, in its initial activity of formulating key flashpoint questions about mathematics instruction, identified the question, "Is teacher-directed instruction more effective than student-centered instruction?" as one needing particular attention because of pressures faced by teachers being urged to use one or the other of these styles exclusively.

The terms "teacher-directed instruction" and "student-centered instruction" are sometimes used as labels to stake out starkly contrasting views in discussions about mathematics teaching. In their purest forms, these labels convey images of instruction that are in some sense polar opposites. For some, these terms convey differences of perspective about whether the goals of teachers or the needs of students should have primacy in determining what specific mathematics teaching interventions will be used in the mathematics classroom. Some have interpreted "student-centered" instruction to mean that students, rather than teachers, control the direction and content of the mathematical discussion, or that students are expected to somehow learn all mathematics on their own, by teaching one another. "Teacher directed" instruction has been interpreted in similarly extreme ways, to mean that teachers are not responsive to, or aware of, students' learning issues, and instead dispense mathematics instruction in a way that is disconnected from the learners. The distinction has been summarized by some with the ubiquitous "sage on the stage rather than a guide on the side" maxim. The idea of the "guide on the side" is often associated with intentions of mathematics education reforms in the past two decades concerning the role of the teacher.

The National Research Council (NRC) report *Adding It Up* acknowledges the challenge of such labels in discussing teaching: "Much debate centers on forms and approaches to teaching: 'direct instruction' versus 'inquiry,' 'teacher-centered' versus 'student-centered,' 'traditional' versus 'reform.' These labels make rhetorical distinctions that often miss the point regarding the quality of instruction. The quality of instruction is a function of teachers' knowledge and use of mathematical content, teachers' attention to and handling of students, and students' engagement in and use of mathematical tasks." (2001, p. 315). This section undertakes a circumscribed treatment of what have perhaps become positions that have hardened into ideologies that rarely offer pragmatic guidance to teachers on how they should teach.

A. Literature Review

The caricatures of teacher-directed and student-centered instruction that have sometimes emerged in debates on this subject are not validated in the versions of teacher-directed and student-centered instruction that were examined in the studies reviewed. Indeed, teacher-directed instruction involves assessment and careful attention to student progress—students were very much involved in the versions of teacher-directed instruction described in these studies. And, teachers have a key role in the versions of student-centered instruction described here as well—they choose tasks, direct discussion, and work toward mathematical goals. The Task Group found no examples of studies in which students were teaching themselves or each other without any teacher guidance; nor did the Task Group find studies in which teachers conveyed mathematics content directly to students without any attention to their understanding or response. The fact that these terms, in practice, are neither clearly nor uniformly defined, nor are they true opposites, complicates the challenge of providing a review and synthesis of the literature.

The literature review presented below will not end the debate over student-centered and teacher-centered instruction. Instead, it offers a summary of what is currently known about effective instructional practices in mathematics as they relate to teacher-directed or student-centered approaches, drawing on an exhaustive search based on terms that have been used in the literature to describe both teacher-directed and student-centered instructional approaches. Using the search terms provided in Appendix B, only studies of how instruction influences mathematics achievement were included. Studies of how instructional approaches affect students' motivation, social skills, attitudes toward mathematics, or other noncognitive outcomes were not reviewed. The search found 40 randomized experiments or quasi-experiments that were determined to have a rigorous design and to be relevant to the topic.

Blanket statements endorsing a philosophy of mathematics education will not be found. Even when examining high-quality studies, considering context is crucial to properly interpreting results. In other words, some approaches may be shown to be effective, but confidence in their effectiveness is only warranted under specified conditions. Factors such as the age of students, the mathematical content that is taught, the duration of the instructional program, the preparation of the teachers, and the outcomes that are sought must be taken into account.

Consequently, this literature review comes with a warning. Educators should be leery of sweeping claims that "best practices" in mathematics instruction are known and supported by research. Most efforts to promote any single all-encompassing style of instruction, to the exclusion of any others, are based on beliefs, not science, and much of the research cited to promulgate those beliefs does not meet minimal standards of quality. A body of high-quality research simply does not exist to answer such broad questions as whether teacher-directed or student-centered instruction should be dominant in teaching mathematics.

1. What Is Meant by Teacher-Directed Instructional Strategies?

Interpretations of teacher-directed instructional strategies gleaned from the literature vary widely. Common to most is the notion that the teacher has complete control of the instruction. Perhaps the best-known instantiation of teacher-directed instructional strategies as conceptualized in the late 1960s and 1970s was in the context of Project Follow Through, and was called Direct Instruction (Gersten & Carnine, 1984). Project Follow Through, a part of President Lyndon Johnson's War on Poverty in 1967, has been reported to be the largest and most expensive federally funded experiment in education ever conducted (Becker, 1977; Gersten et al., 1984). There were 17 distinct instructional models represented in the Follow Through evaluation (Stebbins, St. Pierre, Proper, Anderson, & Cerva, 1977, p. 2; see also Stallings, 1975, and Stallings & Kaskowitz, 1974), and Direct Instruction was one of these models.

Direct Instruction was a behaviorally oriented educational program using a tightly controlled teaching methodology and highly structured instructional materials. The instruction was programmed, emphasizing children's learning of intelligent behavior through programmed questions and answers provided in a fast-paced fashion. "Teachers present specified questions.… Proper responses are reinforced and incorrect answers are corrected according to specified procedures" (Stebbins et al., 1977, p. 65).

According to Kameenui, Carnine, Darch, and Stein (1986), the model of Direct Instruction, as described by Gersten and Carnine (1984) "employs clearly articulated teaching sequences that contain explicit, step-by-step teacher modeling and a means of assessing student mastery at each step of development" (p. 635). Meyer, Gersten, and Gutkin (1983) describe the component of the "Direct Instruction Model:" "a) consistent focus on academic objectives; b) high allocations of time to small-group instruction in reading, language, and math; c) the tight, carefully sequenced Distar curriculum; … e) a comprehensive system for monitoring both the rate at which students progress through the curriculum and their mastery of the material covered" (p. 243).

In work of the same period, Good and colleagues described and studied what they termed "active mathematics teaching" (see Good & Grouws, 1977, 1979; Good, Grouws, & Ebmeier, 1983). Guidelines for instruction in this program indicate a highly structured and prescribed instructional sequence, including: daily review, development, seatwork, as well as homework assignments and special reviews. The development sequence includes explanations, demonstrations, and illustrations, as well as repetition and elaboration (Brophy & Good, 1986, p. 348). Kameenui et al. (1986) provide details about what the development component of active mathematics teaching involves: "The direct approach to development views the teacher as one who controls the instructional goals and pace, chooses the appropriate materials, and provides immediate and academically oriented feedback to the learner." And, in this same vein, work by Slavin (1980) has examined "focused instruction," which involves a "highly structured schedule of teaching, worksheet work, and quizzes." (As described in Beady, Slavin, & Fennessey, 1981, p. 519).

More recently, reform documents of the past two decades have argued against teacher-directed instruction, not the same very specific, programmed kind of direct instruction of the 1960s and 1970s but rather a more general type of instruction in which the teacher is the primary authority. For instance, the National Council of Teachers of Mathematics *Curriculum and Evaluation Standards for School Mathematics* notes: "In many classrooms, learning is conceived of as a process in which students passively absorb information, storing it in easily retrievable fragments as a result of repeated practice and reinforcement" (National Council of Teachers of Mathematics, 1989, p. 10). It is contrasted with an instructional style that emphasizes a "constructive, active view of the learning processes" (p. 10)—which is not exactly aligned with the student-centered view of the 1960s and 1970s.

In summary, the hallmarks of teacher-directed instructional strategies include clearly prescribed instructional sequences, consistent focus on content objectives, emphasis on explanation, assessment and correction of errors, feedback to students and assignments and review, in which the teacher is doing all of these things. In addition, teacher-directed instruction can be manifested in the way the classroom is organized, and is often associated with whole group instruction. Most important is that the teacher is doing the teaching.

For the purpose of this review, "teacher-directed" instruction is viewed as instruction in which primarily the teacher is communicating the mathematics to the students directly and in which the majority of interactions about the mathematics are between the teacher and the student.

2. What Is Encompassed in "Student-Centered" Approaches to Mathematics Instruction?

More than a century ago John Dewey urged educators to consider the notion of what some have called "child-centered" education: "The teacher is not in the school to impose certain ideas or to form certain habits in the child, but is there as a member of the community to select the influences which shall affect the child and to assist him in properly responding to these influences" (Dewey, 1897, pp. 77–80).

The emphasis on the centrality of the student in education has been interpreted in various ways in mathematics education over several decades, most recently in the standards reform movement of the late 1980s and subsequent extensions. Common to most of these interpretations is the notion that the students' experience, motivation, interest, and knowledge needs to be a central consideration in the teachers' design and implementation of instruction. In addition, the focus on teachers' relationships with students is sometimes central. In recent years, numerous policies and programs have promoted a student-centered emphasis, invoking various theories of learning. Constructivism is one of these theories. See Cobb (2007) for a discussion of the ways in which theoretical and philosophical perspectives influence mathematics education.

At least three of the Project Follow Through instructional strategies can be classified as student-centered. The Tucson Early Education Model (TEEM) was "based on the concept that each child has a unique growth pattern with individual rates and styles of learning" (Stebbins et al., 1977, p. 41). TEEM took as a premise that formal learning should have as its basis the experiences young children bring to the classroom. "Some classroom activities are selected and structured by the teacher, and others are chosen by the children" (p. 41). The second model, the Cognitively Oriented Curriculum model, was a Piagetian developmental model focused on developing children's ability to reason. The goals included helping children sustain independent activity, define and solve problems, assume responsibility for decisions and actions, and work cooperatively (p. 2, p. 89). And, the Education Development Center (EDC) Open Education approach, with its roots in the philosophy of the British Infant Schools and the developmental theories of Piaget, provided children with a wide range of materials for learning.

Another clearly defined approach to student-centered instruction was developed by Flanders and his colleagues, based on the Flanders Interaction Analysis Categories (FIAC) (see Flanders, 1970; discussed in Brophy and Good, 1986). According to Brophy and Good, "Flanders believed that there was too much teacher talk and not enough student talk in most classrooms, so that teachers should be more 'indirect'" (p. 333). This style of instruction involved examining pupil attitudes and emphasized "asking questions, accepting and clarifying ideas or feelings, and praising or encouraging as indirect techniques" (p. 333). The student-centered interventions of this time period, often aimed at primary and early elementary age children, featured elements of free choice and developmental readiness.

Lampert (1990) has compared school mathematics with knowledge in the discipline of mathematics, noting that "few teachers engage students in a public analysis of the assumptions that they make to get their answers" (p. 32). She summarizes the assumptions of reform documents (National Research Council, 1989; National Council of Teachers of Mathematics, 1989) as follows: "Reform documents recommend that mathematics students should be making conjectures, abstracting mathematical properties, explaining their reasoning, validating their assertions, and discussing and questioning their own thinking and the thinking of others" (p. 33). This might be viewed as a description of a "student-centered" approach, although of course such approaches could be in place in a teacher-directed classroom as well. Socratic teaching methods, for example, feature teacher-directed dialogues between teachers and students. Close questioning requires students to justify thinking out loud and to explain the logic behind their arguments and conclusions.

In the National Research Council report *How People Learn* (National Research Council, 2000), the term "learner centered" is used to:

> *…refer to environments that pay careful attention to the knowledge, skills, attitudes, and beliefs that learners bring to the educational setting. The term includes teaching practices that have been 'culturally responsive,' 'culturally appropriate,' 'culturally compatible,' and 'culturally relevant' (Ladson-Billings, 1995). The term also fits with 'diagnostic teaching' (Bell et al., 1980): attempting to discover what the student is thinking in relation to the problems on hand, discussing their misconceptions sensitively, and giving them situations to go on thinking about which will enable them to readjust their ideas (Bell, 1982). Teachers who are learner centered recognize building on the conceptual and cultural knowledge that students bring with them to the classroom (pp. 133–134).*

To be sure, some depictions of student-centered instruction emphasize a passive role for teachers. The Bureau of Labor Statistics, for example, in its *Occupational Outlook Handbook* describes the job of teaching as follows:

> *Teachers act as facilitators or coaches, using classroom presentations or individual instruction to help students learn and apply concepts in subjects such as science, mathematics, or English. They plan, evaluate, and assign lessons; prepare, administer, and grade tests; listen to oral presentations; and maintain classroom discipline. Teachers observe and evaluate a student's performance and potential and increasingly are asked to use new assessment methods. For example, teachers may examine a portfolio of a student's artwork or writing in order to judge the student's overall progress. They then can provide additional assistance in areas in which a student needs help. Teachers also grade papers, prepare report cards, and meet with parents and school staff to discuss a student's academic progress or personal problems.*

> *Many teachers use a 'hands-on' approach that uses 'props' or 'manipulatives' to help children understand abstract concepts, solve problems, and develop critical thought processes. For example, they teach the concepts of numbers or of addition and subtraction by playing board games. As the children get older, teachers use more sophisticated materials, such as science apparatus, cameras, or computers. They also encourage collaboration in solving problems by having students work in groups to discuss and solve problems together. To be prepared for success later in life, students must be able to interact with others, adapt to new technology, and think through problems logically (U.S. Department of Labor, Bureau of Labor Statistics, 2008).*

In summary, the elements of student-centered mathematics instruction as described in contemporary treatments include emphasis on student responsibility and independence; acknowledgment of students' experiences, prior knowledge, and interests and motivations in the design of mathematics instruction; and the centrality of students' thinking and students teaching other students in the classroom. Teachers facilitate, encourage, and coach but do not explicitly instruct by showing and explaining how things work.

For the purposes of this review, "student-centered" instruction is viewed as instruction in which primarily students are doing the teaching of the mathematics and that the majority of the interactions about the mathematics occurs between and among students.

The vague and often overlapping ways in which "teacher-directed" and "student-centered" are used in the literature, not to mention in contemporary discourse, present challenges for any attempt to summarize research on the topic. A major source of the ambiguity stems from the use of these adjectives to modify several different nouns. As illustrated in the citations above, by their very nature nouns such as "education," "environments," "practices," or "learning" comprise a collection of activities. The Task Group chose to focus on one element—instruction—and to search for studies that contrast who is doing the teaching—teachers or students? The contrast never exists in an absolute sense, of course, but in degree. All of the studies in our review compare an instructional regime in which teachers do more teaching (and therefore students less) with one in which students do more teaching and teachers less.

This focus was chosen because teachers told the Panel that they understand the expectations of administrators in their districts are that they teach exclusively in teacher-directed ways, essentially as it has been defined here. And, other teachers have said that their administrators are critical unless they are teaching in student-centered ways, again as it has been defined here. Thus, this review was undertaken to highlight these distinctions in ways that will hopefully help policymakers and teachers to engage in practice that is evidence based.

In accordance with the definitions of teacher directed and student centered being used, the focus is on the nature of the mathematics **instruction** (literally the interactions between the teachers and the students about mathematics). Not included are studies in which the nature of the **curriculum** (the materials for learning) might be construed as more or less teacher-directed or student-centered. Note that most current interpretations of what it means to be teacher-directed or student-centered conflate issues of instruction and curriculum. For example, the use of practice worksheets (a curricular device) might be associated with a teacher-directed approach but indeed could be highly student-centered in its design.

Within the review, a number of studies were found that directly compare a form of teacher-directed instruction to a form of student-centered instruction. These studies are discussed in the first section. Later sections address studies that have looked at student-centered classroom organizational approaches of cooperative groups and peer-tutoring approaches.

Methodological considerations specific to this section can be found in Appendix A.

3. Comparisons of Student-Centered and Teacher-Directed Approaches to Instruction

Research Studies. Eight studies meeting the criteria to be considered Category 1 studies were located that compared student-centered and teacher-directed approaches to instruction (see Table 1). The pattern of effects is quite complex. It is not possible to undertake a meta-analysis of these studies because the interventions are all of such distinct types, according to the above categorization, that pooling effect sizes is not meaningful. The specific interventions studied in the Project Follow Through evaluation study (Stebbins et al., 1977) are treated in a separate section.

Table 1: Studies That Investigated the Effects of Teacher-Directed and Student-Centered Instruction on Mathematics Achievement

Study	Design	Sample	Duration/Content	Contrast	Measure	Subgroup	Hedge's g	Standard error	
Brenner et al., 1997[a]	RCT	128 students in six intact pre-algebra classes in three junior high schools in southern California	20 days/ Meaningful thematic contexts used in pre-algebra concepts	Guided discovery approach vs. Traditional textbook	Pooled problem solving outcomes: word problem solving (ES = 0.110), function word problem (ES = 0.393), and equation solving (ES= -0.281) measures	Overall	0.074	(ns)	0.399
					Pooled representation outcomes: function word problems (ES = 0.877*) and word problems (ES = 0.623)	Overall	0.750	~	0.403
Ciccelli, 1982	RCT	64 fifth-grade students	40 minutes per day for 9 days/ Probability and graphing	Direct vs. Nondirect instruction	Math achievement test	Low ability	-0.614	(ns)	0.569
						Medium ability	0.156	(ns)	0.311
						High ability	-0.374	(ns)	0.535
Fuchs et al., 2006[c]	RCT	445 third-grade students in 30 classrooms in seven schools in an urban district	16 weeks/ Mathematical problem solving strategies	Schema broadening instruction vs. Control	Pooled transfer measures	Overall	0.545	(ns)	0.439
				Schema broadening instruction-real life vs. Control	Pooled transfer measures	Overall	1.077	*	0.464
Hopkins et al., 1997	Quasi	34 third-grade and 40 fifth-grade students	1 30-minute session/ Arithmetic	Didactic vs. Constructivist approach	Arithmetic computation test	Boys	0.155	(ns)	0.345
						Girls	1.142	***	0.327
Kameenui et al., 1986 – Study 3	RCT	24 fourth-grade students	11 daily 35-minute sessions/ Division	Direct Instruction (Project Follow Through) vs. Control	Math achievement test	Overall	0.444	(ns)	0.399
Muthukrishna & Borkowski, 1995	RCT	54 third-grade students	14 consecutive class days/ Addition and subtraction word problems	Guided discovery approach vs. Direct strategy instruction	Pooled near transfer outcomes (classifications: ES = -0.006, sequence: ES = -0.346, comparison: ES = -0.383)	Overall	-0.245	(ns)	0.276
					Pooled far transfer outcomes (form: ES =0.576*, context: ES =0.380)	Overall	0.478	~	0.278

Continued on p. 6-19

Table 1, continued

Study	Design	Sample	Duration/Content	Contrast	Measure	Subgroup	Hedge's g	Standard error	
Rittle-Johnson, 2006	RCT	85 third- through fifth-grade students in an urban parochial school	1 40-minute session/ Mathematical equivalence	Discovery learning and prompts to explain vs. Direct instruction and prompts to explain	Procedural learning test	Overall	-0.272	(ns)	0.301
					Procedural transfer test	Overall	-0.125	(ns)	0.300
					Conceptual knowledge test	Overall	-0.458	(ns)	0.304
				Discovery learning and no prompts vs. Direct instruction and no prompts	Procedural learning test	Overall	-0.719	*	0.313
					Procedural transfer test	Overall	-0.085	(ns)	0.303
					Conceptual knowledge test	Overall	0.050	(ns)	0.303
Rudnitsky et al., 1995[a]	RCT	401 third- and fourth-grade students in 21 classrooms in six schools	18 days/ Addition and subtraction word problems	Writing and discussion vs. Practice and explicit heuristics	Near transfer posttest	3rd-grade females	-0.073	(ns)	0.428
						3rd-grade males	0.392	(ns)	0.429
						4th-grade females	-0.138	(ns)	0.431
						4th-grade males	0.462	(ns)	0.417
Project Follow Through Evaluation									
Stebbins et al., 1977—Direct Instruction Model[b]	Quasi	316 Project Follow Through and 317 Non-Project Follow Through students enrolled in program from kindergarten through third grade in five districts (New York, NY; Grand Rapids, MI; W. Iron Co., MI; Flint, MI; and Providence, RI)	Kindergarten through 3rd grade/ General elementary school mathematics curriculum	Direct Instruction Follow Through vs. Non-Follow Through	Overall Metropolitan Achievement Test (MAT) outcome: computations (ES = 0.315*), concepts (ES = -0.064), and problem solving (ES = 0.017) measures	Overall	0.105	(ns)	0.142
Stebbins et al., 1977—Cognitive Curriculum Model[b]	Quasi	177 Project Follow Through and 337 Non-Project Follow Through students enrolled in program from kindergarten through third grade in five districts (New York, NY; Okaloosa Co., FL; Greeley, CO; Seattle, WA; and Chicago, IL)	Kindergarten through 3rd grade/ General elementary school mathematics curriculum	Cognitively Oriented Curriculum Follow Through vs. Non-Follow Through	Overall Metropolitan Achievement Test (MAT) outcome: computations (ES = -0.318~), concepts (ES = -0.355*), and problem solving (ES = -0.295~) measures	Overall	-0.357	*	0.167
Stebbins et al., 1977—EDC Open Education Model[b]	Quasi	248 Project Follow Through and 487 Non-Project Follow Through students enrolled in program from kindergarten through third grade in five districts (Philadelphia, PA; Burlington, VT; Lackawanna Co., PA; Morgan Comm. Sch., DC; and Paterson, NJ)	Kindergarten through 3rd grade/ General elementary school mathematics curriculum	EDC Open Education Follow Through vs. Non-Follow Through	Overall Metropolitan Achievement Test (MAT) outcome: computations (ES = 0.052), concepts (ES = -0.081), and problem solving (ES = -0.073) measures	Overall	-0.037	(ns)	0.140

~ p < .10, * p < .05, ** p < .01, *** p < .001
[a] Data were adjusted for clustering that occurred within classrooms.
[b] Data were adjusted for clustering that occurred within schools.
[c] These studies use classroom-level analyses.

The studies that produced significant effect sizes in contrasts comparing teacher-directed to student-centered instruction will be discussed. Hopkins, McGillicuddy-De Lisi, and De Lisi (1997) investigated different ways of teaching third- and fifth-grade students how to compute with whole numbers and fractions. The experiment consisted of a single 30-minute session in which students were taught individually. The researchers were interested in determining if didactic or constructivist instruction, in this case instruction emphasizing particular mathematical practices, benefits girls and boys differentially when learning mathematical computation. Two groups of children—matched on a pretest of computation skills, grade, and gender—were formed at third and fifth grades. In a single 30-minute session, conducted with individual students, students were taught how to solve six computation problems involving addition, subtraction, multiplication, and division with whole numbers and fractions. Whole number operations included multiplying three digit by three digit numbers and the items with fractions included addition and subtraction of mixed numbers without a common denominator. One group was instructed using a didactic approach in which the mathematical practices of algorithms, rules, and solution methods were explicitly taught. A constructivist teaching strategy was used with the other group. In that setting, teachers suggested alternative mathematical practices, including ways to organize tasks, recasting children's comments with tag questions requesting clarification, and providing demonstrations that guided students to discover solutions. Both groups were then post-tested on computation. No effect was found for boys, but girls in the didactic instructional groups made statistically significant gains over girls who received constructivist teaching (ES = 1.142). The gains for girls were apparent at both grade levels.

The literature search uncovered two high-quality studies that found evidence of far transfer, under very limited conditions. Both studies investigated how to teach problem solving strategies, and both studies found guided discovery (a particular interpretation of student-centered instruction) more effective than direct instruction methods. Muthukrishna and Borkowski (1995) conducted an experiment consisting of 14 lessons teaching third graders the number family (or part-whole) strategy for solving word problems with addition and subtraction. The strategy involves a part-whole schema in which one larger quantity (known or unknown) is thought of as a whole comprising two smaller quantities (known or unknown) that are parts. For example, a student who knows that 1 + 5 = 6 can conceptualize 6 as a whole made up of 1 and 5 as parts, with two subtraction facts, 6 - 1 = 5 and 6 - 5 = 1, derived from the addition fact. Students were taught that the unknown in a word problem involving addition and subtraction is either a whole or a part. Students were randomly assigned to four conditions for instruction: direct strategy teaching, guided discovery, a combination of direct teaching and guided discovery, and a control condition. The number family schema was not taught to the control group, and students in the control condition primarily received instruction from their regular classroom teacher. The other groups were instructed by the experimenter and an assistant. A typical guided discovery lesson consisted of 20 minutes working in pairs followed by whole class discussion of students' solutions. Students in the guided discovery group worked with a variety of manipulative materials and did not engage in individual paper and pencil activities. The post-test consisted of addition and subtraction word problems assessing both near and far transfer of skills.

Additional aspects of the study may be of use in interpreting the results. Although a 14-day period is relatively brief for studying instructional methods, it is a long time to teach a single problem solving strategy to third-graders. Note also that the skill emphasized here—using the number family strategy—was used to solve mathematics problems that are typically far below the grade level of the students in the study. Students who had mastered solving the problem types used in the study were excluded, as were students who lacked the basic skills to solve addition and subtraction word problems not requiring regrouping or carrying.

Calculation of effect sizes on the near transfer outcomes, contrasting the discovery group and the direct instruction group, revealed no significant effect sizes. On the far transfer problems, the guided discovery group outperformed the direct strategy teaching group on the test for form transfer (problems of a different form than those in instruction) with a significant effect size (ES = 0.576). There was no significant effect on the test for context (problems presented in a context different from what had been used in instruction) when contrasting the same two groups. Nonetheless, the pooled effect size on the far transfer measures approached significance (ES = 0.478), favoring the guided discovery group.

A study by a team of researchers at the University of California at Santa Barbara (Brenner et al., 1997) investigated middle school pre-algebra students who were learning how to represent function problems in multiple formats. Problem representation involves constructing and using mathematical representations in words, graphs, tables, and equations, a difficult task particularly for many students making the transition from arithmetic to algebra. In pre-algebra textbooks, function word problems are typically represented by equations, tables consisting of ordered pairs, and graphs.

The intervention consisted of a 20-day instructional unit taught to 128 seventh- and eighth-graders in three schools. Three teachers each had two pre-algebra classes; one class was randomly assigned to the treatment and the other to the control condition. Students in the treatment groups worked in heterogeneous groups and used manipulative materials. Teachers used a guided discovery approach in which students were encouraged "to explore different representations and to develop their own understanding of each one" (Brenner et al., p. 668). In the control groups, students were taught with textbooks and teachers used traditional direct instructional methods.

Results were mixed. Five outcomes were tested in the study: word problem solving, function word problems, equation solving, function representation, and word problem representation. One effect size reached statistical significance. Students in the student-centered strategies groups were better able, at a significant level, to represent function problems in multiple ways (ES = 0.877). The subtest of the assessment instrument measuring this outcome awarded two points per item. For example, students were given the problem: "Mary Wong just got a job working as a clerk in a candy store. She already has $42. She will earn $7 per hour. How many hours will she have to work to have a total of $126?" (Brenner et al., p. 671). For the subtest for representation, students received one point for drawing a diagram, chart, table, or graph to represent the problem and one for writing the correct equation in the form of $y = mx + b$. Having the correct answer had no bearing on the score for the representation subtest. The effect size on the word problem representation outcome measure (ES = 0.623) did not reach statistical significance, and the pooled far transfer outcomes (using the two representation outcome measures) bordered on statistical significance (ES = 0.750). On the subtests reflecting ability to find problem solutions correctly, effect sizes were not significant, but favored the guided discovery group with effect size 0.393 on function word problem outcome, and the traditional textbook group (ES = -0.281) on equation solving measures.

In a 2006 paper, Rittle-Johnson reported on a comparison of teacher-directed instruction and student-centered instruction. In the comparison of the discovery learning and direct instruction conditions without prompts, there was a significant effect size favoring the direct instruction condition on the procedural learning test.

Her formulation of teacher-directed is based in information-processing and cognitive theories about working memory capacity, while her view of student-centered instruction draws on Piaget and current reformers who emphasize the importance of discovery learning. In this study of third through fifth grade students, children were assigned randomly to one of four conditions, based on two factors: "direct instruction versus discovery learning" and prompts for self-explanation vs. no prompts. Self-explanation is "generating explanations for oneself" (Rittle-Johnson, 2006, p. 1). The mathematical focus was on equivalence—the ideas that equations represent balance and that the same quantity is on both sides. Rittle-Johnson points out the importance of this idea as a precursor to algebra.

Assessments involved a pretest on mathematical equivalence problems, a posttest given immediately after the intervention, and then delayed posttest. The intervention was done in a single session, in which students worked one-on-one with a researcher in a 40-minute session. Children solved problems, reported on their solution strategies, and were provided with feedback. In the direct instruction condition students were told explicitly how to solve the problems. In the student-centered condition no instruction was given, and children were asked to "think of a new way to solve the problem." Prompts for doing self-explanation were introduced in the two groups in that condition. The posttests measured procedural learning, procedural transfer, and conceptual understanding. Comparison of the discovery learning and prompts group to the direct instruction with prompts group yielded nonsignificant but negative (i.e., favoring the direct instruction group) results on all three outcome measures (note ES = -0.272 and -0.458 on the procedural and conceptual tests, respectively).

Four additional studies involved comparisons, in some form, of teacher-directed and student-centered strategies, in which no significant effect sizes were found. The studies by Cicchelli (1982) and Kameenui et al. (1986) introduced a direct instruction-type treatment and compared to more student-centered instruction. The Rudnitsky et al. (1995) and Fuchs et al. (2006) studies both implemented clearly specified student-centered instruction and compared to more of a "business as usual" condition. Effect sizes in Kameenui et al., while non-significant (ES = 0.444) favored the direct instruction condition. In the Ciccelli study, the effect favored the nondirect instruction condition for low and high ability students (ES = -0.614 and -0.374, respectively) and the direct instruction group for medium ability students (ES = 0.156).

Because this set of studies differs in terms of the nature of the intervention, pooled effect sizes have not been calculated. In summary, note that there is no conclusive evidence from this set of studies to support either a teacher-directed or student-centered approach to mathematics instruction.

4. Project Follow Through Evaluation Studies

Three relevant models of the extensive National Longitudinal Evaluation of Project Follow Through (Stebbins et al., 1977) were located that met the criteria for inclusion. At that point in history, the evaluation of Project Follow Through (FT) was the largest and most expensive evaluation of any intervention conducted in education or any of the social sciences. Follow Through's evaluation involved longitudinal quasi-experiments conducted with kindergarteners through third-graders in low-income schools across the country. The outcome measures testing mathematical achievement were the computations, concepts, and problem solving subtests of the Metropolitan Achievement Test. The three Follow Through models included in this report that tested teacher-directed or student-centered instruction and had equivalent groups at baseline were Direct Instruction, Cognitively Oriented Curriculum, and EDC Open Education. Results are reported in Table 1.

Direct Instruction "use(d) a fast moving series of programmed … (i.e., scripted) ... questions and answers … teachers present specified questions to elicit a verbal child response. Proper responses are reinforced and wrong answers are corrected according to specified procedures" (Stebbins et al., 1977, p. 65). Programmed instruction materials are used, and students work in small homogeneous groups; frequent criterion-referenced tests are provided. The Task Group considers this to be a teacher-directed intervention.

The EDC Open Education approach states, "Children learn at individual rates and in individual ways…" (Stebbins et al., 1977, p. 113). The approach has its roots in British infant schools and in the developmental theories of Piaget. The instruction occurs in an open setting and children are provided with a wide range of materials for learning. This is a student-centered model. The Cognitively Oriented Curriculum model, also a developmental model, was aimed at "developing children's ability to reason." The curriculum is based on the use of learning centers in which "children choose their activities and work with teachers in small groups" (Stebbins et al., p. 89). This too is a student-centered intervention.

Concurrent with the impact evaluation, Stallings (1975), conducted an extensive observational study of the activities in the Follow Through classrooms and their corresponding control group (i.e., business as usual) classrooms. Using a complex, reliable observational system, they were able to predict which classrooms were FT and which were control, and to discriminate between each example (FT) and control classroom with over 80% accuracy. They consistently found significant differences between each FT approach and its control condition, and between the various FT models.

When Stallings (1975) compared the Direct Instruction approach with "non Follow Through" instruction, which would have been a version of "business as usual" at the time they were essentially comparing a highly structured, teacher-directed intervention (based on principles of instructional design and concept development derived from learning theory and applied behavior analysis, to say nothing of the genius of Englemann) with a more general type of teacher-directed instruction.

The 1977 project evaluation (Stebbins et al., 1977) found significant effect sizes favoring Direct Instruction on the Computation subtest only (ES = 0.315). No other effect sizes were significant, nor was the overall effect on the Metropolitan Achievement Test Mathematics composite significant.

Another Project Follow Through model study compared the Cognitively Oriented Curriculum to a non-Follow Through business as usual condition. The effect size on the Concepts subtest was significant (-0.355), and the effect sizes on the Computation and Problem Solving measures were marginally significant (ES = -0.318 and ES = -0.295). Note that all effect sizes were negative, favoring the teacher-directed control condition. The final study in this group involved a comparison of the Open Education program to a business-as-usual control condition. There were no significant effect sizes.

No pooling was done of these studies.

5. Conclusion

The studies produced a mix of significant effect sizes favoring student-centered instruction, and others favoring teacher-directed instruction, together with findings of no significant effects. As a result, the research does not lead to any conclusive result about the value of student-centered instructional strategies in comparison to teacher-directed instructional strategies. Under some conditions, with some groups of students, and for some kinds of outcomes, an isolated study may find that either teacher-directed or student-centered strategies are preferable. In general the evidence does not provide a case for favoring or promoting either strategy over the other. The Task Group points out that in only one of the studies reviewed is “teacher-directed” instruction the experimental treatment.

B. Cooperative Learning and Peer Tutoring

Cooperative learning strategies offer students an opportunity to learn from and with other students. However, the means by which cooperative learning plays out in classrooms varies along many dimensions. For instance, tasks assigned in cooperative learning groups range from practice on teacher-taught skills to learning methods of problem solving. Students can be grouped homogeneously or heterogeneously by ability. Students may be assigned specific roles within a cooperative group or they may decide for themselves how to accomplish a group task. Group and individual accountability operate differently in different cooperative learning settings. Slavin (1991) describe a cooperative learning strategy as when groups work to earn some type of recognition or award based on the individual learning of every group member. Group members’ individual learning is measured by success on assignments, quizzes, and tests. Students are motivated to help each other learn so that individual achievement increases and, as a result, the group receives awards or recognition. Cooperative learning may include individual accountability or group reward structures.

Good, Mulryan, and McCaslin (1992) conclude that small-group instructional approaches are supported by research that indicated students need to be more active. “[Research] suggests that students are too passive and need to become more involved intellectually in classroom activities" (p. 167). They go on to note, "Many writers interpret recent cognitive science research as suggesting the need for less teacher support and for more learner independence (see Nickerson, 1988), [although] what this means in practice is far from clear. For example, does heavy reliance on more talented peers mean less dependency on the teacher? Do different skills and concepts require different amounts of expert modeling and coaching so that simple statements about appropriate practice are highly misleading?" (p. 167).

A number of studies compare cooperative grouping strategies to more traditional whole-class instruction, or in some cases, to individual practice that is part of teacher-directed instruction. Our review is organized by the following categories: studies of very specific approaches to cooperative learning [Team Assisted Individualization (TAI), Student Teams-Achievement Division (STAD), Peer Assisted Learning (PALS)]; studies of other collaborative learning strategies; combination strategies involving cooperative or collaborative learning; and cooperative learning approaches in the context of technology.

Note that our search of the literature and analyses are not concerned with affective outcomes, only on measures of mathematics achievement.

1. Specific Approaches to Cooperative Learning Team Assisted Individualization (TAI)

This strategy combines individualization with cooperative work. In TAI, students are grouped in heterogeneous teams of four or five persons. Each student receives a set of mathematics problems tailored to individual performance on a diagnostic test. Students help each other when needed and check each others' work. Rewards are based on group performance on assignments, quizzes, and tests. Tests at the end of the unit are taken individually.

Six studies, described within four separate papers, met our criteria for review and examined the effect of TAI on some type of mathematical outcome. The pooled effect size for computation outcomes on student-level analyses was significant (ES = 0.377), favoring the TAI condition. Slavin, Leavey, and Madden (1984) report on two studies with elementary school students focused on computation with decimals and fractions, and with word problems. In one randomized controlled trial (RCT) study, TAI was contrasted with whole-class lectures and group-paced instruction. The second study, a quasi-experiment, involved fourth- through sixth-graders, with the same type of control condition. A third study (Slavin, Leavey, & Madden, 1984), involving 1,371 students in Grades 3 to 5, again compared TAI with whole class lectures. The fourth paper in this set (Xin, 1999) involved third-grade students in an RCT focusing on arithmetic topics including basic fact families, calculation, coins, and place value. A large number of mainstreamed special education students (14%) were involved. The TAI treatment was coupled with a CAI component; the control condition was whole class instruction coupled with the same CAI.

All of these studies allowed for student-level analyses to examine effect sizes on an outcome measure of computation, the California Test of Basic Skills-Computations (CTBS) for the Slavin studies, and the Stanford Achievement Test-Math for the Xin study. Five contrasts were examined—the three in the Slavin et al. studies, and results for two groups in the Xin study (regular education, and LD students). No effect sizes were significant separately, except for the significant effect size of 0.595 comparing TAI with computer-assisted instruction (CAI) in the Xin study to students in the regular education group.

Two additional studies, both RCTs, are reported in Slavin and Karweit (1985). One worked with students in Grades 4 through 6, the other with third- through fifth-graders, using TAI over a period of 18 and 16 weeks. The mathematics topics were decimal and fraction arithmetic, introduction to algebra, and word problems. The control conditions were the Missouri Mathematics Program (a form of teacher-directed instruction) and a business as usual control condition. Classroom-level effects (ES = 0.709, 0.562) on computation outcomes were significant and favored the TAI intervention on computation scores. In three of these studies a concept outcomes measure was also included. Slavin et al. (1984b) and Slavin & Karweit (1985)—Studies 1 and 2 included outcomes on the CTBS-Concepts test. Three contrasts were computed, all proving to be nonsignificant; the pooled effect size of the classroom-level analyses (0.018) also was nonsignificant.

The studies are summarized in Tables 2a and 2b. It can be concluded that the implementation of TAI for students in Grades 3 through 6, in comparison to a form of whole class instruction, benefits computation skills. Note that this finding applies only to the very particular cooperative group strategy of TAI and only to computation, not concepts or problem solving.

2. Student Teams-Achievement Division (STAD)

This form of cooperative learning developed by Slavin and colleagues, involves four-to-five member homogeneous teams studying together after teacher presentation. Individual quizzes are taken and rewards are at the team level.

Four studies of STAD, all randomized controlled experiments, met our criteria for review. They are summarized in Table 3. No significant effect sizes were produced in this set, although all effect sizes were positive, favoring the STAD intervention. Jacobs (1996) examined the performance of third- through fifth-graders, content not specified, in a STAD condition and then in an individual student accountability condition, and produced non-significant effects favoring STAD of 0.573, 0.484, and 0.454, for third-, fourth-, and fifth-graders respectively. STAD did not provide any particular advantage to student participants in comparison with more teacher-directed classroom strategies as implemented in these four studies.

3. Peer Tutoring Approaches

The studies in this section examine the impact of a small group instruction approach that features peers learning from and with their peers, in variations of peer-tutoring. One particular version, *Peer-Assisted Learning Strategies* (PALS) (http://kc.vanderbilt.edu/pals/), "is a version of classwide peer tutoring. Teachers identify which children require help on specific skills and [whom] the most appropriate children are to help other children learn those skills. Using this information, teachers pair students in the class, so that partners work simultaneously and productively on different activities that address the problems they are experiencing. Pairs are changed regularly, and over a period of time as students work on a variety of skills, all students have the opportunity to be 'coaches' and 'players'." (http://kc.vanderbilt.edu/pals/). The strategy also creates opportunities for a teacher to circulate in the class, observe students, and provide individual remedial lessons.

Five studies that examine various versions of peer tutoring met our criteria for review, allowing for calculation of effect sizes for 20 different contrasts at the student and classroom levels, and for computation and concepts outcome measures. Three of the studies in this section use PALS. The studies are summarized in Table 4. Pooled effect sizes examining computation outcomes at the classroom level were significant (ES = 0.431), and approached significance at the student level, favoring the peer tutoring interventions. In none of these studies were significant effect sizes produced on concept outcome measures. Nor, when doing student-level analyses, were any individual effect sizes on computation significant.

The pooled effect size on the computation outcomes for the three studies that used student-level analyses was 0.238, which approached statistical significance. In a 15-week randomized controlled study of kindergartners (Fuchs et al., 2001) PALS was compared to a control condition that was described as teacher-directed lessons and demonstrations. Positive but not significant effects on the Stanford Early School achievement test in student-level analyses were detected for special education students, low-achieving students, and medium-achieving students (effect sizes respectively, of 0.431, 0.374, and 0.436). In a study of first-graders comparing PALS with a business-as-usual basal core curriculum (Fuchs, Fuchs, Yazdian, & Powell, 2002), no statistically significant effects were found. Ginsburg-Block and Fantuzzo (1998) implemented an RCT with low-achieving third- and fourth-grade students in an urban elementary school using a reciprocal peer tutoring (RPT) model (Palincsar & Brown, 1984) on the mathematics achievement of low SES students. Results were nonsignificant but favored the peer collaboration condition (ES = 0.590).

The effect sizes of the two studies that included classroom-level analyses were also pooled (Fuchs et al., 1995; Fuchs et al., 1997). These examined the impact of peer-assisted strategies, allowing for effect size calculations on seven different contrasts. The 1995 RCT study was done in second through fourth grade classrooms, over a 23-week period in which two 25–30 minute sessions per week were done using PALS, integrated with regular assessments. The control condition was teacher-mediated instruction, and the topic focus arithmetic operations. The 1997 study, also an RCT, also worked with second- through fourth-grade classrooms using peer-mediated versus teacher-mediated instruction (Fuchs et al., 1997). They investigated the effects of a peer tutoring program in which students received explicit instruction on how to provide elaborated help. Students were taught to provide explanations that would encourage peers to solve problems for themselves (instead of simply giving answers), referred to as "elaborated PMI." These interventions were modeled separately, with students assigned to two experimental treatments—PMI-elaborated and PMI elaborated plus conceptual. Both studies had both computation and conceptual outcome measures.

In both studies contrasts were calculated for LD, low-achieving, and average-achieving students. In the 1997 study there was also a comparison of high-achieving students. The results are interesting and mixed. In the 1995 study, the only significant effect size was for low-achieving students (ES = 0.728), on the computation measure, favoring PALS. Only two other effect sizes approached significance in this set: in the 1997 study, for both LD and low-achieving students, on the computation tests, the effect sizes were appreciable (0.663 and 0.704), on the computation outcome measure only, favoring the peer-assisted condition. No other effect sizes were significant, although all but one (in the 1997 study, LD on the concepts outcome measure) were positive. However, when the seven effect sizes for classroom-level computation outcome measures were pooled for these two studies, the result was a significant effect size (ES = 0.431).

In summary, it appears that peer tutoring strategies may be promising in teaching young children mathematical operations (which may not be exclusively computation oriented). However, this finding must be treated cautiously because the evidence is only suggestive. For the pooled effects of the Fuchs et al. (1995) and Fuchs et al. (1997) studies, there are some important limitations. The two studies were both in Grades 2–4, involve learning whole number operations, were possibly conducted using the same sample of schools, and were conducted by the same research team. The 1995 Fuchs et al. study did not include high-achieving students. Moreover, in the 1995 study, teachers in the peer tutoring condition were regularly provided with formative assessment data to guide instruction, but teachers in the control condition were not. Thus, the study does not provide a clear contrast between peer tutoring and a more teacher-directed form of instruction. The extent to which the positive effects that were detected were produced by formative assessment, by peer tutoring, or by an interaction of the two interventions cannot be determined.

4. Other Collaborative Learning Strategies

Five studies of other collaborative learning strategies met the criteria for inclusion, all of them RCTs. The studies are referred to as "collaborative learning" because they do not feature interventions as structured as the cooperative learning techniques featured above, but they all utilize methods of student grouping that involve student-to-student collaboration in learning. Two of the seven contrasts computed yielded effects that were significant.

A study by Barron (2000) produced statistically significant effect sizes favoring the collaborative condition in solving complex video-based mathematical problems. Sixth-graders enrolled in a public magnet school serving academically talented children were assigned randomly to either a group or individual condition. The task was to solve video-based problems from *The Adventures of Jasper Woodbury* series. Students first viewed the 15-minute episode, "Journey to Cedar Creek," which describes several dilemmas facing a character who is considering the purchase of a boat. In the second session, students were asked to solve these problems either individually or in teams of three by completing exercises in a workbook. In the third session, students were asked to solve the problems again, this time individually, regardless of the condition to which they had been assigned in the previous session. In the fourth and final session, students viewed a 5-minute video posing a parallel problem to assess near transfer of the acquired skills.

Students who had worked in triads solved more of the problems correctly than students who had worked individually at significant levels (ES = 0.472). The effect on a transfer task approached the level of significance (ES = 0.392).

In a study of the effects of a communal environment on African-American students' learning of mathematical estimation (Hurley et al., 2005), effects were significant. The researchers drew a sample of fifth-grade African-American students from two urban public schools. The students had to fall within the middle 75% for their classroom and school in terms of academics and behavior. The children in the sample came from a low SES background as measured by both their school's participation in Title I and the students' participation in the free and reduced-price lunch program.

Boys and girls were divided equally between two treatments though within gender students were randomly assigned. One experimental condition was highly communal; students learned estimation working in groups of three. Students sat at the same table and shared one set of materials. Each study session included the experimenter reading a communal prompt to the students. Students were asked to hold hands and were reminded that they were members of a group and should work hard and help one another because they were members of the same school and community. The other condition was low-communal; students studied alone in sets of three. Each student had his or her own set of materials and sat at his or her own desk. These study sessions included an individualized prompt to remind students they could earn a reward if their scores increased and to work hard on their own to improve their scores. Before and after their study sessions, all students took a 15-question test on mathematics estimation. The intervention was very brief (20 minutes). The effect size (ES = 0.655) favoring the triads groups was significant.

The studies by Janicki and Peterson (1981), Kramarski and Mevarech (2003), and Peklaj and Vodopivec (1999) all compared some form of cooperative group strategies, with no significant effects (see Table 5).

5. Strategies Combining Collaborative or Cooperative Learning With Other Approaches

Also meeting the criteria were three studies that examined cooperative learning strategies used in conjunction with other instructional practices. Because the cooperative learning elements were mixed with other modifications in practice, the Task Group was not able to isolate the effects of cooperative learning alone.

Only one of the studies yielded a statistically significant effect size, on a computation outcome measure. A quasi-experimental study by Busato et al. (1995) investigated Adaptive Instruction and Cooperative Learning (AGO),[1] a Dutch model that includes curricular adaptations, whole class instruction to introduce a topic, small group cooperation, regular assessments, individual work with the possibility of students helping each other, remedial groups working with direct guidance from a teacher, and whole class reflection. The study involved 572 middle school students in the Netherlands, and the intervention lasted for one month. The mathematical topic was pre-algebra ideas, and the interventions were AGO versus "a more traditional instructional method (mainly without group work)" (p. 671). The effect size for boys was significant (0.681) and for girls approached significance (0.583).

[1] AGO refers to the Dutch model called Adaptief Groeps-Onderwijs.

A quasi-experiment by Stevens and Slavin (1995) involved 1,012 elementary school students in Grades 2–6 in a year-long comparison of cooperative learning vs. whole class instruction. The nonsignificant effect on the California Achievement Test was negative, favoring the whole group instruction, on the applications test. The effect on the CAT computation (ES = 0.120) approached significance favoring the cooperative learning treatment. And finally, the Brenner et al. study (1997) discussed previously, small group instruction was compared to a control condition. The effects, favoring the small group condition, were not significant. The effect size for the pooled problem solving outcomes was 0.074.

6. Cooperative Learning Strategies in the Technology Context

Here the Task Group presents a review of the research that examined the use of cooperative learning strategies in the context of technology-based instruction. Use of some form of collaborative learning in computer-based instruction (CBI) is suggested by several positive reports from preschool to college (e.g., Leron & Lavy, 2004; Light & Blaye, 1990; Nastasi & Clements, 1994; Scardamalia et al., 1992; Schofield, 1995; Strommen, 1993). The following section discusses the eight studies identified that examined the use of cooperative learning strategies in the context of computers; the studies are summarized in Table 7. Only two of the studies produced significant effect sizes.

Seven studies investigated learning on computers in groups versus learning on computers individually (Hooper, 1992; Hooper, Temiyakam, & Williams, 1993; Mevarech et al., 1991; Mevarech, 1993, 1994; Slavin & Karweit, 1984; and Xin, 1999). All participating students were in elementary school. Durations ranged from very short (1 week) to an academic year. The outcomes measured in most studies were limited to computation, but the Hooper (1992, 1993), and the Slavin and Karweit (1984) studies also assessed mathematical concepts.

It was possible to calculate 15 different effect sizes across these studies. Only those that were significant are highlighted. Weiss et al. (2006) studied kindergarten students in Israel learning about numbers and operations. One treatment was a multimedia environment involving cooperation while the other was a multimedia environment involving an individual learning style. The outcome measure was a skills test on numbers and operations. The significant effect size (-0.862) favored the individual teaching style treatment. In the study by Xin (1999) that compared TAI involving CAI to whole-class computer-assisted instruction, the effect size on the Standard Achievement Test-Math for regular education students was significant (ES = 0.595), favoring the small group-based treatment. Effects for all but two studies were positive. A negative effect was found in the Weiss et al. (2006) study and for low- and high-achieving students in the Mevarech (1993) study. None of the other effect sizes calculated reached significance.

In summary, implications for policy and practice do not indicate a simple solution, such as, “Students should work together on computers.” Positive effects of cooperative learning in technological contexts can be obtained, but they may be limited in size, especially when using simple CAI programs, and may depend on teachers’ management and guidance of positive interactions and collaboration.

C. Summary and Conclusions

At this time a body of scientifically sound research does not exist that will quell the controversies about the best way to instruct students in mathematics. As evident from the review, however, educators are not completely in the dark about effective practice. Three principal findings emerge from the literature. First, some of the limitations in the studies reviewed in this section will be discussed.

Although all of the studies discussed here met the technical criteria for inclusion, there were a number of issues relative to the studies that could be addressed in future research. In some cases, for instance, the treatments were of such brief duration that it is difficult to interpret the conclusions. In many cases the control condition is inadequately specified leaving the reader to make assumptions about exactly what is being compared to what. In the case of teacher-directed and student-centered instruction, the terms are so vaguely defined to begin with that this is a serious problem in using the research literature to draw conclusions that may be useful for policy. Only some of these studies actually measure and document the nature of the intervention, leaving questions about the fidelity and extent of implementation and thus lack of clarity about what might be causing the results. This body of work tends to include examinations of specific groups of students, which is informative relative to specific groups, but needs to be balanced with studies that look at broader populations. Finally, in some cases the team evaluating the effectiveness of the intervention also invented the method (although it is emphasized that all studies included in this report met the stringent criteria for inclusion).

The review does allow us to make some key conclusions. First, Team Assisted Individualization (TAI), a cooperative learning strategy, has been shown to be effective in teaching computation skills. The finding does not extend to problem-solving skills or mathematical concepts. It is critical to note that the strategy involves much more than simply putting students into groups. Students first take diagnostic tests, and teachers utilize the results to prepare individualized sets of worksheets that target weak computation skills. Working in heterogeneous groups of four or five students, students are encouraged to work together to ensure that all students in the group attain mastery. Teachers work with small groups of students, pulled from different teams who are working on the same skill (e.g., division of decimals). Cooperation within groups is reinforced by group rewards given on the basis of final tests (for a more detailed description of TAI, see Slavin, Madden, and Leavey (1984) and Slavin and Karweit (1985)).

Why does TAI work? Researchers of TAI have argued that several elements of the technique may enhance learning: students receive immediate feedback from peers (as compared to delayed feedback from teachers during whole class instruction); materials present mathematical skills in a logical, hierarchical sequence; students' deficient areas are assessed, identified, and targeted with individualized materials; a group reward structure motivates students and encourages teamwork; the intervention blends teacher-directed and student-centered instruction. More research is needed to identify the precise mechanisms of TAI's effectiveness.

A second cooperative learning strategy, generally known as peer tutoring, also showed signs of promise, with a significant pooled effect size favoring the peer-assisted condition. This finding must be treated cautiously, in that it involves only two studies, is limited to the study of whole number operations by students in Grades 2-4, and only reaches statistical significance at the class level. In one of the studies, formative assessment was a key component of the intervention. Studies on which student level effects could be calculated did not produce statistically significant findings. As with TAI, the treatment is highly specific and involves far more than having students work in pairs. In both of these sets of studies, it is important to underscore that significant effect sizes were found only for computation or operations, not for mathematical concepts or problem solving.

The second main finding pertains to problem solving. Three studies were reviewed that documented successful far transfer of problem solving skills after extensive instruction. Muthukrishna and Borkowski (1995) studied third-graders who were taught a part-whole problem solving strategy in 14 lessons. Brenner et al. (1997) investigated middle school pre-algebra students learning how to represent function problems in multiple formats. The program consisted of 20 lessons. In both experiments, students in the guided discovery condition outperformed students in the traditional instruction condition on measures of far transfer. These effects were not statistically significant but they bordered on significance. Educators considering whether to implement these interventions would have to weigh the limitations of the outcomes—problem solving strategies that are restricted to particular topics in mathematics—with the amount of instructional time spent to attain them. Whether the benefits of guided discovery extend to content beyond the areas examined in these studies, or can be accomplished in less time, has not been studied.

In contrast, there were three studies (not counting the cooperative group studies) in which significant effects were found, favoring the teacher-directed instructional approach), for performance on computation outcome measures. Hopkins et al. (1997) found that the "didactic" treatment led to better performance by girls on an arithmetic computation test. And, in the Project Follow Through Evaluation, Stebbins et al. (1977) found significant effects favoring direct instruction, and nearly significant effects favoring the control treatment (in contrast to the Cognitively Oriented Curriculum) on the computation outcome measure. It is possible that under some conditions, with certain mathematical emphases and particular groups of students, the teacher-directed approaches can lead to better performance on computational assessments than more student-centered approaches.

That leads to the final principal finding. Much more research is needed that directly compares the effectiveness of student-centered and teacher-directed instruction, and that provides clear operational definitions for these terms. In particular, research is needed with teacher-centered instruction as the experimental condition. Almost all of the research reviewed here investigated experimental modes of instruction that are student-centered—whether guided discovery, cooperative learning, or peer tutoring—with the control condition described as "teacher-directed" or "traditional" or "direct instruction." Experiments with better specified teacher-directed interventions would enhance our understanding of how to improve classroom instruction in mathematics. A comprehensive program of research might succeed in transforming what has been a clash of ideologies into a search for effective practice.

Table 2a: Studies That Investigated the Effects of Team Assisted Individualization (TAI) on Computation Outcomes

Study	Design	Sample	Duration/ Content	Contrast	Measure	Subgroup	Hedge's g	Standard Error	
Computation Outcomes									
Student-level analyses									
Slavin, 1984—Study 1[b]	RCT	504 students in Grades 3–5, including 6% who were receiving special education services, in 18 mathematics classes in six schools in a suburban Maryland school district	8 weeks/ Addition, subtraction, multiplication, division, numeration, decimals, fractions, and word problems	TAI vs. whole class lectures and group paced instruction	CTBS—Computations	Overall	0.103	(ns)	0.460
Slavin, 1984—Study 2[b]	Quasi	375 students in Grades 4–6, including 27% who were receiving special education services, in 16 mathematics classes in four schools in a suburban Maryland school district	10 weeks/ Addition, subtraction, multiplication, division, numeration, decimals, fractions, and word problems	TAI vs. whole class lectures and group paced instruction	CTBS—Computations	Overall	0.109	(ns)	0.460
Slavin et al., 1984[b]	Quasi	1,371 students in Grades 3–5, including 8% that received special education services, in 59 mathematics classes in eight schools in a suburban Maryland school district	24 weeks/ Unspecified math curriculum (likely same topics as above)	TAI vs. whole class lectures and group paced instruction	CTBS—Computations	Overall	0.147	(ns)	0.331
Xin, 1999	RCT	118 third-grade students in six mathematics classes in three schools	Daily for one semester/ Basic fact families including addition, subtraction, multiplication, and division; coin recognition, place value, concepts, number patterns	TAI w/CAI vs. whole class w/CAI	Stanford Achievement Test—Math	Regular education	0.595	**	0.210
						Learning disability	0.338	(ns)	0.390
Heterogeneity									
Q-value	*df (Q)*	*P-value*	*I-squared*	*Pooled ES (four studies, five effect sizes)*			*Hedge's g*	*Standard Error*	
2.265	*4*	*0.271*	*0.687*				*0.377*	**	*0.145*
Classroom-level analyses									
Slavin & Karweit, 1985—Study 1	RCT	345 students in Grades 4–6 in 15 mathematics classes	18 weeks/ Addition, subtraction, multiplication, division, numeration, decimals, fractions, ratios, statistics, introduction to algebra, and word problems	TAI vs. Missouri Mathematics Program	CTBS—Computations	Overall	0.709	***	0.143
Slavin & Karweit, 1985—Study 2	RCT	480 students in Grades 3–5 in 22 mathematics classes in and around Hagerstown, MD	16 weeks/ Addition, subtraction, multiplication, division, numeration, decimals, fractions, ratios, statistics, introduction to algebra, and word problems	TAI vs. As-Is Control	CTBS—Computations	Overall	0.562	***	0.138
Heterogeneity									
Q-value	*df (Q)*	*P-value*	*I-squared*	*Pooled ES (two studies, two effect sizes)*			*Hedge's g*	*Standard Error*	
0.548	*1*	*0.459*	*0.000*				*0.633*	***	*0.099*

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

[a] Data were adjusted for clustering that occurred within classrooms.

[b] Data were adjusted for clustering that occurred within schools.

Table 2b: Studies That Investigated the Effects of Team Assisted Individualization (TAI) on Concepts Outcomes

Study	Design	Sample	Duration/ Content	Contrast	Measure	Subgroup	Hedge's g	Standard Error	
Concepts Outcomes									
Student-level analyses									
Slavin, 1984[b]	Quasi	1,371 students in Grades 3–5, including 8% that received special education services, in 59 mathematics classes in eight schools in a suburban Maryland school district	24 weeks/ Unspecified math curriculum (likely same topics as above)	TAI vs. whole class lectures and group paced instruction	CTBS—Concepts	Overall	0.098	(ns)	0.331
Classroom-level analyses									
Slavin & Karweit, 1985—Study 1	RCT	345 students in Grades 4–6 in 15 mathematics classes	18 weeks/ Addition, subtraction, multiplication, division, numeration, decimals, fractions, ratios, statistics, introduction to algebra, and word problems	TAI vs. Missouri Mathematics Program	CTBS—Concepts	Overall	-0.003	(ns)	0.139
Slavin & Karweit, 1985—Study 2	RCT	480 students in Grades 3–5 in 22 mathematics classes in and around Hagerstown, MD	16 weeks/ Addition, subtraction, multiplication, division, numeration, decimals, fractions, ratios, statistics, introduction to algebra, and word problems	TAI vs. As-Is Control	CTBS—Concepts	Overall	0.038	(ns)	0.134

Heterogeneity

Q-value	*df (Q)*	*P-value*	*I-squared*		*Hedge's g*	*Standard Error*	
0.045	*1*	*0.832*	*0.000*	*Pooled ES (two studies, two effect sizes)*	*0.018*	*(ns)*	*0.097*

~ p < .10, * p < .05, ** p < .01, *** p < .001
[a] Data were adjusted for clustering that occurred within classrooms.
[b] Data were adjusted for clustering that occurred within schools.

Table 3: Studies That Investigated the Effects of Student Teams-Achievement Divisions (STAD)

Study	Design	Sample	Duration/ Content	Contrast	Measure	Subgroup	Hedge's g	Standard Error	
Alkhateeb & Jumaa, 2002[a]	RCT	111 eighth-grade students in four classes in two schools in the United Arab Emirates	3 weeks/ Algebraic expressions	STAD vs. Whole Class Instruction	Algebra test	Overall	0.108	(ns)	0.479
Jacobs, 1996[a]	RCT	266 students in Grades 3–5 at a large private Christian/ fundamentalist elementary school in the Southeast	9 weeks/ Unspecified 3rd grade curricular unit	STAD vs. Direct Instruction with rewards and student individual accountability	Curriculum specific math test	Third grade	0.573	(ns)	0.492
						Fourth grade	0.484	(ns)	0.487
						Fifth grade	0.454	(ns)	0.486
Madden & Slavin, 1983[a]	RCT	183 third-, fifth-, and sixth-grade students, including 40 special education students, in six math classes in the Baltimore City schools	7 weeks/ Unspecified 3rd-, 5th-, and 6th-grade curricular units	STAD vs. Focused Instruction (whole class lectures, individual practice, quizzes, and individual recognition)	Curriculum specific math test	Overall	0.124	(ns)	0.402
Slavin & Karweit, 1984[a]	RCT	588 ninth-grade students in 25 math classes in 16 inner city Philadelphia junior and senior high schools	One school year/ Unspecified 9th-grade general math curriculum	STAD vs. Focused Instruction (students worked individually and did not receive team recognition)	CTBS, Shortened version (computation, concepts and applications subscales)	Overall	0.113	(ns)	0.221
Heterogeneity									
Q-value	*df (Q)*	*P-value*	*I-squared*	*Pooled ES (four studies, six effect sizes)*			*Hedge's g*	*Standard Error*	
1.384	*5*	*0.926*	*0.000*				*0.227*	(ns)	*0.152*

~ p < .10, * p < .05, ** p < .01, *** p < .001
[a] Data were adjusted for clustering that occurred within classrooms.
[b] Data were adjusted for clustering that occurred within schools.

Table 4: Studies That Investigated the Effects of Peer Assisted Learning

Study	Design	Sample	Duration/ Content	Contrast	Measure	Subgroup	Hedge's g	Standard Error	
Computation Outcomes									
Student-level analyses									
Fuchs et al., 2001[a]	RCT	168 Kindergarten students in 20 classes in five schools in a Southeastern metropolitan area	15 weeks/ Kindergarten core curriculum	PALS vs. Teacher-directed lessons and demonstrations	Stanford Early School Achievement Test	Special education	0.431	(ns)	0.524
						Low achieving	0.374	(ns)	0.454
						Medium achieving	0.436	(ns)	0.270
						High achieving	-0.162	(ns)	0.380
Fuchs et al., 2002[a]	RCT	327 first-grade students in 20 classrooms in a Southeastern metropolitan public school system	16 weeks/ Addition, subtraction, counting, sets, geometry, and measuring	PALS vs. As-is basal core curriculum	Stanford Achievement Test	Overall	0.055	(ns)	0.223
Ginsburg-Block & Fantuzzo, 1998	RCT	104 low-achieving third- and fourth-grade students in an urban elementary school	Two 30-minute sessions per week for 7 weeks/ Addition, subtraction, multiplication, and division computation and word problems	Peer collaboration dyads vs. Control	Curriculum based computation test	Overall	0.590	(ns)	0.389
Heterogeneity									
Q-value	*df (Q)*	*P-value*	*I-squared*	*Pooled ES (three studies, six effect sizes)*			*Hedge's g*	*Standard Error*	
3.361	*5*	*0.645*	*0.000*				*0.238*	~	*0.134*
Classroom-level analyses									
Fuchs et al., 1995	RCT	40 Grade 2–4 classrooms in nine elementary schools in a Southeastern, urban school district	Two 25-30 minute sessions per week for 23 weeks/ Grade level's annual operations curriculum	PALS integrated with regular assessments vs. Teacher-mediated instruction	Acquisition learning: Math Operations Test—Revised	Learning disability	0.260	(ns)	0.311
						Low achieving	0.728	**	0.320
						Average achieving	0.297	(ns)	0.312
Fuchs et al., 1997	RCT	40 Grade 2–4 classrooms in a Southeastern metropolitan public school system	18 weeks/ Number concepts, counting, word problems, charts/graphs, money, measurement, geometry, and computation	Peer-mediated instruction vs. Teacher-mediated instruction	Comprehensive Mathematics Test—Operations	Learning disabilities	0.663	~	0.386
						Low achieving	0.704	~	0.388
						Average achieving	0.177	(ns)	0.378
						High achieving	0.242	(ns)	0.378
Heterogeneity									
Q-value	*df (Q)*	*P-value*	*I-squared*	*Pooled ES (two studies, seven effect sizes)*			*Hedge's g*	*Standard Error*	
2.904	*6*	*0.821*	*0.000*				*0.431*	**	*0.132*

Continued on p. 6-37

Table 4, continued

Study	Design	Sample	Duration/ Content	Contrast	Measure	Subgroup	Hedge's g	Standard Error	
Concepts Outcomes									
Classroom-level analyses									
Fuchs et al., 1995	RCT	40 Grade 2–4 classrooms in nine elementary schools in a Southeastern, urban school district	Two 25–30 minute sessions per week for 23 weeks/ Grade level's annual operations curriculum	PALS integrated with regular assessments vs. Teacher-mediated instruction	Acquisition learning: Math Concepts and Applications	Learning disability	0.199	(ns)	0.311
						Low achieving	0.063	(ns)	0.310
						Average achieving	0.307	(ns)	0.312
Fuchs et al., 1997	RCT	40 Grade 2–4 classrooms in a Southeastern metropolitan public school system	18 weeks/ Number concepts, counting, word problems, charts/graphs, money, measurement, geometry, and computation	Peer-mediated instruction vs. Teacher-mediated instruction	Comprehensive Mathematics Test-Concepts	Learning disabilities	-0.016	(ns)	0.377
						Low achieving	0.515	(ns)	0.383
						Average achieving	0.139	(ns)	0.377
						High achieving	0.099	(ns)	0.377
Heterogeneity									
Q-value	*df (Q)*	*P-value*	*I-squared*	*Pooled ES (two studies, seven effect sizes)*			*Hedge's g*	*Standard Error*	
1.406	*6*	*0.965*	*0.000*				*0.186*	*(ns)*	*0.130*

~ p < .10, * p < .05, ** p < .01, *** p < .001
[a] Data were adjusted for clustering that occurred within classrooms.
[b] Data were adjusted for clustering that occurred within schools.

Table 5: Studies That Investigated Other Cooperative Learning Strategies

Study	Design	Sample	Duration/ Content	Contrast	Measure	Subgroup	Hedge's g	Standard Error	
Barron, 2000	RCT	96 sixth-grade students in a public magnet school for academically talented children	Four 1-hour sessions/ Contextual problem solving	Problem solving collaboratively in triads vs. Problem solving individually	Solutions to story problems—mastery	Overall	0.472	*	0.205
					Solutions to story problems—transfer	Overall	0.392	~	0.204
Hurley et al., 2005	RCT	78 African-American fifth-grade students in two urban public schools	One 20-minute session/ Math estimation	Triads worked together in a high-communal setting vs. Individuals worked in a low-communal setting	Math estimation task	Overall	0.655	**	0.230
Janicki & Peterson, 1981[a]	RCT	117 fourth- and fifth-grade students	2 weeks/ Fractions	Small group direct instruction vs. Individual direct instruction	Researcher developed test on fractions	Overall	-0.041	(ns)	0.188
Kramarski & Mevarech, 2003[a]	RCT	384 eighth-grade students in 12 classrooms in four Israeli junior high schools	2 weeks/ Linear graphing	Metacognitive and cooperative groups vs. Metacognitive and individual	Graph interpretation test	Overall	0.355	(ns)	0.387
				Cooperative groups vs. Individual work	Graph interpretation test	Overall	0.105	(ns)	0.388
Peklaj & Vodopivec, 1999[a]	RCT	373 fifth-grade students in 15 classes from nine primary schools in Slovenia	One lesson per week for seven months/ Basic concepts, measure transformation, calculations, problem solving	Cooperative learning vs. Individual work	Teacher developed math test	Overall	0.317	(ns)	0.251

~ p < .10, * p < .05, ** p < .01, *** p < .001
[a] Data were adjusted for clustering that occurred within classrooms.
[b] Data were adjusted for clustering that occurred within schools.

Table 6: Studies That Investigated Multiple Strategies—Cooperative Learning Combined With Other Instructional Practices

Study	Design	Sample	Duration/Content	Contrast	Measure	Subgroup	Hedge's g	Standard Error	
Brenner et al., 1997[a]	RCT	128 students in six intact pre-algebra classes in three junior high schools in southern California	20 days/ Pre-algebra ideas such as the functional relationship between two variables and contextual translation and application	Anchored instruction using small groups vs. Control	Pooled problem solving outcomes: word problem solving (ES = 0.110), function word problem (ES = 0.393), and equation solving (ES = -0.281) measures	Overall	0.074	(ns)	0.399
Busato et al., 1995[a]	Quasi	572 middle school students in 23 classes in six schools in the Netherlands	Unspecified duration/ Existing Dutch math curriculum	AGO model curriculum vs. Traditional curricula	Test of math achievement	Boys	0.681	*	0.227
						Girls	0.583	~	0.235
Stevens & Slavin, 1995[c]	Quasi	1,012 elementary students in Grades 2–6 in five schools in a suburban Maryland school district. Two of the schools were cooperative elementary schools and three schools were more traditional elementary schools	1 year/ Math computation	Cooperative learning school vs. Whole class instruction	California Achievement Test—CAT computations	Overall	0.120	~	0.068
					California Achievement Test—CAT—application	Overall	-0.050	(ns)	0.068

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

[a] Data were adjusted for clustering that occurred within classrooms.

[b] Data were adjusted for clustering that occurred within schools.

[c] To be more comparable with other studies, the data presented for this study are data only after the first year, although two years of data were available.

Table 7: Studies That Investigated Cooperative Learning Strategies in the Context of Computers

Study	Design	Sample	Duration/Content	Contrast	Measure	Subgroup	Hedge's g	Standard Error	
Hooper, 1992	RCT	115 fifth- and sixth-grade average or high-ability students from a suburban middle school	1 week/ Calculation of number of sides of a three-dimensional object and classifying objects as examples or nonexamples of a concept	Cooperative learning with computer based instruction (CBI) vs. Individual learning with CBI	Math test including fact, application, generalization, and problem-solving questions	Overall	0.335	~	0.201
Hooper, 1993	RCT	175 fourth-grade average or high-ability students from six classrooms in a suburban middle school	3 weeks/ Calculations using the four basic arithmetic operations using symbols to represent constants and operations	Cooperative learning with computer based instruction (CBI) vs. Individual learning with CBI	Math test including fact, application, generalization, and problem-solving questions	Overall	0.305	~	0.157
Mevarech et al., 1991	RCT	149 sixth-grade students in five classrooms in one Israeli school	One trimester/ Basic operations with positive integers and fractions	Pairs w/CAI vs. Individual w/CAI	TOAM achievement (computerized diagnostic)	Low achieving	0.266	(ns)	0.285
						Medium achieving	0.268	(ns)	0.285
						High achieving	0.120	(ns)	0.289
Mevarech, 1993[a]	RCT	110 third-grade students in two Israeli public schools	One trimester/ Traditional 3rd grade curriculum in arithmetic	Pairs w/CAI vs. Individual w/CAI	Arithmetic achievement test	Low achieving	-0.028	(ns)	0.479
						High achieving	-0.506	(ns)	0.480
Mevarech, 1994[a]	RCT	344 third-grade and 279 sixth-grade students in 19 classrooms in five schools in a suburb of Tel Aviv, Israel	Two 20-minute sessions per week for one academic year/ Basic skills with mathematics operations, comprehension of numerical systems, understanding mathematical rules, solving word problems	Integrated learning system (ILS) in homogenous pairs vs. ILS individually	ILS diagnostic test	3rd grade Low Achievers	0.287	(ns)	0.322
						3rd grade High Achievers	0.123	(ns)	0.319
						6th grade Low Achievers	0.266	(ns)	0.337
						6th grade High Achievers	0.348	(ns)	0.336
Slavin & Karweit, 1984[a]	RCT	588 ninth-grade students in 25 math classes in 16 inner city Philadelphia junior and senior high schools	One school year/ Unspecified 9th grade general math curriculum	STAD vs. Focused Instruction (students worked individually and did not receive team recognition)	CTBS, Shortened version (computation, concepts and applications subscales)	Overall	0.113	(ns)	0.221

Continued on p. 6-40

Table 7, continued

Study	Design	Sample	Duration/Content	Contrast	Measure	Subgroup	Hedge's g	Standard Error	
Weiss et al., 2006	RCT	116 students in six Kindergarten classes of medium to high SES in Israel	28 hours over 5 months/ Mathematical skills about numbers and operations from 1 to 10	Multimedia environment in a cooperative learning teaching style vs. Multimedia environment in an individual learning teaching style	Skills test on numbers and operations	Overall	-0.862	***	0.238
Xin, 1999	RCT	118 3rd-grade students in six mathematics classes in three schools	Daily for one semester/ Basic fact families including addition, subtraction, multiplication, and division; coin recognition, place value, concepts, number patterns	TAI w/ CAI vs. whole class w/ CAI	Stanford Achievement Test-Math	Regular education	0.595	**	0.210
						Learning disability	0.338	(ns)	0.390
Heterogeneity									
Q-value	*df (Q)*	*P-value*	*I-squared*	*Pooled ES (8 studies, 15 effect sizes)*			*Hedge's g*	*Standard Error*	
27.457	*14*	*0.017*	*49.011*				*0.157*	*0.101*	

~ p < .10, * p < .05, ** p < .01, *** p < .001
[a] Data were adjusted for clustering that occurred within classrooms.
[b] Data were adjusted for clustering that occurred within schools.

BIBLIOGRAPHY

Alkhateeb, H.M., & Jumaa, M. (2002). Cooperative learning and algebra performance of eighth-grade students in United Arab Emirates. *Psychological Reports, 90,* 91–100.

Barron, B. (2000). Problem solving in video-based microworlds: Collaborative and individual outcomes of high-achieving sixth-grade students. *Journal of Educational Psychology, 92*(32), 391–398.

Beady, C.L., Slavin, R.E., & Fennessey, G.M. (1981). Alternative student evaluation structures and a focused schedule of instruction in an inner-city junior high school. *Journal of Educational Psychology*, *73,* 518–523.

Becker, W.C. (1977). Teaching reading and language to the disadvantaged—What we have learned from field research. *Harvard Educational Review*, *47*, 18–543.

Bell, A.W., O'Brien, D., & Shiu, C. (1980). Designing teaching in the light of research on understanding. In R. Karplus (Ed.), *Proceedings of the fourth international conference for the psychology of mathematics education*, Berkeley, CA: The International Group for the Psychology of Mathematics. (ERIC Document Reproduction Service No. ED 250 186).

Bell, A.W. (1982). Diagnosing students' misconceptions. *The Australian Mathematics Teacher, 1*, 6–10.

Brenner, M., Mayer, R., Moseley, B., Brar, T., Duran, R., Reed, B.S., & Webb, D. (1997). Learning by understanding: The role of multiple representations in learning algebra. *American Educational Research Journal, 34(4*), 663–689.

Brophy, J.E., & Good, T.L. (1986). Teacher behavior and student achievement. In M.C. Wittrock (Ed.), *Handbook of research on teaching*. New York: Macmillan.

Busato, V., Tendam, G., Vandeneeden, P., & Terwel, J. (1995). Gender-related effects of cooperative learning in a mathematics curriculum for 12–16 year olds. *Journal of Curriculum Studies, 27*(*6*), 667–686.

Cicchelli, T. (1982). Effects of direct and indirect instruction patterns and prior achievement on post course achievement. *Journal of Instructional Psychology, 9(4),* 176–189.

Cobb, P. (2007). Putting philosophy to work: Coping with multiple theoretical perspectives. In F. Lester (Ed.), *Second handbook of research on mathematics teaching and learning (Vol. 1, pp. 3–38).* Greenwich, CT: Information Age Publishing.

Dewey, J. (1897). My pedagogic creed. *School Journal*, *54,* 77–80.

Flanders, N.A. (1970). *Analyzing teaching behavior*. Reading, MA: Addison-Wesley Publishing Company.

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

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 & Practice, 17*(2), 90–106.

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

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

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

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

Gersten, R., & Carnine, D. (1984). Direct instruction mathematics: A longitudinal evaluation of low-income elementary school students. *Elementary School Journal, 84(4)*, 395–407.

Ginsburg-Block, M. & Fantuzzo, J. (1998). An evaluation of the relative effectiveness of NCTM standards-based interventions for low-achieving urban elementary students. *Journal of Educational Psychology, 90(3)*, 560–569.

Good, T.L., & Grouws, D.A. (1977). Teaching effects: A process-product study in fourth-grade mathematics classrooms. *Journal of Teacher Education, 28(3)*, 49–54.

Good, T.L., & Grouws, D.A. (1979). The Missouri mathematics effectiveness project. *Journal of Educational Psychology, 71(3)*, 355–362.

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

Good, T.L., Mulryan, C., & McCaslin, M. (1992). Grouping for instruction in mathematics: A call for programmatic research on small-group processes. In D. A. Grouws (Ed), *Handbook of research on mathematics teaching and learning*. New York: Macmillan.

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

Hooper, S., Temiyakarn, C., & Williams, M.D. (1993). The effects of cooperative learning and learner control on high and average ability students. *Educational Technology Research and Development, 41(2)*, 5–18.

Hopkins, K.B., McGillicuddy-De Lisi, A.V., & De Lisi, R. (1997). Student gender and teaching methods as sources of variability in children's computational arithmetic performance. *The Journal of Genetic Psychology, 158(3)*, 333–345.

Hurley, E.A., Boykin, A.W., & Allen, B.A. (2005). Communal versus individual learning of a math-estimation task: African American children and the culture of learning contexts. *Journal of Psychology, 139(6)*, 513–527.

Jacobs, J.A. (1996). Gender inequality and higher education. *Annual Review of Sociology, 22(1)*, 153–185.

Janicki, T.C. & Peterson, P.L. (1981). Aptitude-treatment interaction effects of variations in direct instruction. A*merican Educational Research Journal, 18(1)*, 63–82.

Kame'enui, E., Carnine, D., Darch, C., & Stein, M. (1986). Two approaches to the development phase mathematics instruction. *Elementary School Journal, 86*, 633–650.

Kramarski, B. & Mevarech, Z.R. (2003). Enhancing mathematical reasoning in the classroom: The effects of cooperative learning and metacognitive training. *American Educational Research Journal, 40(1)*, 281–310.

Ladson-Billings, G. (1995). Toward a theory of culturally relevant pedagogy. *American Education Research Journal, 35*, 465–491.

Lampert, M. (1990). When the problem is not the question and the solution is not the answer: Mathematical knowing and teaching. *American Educational Research Journal, 27*, 29–63.

Leron, U., & Lavy, I. (2004). The emergence of mathematical collaboration in an interactive computer environment. *International Journal of Computers for Mathematical Learning, 9*, 1–23.

Light, P., & Blaye, A. (1990). Computer-based learning: The social dimensions. In H. C. Foot, M. J. Morgan & R. H. Shute (Eds.), *Children helping children* (pp. 135–147). New York: John Wiley & Sons.

Madden, N.A., & Slavin, R.E. (1983). Effects of cooperative learning on the social acceptance of mainstreamed academically handicapped students. *Journal of Special Education, 17(2)*, 171–182.

Mevarech, Z.R. (1993). Who benefits from cooperative computer-assisted instruction? *Journal of Educational and Computing Research. 9*(4), 451–461.

Mevarech, Z.R. (1994). The effectiveness of individualized versus cooperative computer-based integrated learning systems. *International Journal of Educational Research*. *86* (4), 39–52.

Mevarech, Z.R., Silber, O., & Fine, D. (1991). Learning with computers in small groups: Cognitive and affective outcomes. *Journal of Educational and Computing Research. 7*(2), 233–243.

Meyer, L.A., Gersten, R., & Gutkin, J. (1983). Direct instruction: A project follow through success story in an inner-city school. *The Elementary School Journal, 84*, 241–252.

Muthukrishna, N., & Borkowski, J.G. (1995). How learning contexts facilitate strategy transfer. *Applied Cognitive Psychology, 9*, 425–446.

Nastasi, B.K., & Clements, D.H. (1994). Effectance motivation, perceived scholastic competence, and higher-order thinking in two cooperative computer environments. *Journal of Educational Computing Research, 10(3)*, 249–275.

National Council of Teachers of Mathematics (NCTM) (1989). *Curriculum and evaluation standards for school mathematics.* Reston, VA: National Council of Teachers of Mathematics.

National Research Council (1989). *Everybody counts: A report to the nation on the future of mathematics education.* Washington, DC: National Academy Press.

National Research Council (2000). *How people learn: Brain, mind, experience, and school.* Washington, DC: National Academy Press.

National Research Council. (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, DC: National Academy Press.

Nickerson, R.S. (1988). On improving thinking through instruction. In E.Z. Rothkopf (Ed.), *Review of research in education, 15,* 3–57. Washington, DC: American Educational Research Association.

Palincsar, A.S., & Brown, A.L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction, 1(2)*, 117–175.

Peklaj, C., & Vodopivec, B. (1999). Effects of cooperative versus individualistic learning on cognitive, affective, metacognitive and social processes in students. *European Journal of Psychology of Education, 14(3)*, 359–373.

Rittle-Johnson B. (2006). Promoting transfer: Effects of self-explanation and direct instruction. *Child Development, 77(1)*, 1–15.

Rudnitsky, A., Etheredge, S., Freeman, S.J.M., & Gilbert, T. (1995). Learning to solve addition and subtraction word problems through a structure-plus-writing approach. *Journal for Research in Mathematics Education, 26(5)*, 467–486.

Scardamalia, M., Bereiter, C., Brett, C., Burtis, P.J., Calhoun, C., & Lea, N.S. (1992). Educational applications of a networked communal database. *Interactive Learning Environments, 2(1)*, 45–71.

Schofield, J.W. (1995). *Computers and classroom culture*. Cambridge, MA: Cambridge University Press.

Slavin, R.E. (1980). Effects of student teams and peer tutoring on academic achievement and time on-task. *Journal of Experimental Education, 48*, 252–257.

Slavin, R.E. (1991). Synthesis of research on cooperative learning. *Educational Leadership, 48*, 71–82.

Slavin, R.E., & Karweit, N.L. (1984). Mastery learning and student teams: A factorial experiment in urban general mathematics classes. *American Educational Research Journal, 21(4)*, 725–736.

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

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

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

Stallings, J., (1975). Implementation and child effects of teaching practices in follow through classrooms. *Monographs of the Society for Research in Child Development, 40(7/8)*, 1–133.

Stebbins, L., St. Pierre, R.G., Proper, E.L., Anderson, R.B., & Cerva, T.R. (1977). *Education as experimentation: A planned variation model (Vol. IV A–D)*. Cambridge, MA: Abt Associates.

Stevens, R.J., & Slavin, Robert E. (1995). The cooperative elementary school: Effects on students' achievement, attitudes, and social relations. *American Educational Research Journal*, *32(2)*, 321–351.

Strommen, E.F. (1993). Does yours eat leaves? Cooperative learning in an educational software task. *Journal of Computing in Childhood Education, 4(1)*, 45–56.

U.S. Department of Labor, Bureau of Labor Statistics. (2008). Teachers—Preschool, kindergarten, elementary, middle, and secondary. In *Occupational Outlook Handbook.* Retrieved on December 1, 2007, from http://www.bls.gov/oco/oco069.htm#nature.

Weiss, I., Kramarski, B., & Talis, S. (2006). Effects of multimedia environments on kindergarten children's mathematical achievements and style of learning. *Educational Media International*, *43(1)*, 3–17.

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

III. Effective Instruction for Students With Learning Challenges: A Meta-Analytic Review

Between 5% and 10% of students will experience a serious learning disability in mathematics before completing high school (Barbaresi, Katusic, Colligan, Weaver, & Jacobsen, 2005). Many more will have difficulties in learning mathematics at an acceptable level of proficiency. In this section, the Instructional Practices Task Group addresses the rigorous research on instructional methods that can help these students. An overview of the methodological procedures for the Task Group is provided in Appendix A. Throughout this section, the Task Group used these meta-analytic techniques as noted in the methodology statement. Because of the wide array of instructional approaches explored in the research for this section, multiple meta-analyses were performed to analyze this research.

The Task Group chose to review studies of students with LD separately from studies of low-achieving students because the problems experienced by students with LD are consistently more severe than those experienced by other low-performing students (Fuchs, Fuchs, Mathes, & Lipsey, 2000; Murphy, Mazzocco, Hanich, & Early, 2007). Therefore, educators cannot necessarily assume that techniques that are effective for students with learning disabilities are the most effective or efficient means for teaching struggling students. However, the reader will note that many of the same themes and issues recur across these two bodies of research.

A. Characteristics of Students With Learning Disabilities in Mathematics

Most of the research on the nature of learning disabilities in mathematics has been conducted with younger students and typically involves understanding their gaps in whole number arithmetic. Certain findings have been consistently replicated. Because students with LD display problems in so many areas of mathematics, pinpointing the exact nature of the cognitive difficulty has been an intricate process (Geary, 2003).

However, there are several problems that seem particularly chronic. The first is efficient retrieval of basic arithmetic combinations (mathematics facts) (Jordan, Hanich & Kaplan, 2003). A second is delayed adoption of efficient counting strategies. Students with learning disabilities will tend to count on their fingers well after their peers have outgrown this approach and when forbidden by their teachers they may count with the help of visual placemarkers in the classroom (e.g., stripes on the ceiling or the radiator), or give up in frustration. Most typically developing students learn, prior to entering school, what is commonly called a "counting-on strategy." They learn that if they have to add 7 to 2, this process is mathematically equivalent to adding 2 to 7, and that is much more efficient to make this transformation (i.e., that the most efficient way to find 2 + 7 is to start with the 7 in a mental number line and count up 2, rather than start with the 2 and count up 7). In contrast, students with LD will tend to start at 2 and count up using 7 figures or objects. Thus, they are more likely to make errors by using the tedious procedure. Furthermore, even if their answer is accurate, their strategy for reaching this answer is far from efficient.

It also appears that students with learning disabilities have a very limited working memory (Geary, Hoard, Byrd-Craven, Nugent, & Numtee, 2007; McLean & Hitch, 1999; Swanson & Sachse-Lee, 2001), which affects their ability to keep abstract information in their minds for the purpose of counting and solving specific problems. Finally, students with learning disabilities seem to display problems in many aspects of basic number sense such as comparing magnitudes of numbers by quickly visualizing a number line or transforming simple word problems into simple equations (Jordan, Hanich, & Kaplan, 2003; Fuchs, Compton, Fuchs, Paulsen, Bryant, & Hamlett, 2005). In addition, two studies (DiPerna, Lei & Reid, 2007; Fuchs, 2005) have both found that teachers' ratings of a child's attention span and task persistence, both areas that are often difficult for students with LD, are good indicators of subsequent problems in learning mathematics.

B. Students With Low Achievement in Mathematics

Many more students struggle to learn mathematics than the 5 to 10% who appear to possess a learning disability in mathematics (Badian, 1983; Fuchs et al., 2005; Gross-Tsur et al., 1996; Lewis et al., 1994). Although there are numerous disputes about how to best define and operationalize the general term "learning disabilities," and the more specific term "math disabilities," there is some emerging consensus (Bradley, Danielson, & Hallahan, 2002).

In contrast, there is no consensus as to how to operationalize the term "low achieving," other than students whose performance in mathematics is below grade level expectations. In some cases, e.g., Cardelle-Elawar (1995), all students in a low-income, low-achieving school are considered low achieving. Other studies (e.g., Moore & Carnine, 1989) only select students who perform poorly on a screening test that addresses the topic of the intervention research study.

Several dilemmas and constraints presented themselves when considering studies of low-achieving students not presented in working with studies of students with LD. First, there is no agreed upon operational definition of what is meant by a student struggling to learn mathematics or a low-achieving or "at risk" student (Mazzocco, 2007). Indeed, there is no measurable boundary or cut-off criterion, based on standardized test performance, for considering a student to be math disabled versus experiencing low achievement in mathematics.

Meanwhile, the factors that contribute to the low-achieving designation seem to include some unknown combination of the following:

- deficiencies with previous mathematics instruction and mathematics teachers with limited knowledge of the subject (Sowder, Philipp, Armstrong, & Schappelle, 1998);
- limited experiences at home that informally teach familiarity with number concepts, build and reinforce procedural facility and demonstrate relevance of mathematics to everyday problems (e.g., Griffin, Case, & Siegler, 1994);
- problems with sustaining attention to academic tasks and activities (Fuchs, Compton, Fuchs, Paulson, Bryant, & Hamlett, 2005; DiPerna, Lei, & Reid, 2007; Kolligian & Sternberg, 1987); and,
- weak motivation and maladaptive attribution style (Torgesen, 1994).

C. A Meta-Analytic Review of Research With Students With LD and LA in Mathematics (1976–2007)

There is a dramatically smaller body of research on mathematics instruction compared to reading instruction for 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. However, this was a dramatic improvement over the ratio of 16:1 in the prior decade.

Despite the limited knowledge of the precise nature of learning disabilities in mathematics, especially in areas such as rational numbers, geometry and pre-algebra, researchers have attempted to develop interventions that can teach students with LD. In fact, in the Panel's literature search, the number of high-quality studies examining the effectiveness of various instructional practices for teaching students with LD far surpasses the number of studies conducted with typically developing students.

The Task Group speculates that there are several reasons for this phenomenon. One important factor has been the consistent support for research in the field of special education; annual research budgets for special education often surpassed budgets for research on the education of nondisabled students in academic areas. Even more importantly, the Office of Special Education Programs (OSEP) at the U.S. Department of Education has consistently supported experimental research since the late 1960s.

In addition to studies of students with LD, the Task Group was also able to locate a small number of studies with low-achieving students that met the criteria for rigorous experimental or quasi-experimental research. The children in these studies were defined as either at-risk or experiencing mathematics difficulties based on performance on a screening measure of mathematics skills or teacher ratings or recommendation. None of the participants was formally diagnosed as math disabled by the researchers.

Thus, a reasonable set of studies exists that investigate the effectiveness of various instructional approaches for teaching students with LD using rigorous experimental or quasi-experimental designs, and a smaller, but still adequate set of studies exists that examine various approaches for teaching students who experience difficulties in mathematics, but were not classified as possessing a learning disability. To conduct a meta-analysis, the Task Group needed to center the analysis around a key research question. The question most consistently posed in the research studies reviewed concerned the effectiveness of explicit systematic instruction on the mathematics performance of this group of students.

D. The Nature of This Report

This document summarizes the meta-analyses conducted for the Panel's Instructional Practices Task Group on the nature of effective mathematics instruction for students with LD and for other low-achieving students. To organize these meta-analyses, the Task Group clustered studies into four categories:

- Studies of the impact of systematic *explicit instruction* on the performance of students with LD in mathematics

- Studies of the impact of systematic *explicit instruction* on the performance of low-achieving students in mathematics

- Other approaches for teaching students with LD
    - — Selection of examples to foster development of more sophisticated strategies for quick retrieval of basic arithmetic facts
    - — Use of visual representations as a key component of instruction
    - — Instruction that encouraged students to think aloud

- Other approaches for teaching low-achieving students that are primarily implicit

The following sections provide study characteristics for each of the studies identified and effect sizes for the Category 1 (high-quality) studies on posttest measures and transfer measures (when available). The effect sizes are pooled for the studies that examined the effects of using explicit instruction for students with LD using common meta-analytic standards. Effect sizes for studies in the remaining categories were not pooled because the interventions varied greatly across studies.

All effect sizes have been adjusted for clustering, when appropriate. The Task Group used the U.S. Department of Education's What Works Clearinghouse default Intra-Class Correlation of .20 for the adjustment. For further details on data analysis, see the footnotes accompanying the tables and the Methodological Procedures section in Appendix A.

E. Explicit Strategies Used for Students With Learning Disabilities

Explicit instruction involves teacher-demonstrated step-by-step plans for solving a problem. The teacher demonstrates a specific plan for a set of problems (as opposed to a general problem-solving heuristic strategy) and students are asked to use the same procedures or steps demonstrated by the teacher to solve the problem. For example, Xin, Jitendra, and Deatline-Buchman (2005) provided explicit instruction for using strategies for identifying and solving various word problem types. Students were given prompt sheets for identifying salient features of the word problem types. Students then were taught to map these features onto a schema diagram that represented the problem structure. Next, students used the schema diagram to formulate the appropriate mathematical equation for solving the problem. Each of these steps toward problem solution were explicitly taught strategies for problem solution.

There were nine studies that looked at the effect of explicit strategies for students with LD, met the inclusion criteria, and were methodologically adequate. Results from these studies are presented in three tables. Each table includes the findings for a separate mathematical outcome: word problem solving, computation, or transfer of learning. Table 8 below presents the results from the six studies that investigated the effects of using explicit strategies on improving word problem-solving outcomes with students with learning disabilities. Table 9 presents results of the three high-quality studies that looked at computation outcomes. Table 10 presents the results of the four studies that included a measure of generalization of training. All tables include the pooled effect size, tests for heterogeneity, and tests for statistical significance for the pooled effect size. The pooled effects reveal significant effects of explicit instruction on solving word problem solving (ES = 1.152), computation (ES = 1.285), and transfer (ES = 0.777).

Table 8: Studies That Investigate Explicit Strategies With Students With Learning Disabilities: Word Problem Outcomes

Study	Design	Sample	Duration/Content	Contrast	Measure	Hedge's g		Standard Error
Word Problem Outcomes								
Hutchinson, 1993	RCT	20 LD students in Grades 8–10 from two junior high schools in suburban Vancouver, Canada	40-minute sessions on alternate days for 4 months/ Algebraic word problem solving (relational problems, proportion problems, & two-variable, two-equation problems)	Metacognitive and solution strategies vs. Regular resource class instruction	Pooled B.C. Mathematics Achievement Test (ES = 0.705), Q2 B.C. Achievement Test (ES = 1.724)	1.215	*	0.484
Jitendra et al., 1998	RCT	34 students in Grades 2–5 from four public-schools in the U.S. 25 students were classified as having mild disabilities (LD, educable mentally retarded, or seriously emotionally disturbed), and the remaining nine students had difficulty in math	17–20 40–45-minute sessions/ Addition and subtraction word problems (including change, group, and compare problems)	Explicit step-by-step strategy vs. Traditional basal strategy	Researcher designed word problem solving criterion test	0.557	(ns)	0.342
Milo et al., 2005[a]	Quasi	36 LD students in three special primary schools. Average age of students was 9.10 years	Two weekly lessons for half the year during regular math class/ Addition and subtraction	Directing vs. Guiding instruction	Addition and subtraction word problem test based on problems from the databank of the National Institute for Educational Measurement	0.303	(ns)	0.447
Owen & Fuchs, 2002[a]	Quasi	24 third-grade students with IEPs from 14 classrooms in six schools (20 students had LD, one had MMR, two had speech disorders, and one had ADHD)	Six lessons/ Word problems that involve finding "halves"	Full-dose acquisition and transfer vs. Traditional	Researcher designed word problem solving test	3.385	***	0.888
Wilson & Sindelar, 1991[b]	RCT	62 LD students from nine elementary schools in a medium-sized school district in northern Florida	Fourteen 30-min lessons over 3 weeks/ Addition and subtraction word problems (four types of two- to three-sentence problems)	Strategy plus sequence vs. Sequence only	Researcher designed word problem test	0.782	~	0.470
Xin et al., 2005	RCT	22 middle school students in one school (18 students had LD, one had severe emotional disorders, and three were at-risk for math failure)	12 1-hour sessions/ Multiplicative compare and proportion problems and mixed word problems	Schema-based instruction (SBI) vs. General strategy instruction (GSI)	Researcher designed word problem solving criterion test	1.866	***	0.497
Heterogeneity								
Q-value	*df (Q)*	*P-value*	*I-squared*	*Pooled ES (five studies, five effect sizes)*		*Hedge's g*	*Standard Error*	
14.754	*5*	*0.011*	*66.110*			*1.152*	***	*0.341*

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

[a] Data were adjusted for clustering that occurred within classrooms.

[b] Data were adjusted for clustering that occurred within schools.

Table 9: Studies That Investigate Explicit Strategies With Students With Learning Disabilities: Computation Outcomes

Study	Design	Sample	Duration/Content	Contrast	Measure	Hedge's g	Standard Error	
Computation Outcomes								
Schopman & Van Luit, 1996	Quasi	60 students between the ages of 5 and 7 attending schools for special education (primarily LD) who scored less than 45% correct on a test for number sense (likely in the Netherlands)	Thirteen lessons in 3 months/ Preparatory arithmetic skills (number sense, counting skills, and Piagetian operations)	Directing and guiding vs. Control	Utrech test of number sense	1.023	***	0.286
Tournaki, 2003	RCT	42 LD second-grade students attending self contained special education classes in one school in New York	Eight 15-minute sessions on consecutive school days/ Algebra (three problem types: relational problems, proportion problems, & two-variable, two-equation problems)	Strategy instruction vs. Drill and practice	Researcher designed computation test	1.612	***	0.426
Van Luit & Naglieri, 1999[a]	Quasi	42 9–11-year-old LD students from two schools for special education in the Netherlands	Three 45-minute sessions per week for 17 weeks/ Multiplication and division problems	MASTER program (Mathematics Strategy Training for Educational Remediation) vs. Standard instruction	Researcher designed mathematics achievement test	2.174	*	0.991
Heterogeneity								
Q-value	*df (Q)*	*P-value*	*I-squared*	*Pooled ES (four studies, four effect sizes)*		*Hedge's g*	*Standard Error*	
2.221	*2*	*0.329*	*9.956*			*1.285*	***	*0.256*

~ p < .10, * p < .05, ** p < .01, *** p < .001
[a] Data were adjusted for clustering that occurred within classrooms.

Table 10: Studies That Investigate Explicit Strategies With Students With Learning Disabilities: Transfer Outcomes

Study	Design	Sample	Duration/ Content	Contrast	Measure	Hedge's g	Standard Error	
Transfer Outcomes								
Jitendra et al., 1998	RCT	34 students in Grades 2–5 from four public-schools in the U.S. 25 students were classified as having mild disabilities (LD, educable mentally retarded, or seriously emotionally disturbed), and the remaining nine students had difficulty in math	17–20 40–45-minute sessions/ Addition and subtraction word problems (including change, group, and compare problems)	Explicit step-by-step strategy vs. Traditional basal strategy	Researcher designed generalization test	1.010	**	0.357
Milo et al., 2006[a]	Quasi	36 LD students in three special primary schools. Average age of students was 9.10 years	Two weekly lessons for half the year during regular math class/ Addition and subtraction	Directing vs. Guiding instruction	Researcher designed transfer test	-0.073	(ns)	0.445
Tournaki, 2003	RCT	42 LD 2nd-grade students attending self contained special education classes in one school in New York	Eight 15-minute sessions on consecutive school days/ Algebra (3 problem types: relational problems, proportion problems, & two-variable, two-equation problems)	Strategy instruction vs. Drill and practice	Researcher designed transfer test	0.801	*	0.382
Xin et al., 2005	RCT	22 middle school students in one school (18 students had LD, one had severe emotional disorders, and three were at-risk for math failure)	3–4 times per week, for a total of 12 1-hour sessions/ Multiplicative compare and proportion problems and mixed word problems	Schema-based instruction (SBI) vs. General strategy instruction (GSI)	Researcher designed generalization test	1.334	**	0.467
Heterogeneity								
Q-value	*df (Q)*	*P-value*	*I-squared*	*Pooled ES (four studies, four effect sizes)*		*Hedge's g*	*Standard Error*	
5.496	*3*	*0.139*	*45.416*			*0.777*	**	*0.277*

~ p < .10, * p < .05, ** p < .01, *** p < .001
[a] Data were adjusted for clustering that occurred within classrooms.

From these results, one can infer that explicit instruction is an effective means for building performance in problem solving, computational proficiency, and ability to transfer from items on which students received training to items on which students had not received training, for students with LD.

However, the number of high-quality studies is small, and one would not want to overgeneralize from a set of nine studies that, taken together, are limited by a restricted range of study characteristics. For example, many of the studies were of short or moderate duration. Although the set of studies represents a wide range of age levels (seven of the studies examine elementary schools, while two studies examined middle schools) there are a sparse number of studies for any given age level or any given mathematical topic. The studies reviewed almost exclusively used researcher-developed measures, which tend to yield higher effect sizes than norm-referenced measures of generalized mathematics proficiency (Swanson & Hoskyn, 1998).

1. The Evolving Nature of Explicit Systematic Strategy Instruction

Nonetheless, the positive and significant pooled effect sizes for the studies that investigate the effect of explicit systematic instruction study results on word problems, computation, and transfer outcomes suggests that explicit systematic instruction is a desirable approach for at least some critical aspects of mathematics instruction for students with LD. The question becomes, what exactly is explicit systematic instruction? There is no easy answer to this question. In fact, like most educational labels, this term means very different things to different individuals. In addition, the nature of explicit systematic instruction has evolved over time.

Probably the earliest use of this term (at least during the past four decades) was the pioneering work of Bereiter and Engelmann (1966) in providing preschoolers from low-income families with explicit systematic instruction in number concepts, counting, phonological awareness, and the more formal structure of the language used in school. By the 1980s, the external evaluation of Project Follow Through documented the success of this approach for teaching low-income students in the primary grades, particularly in the area of mathematics (Stebbins, St. Pierre, Proper, Anderson, & Cerva, 1977; Gersten & Carnine, 1984). As a result, many advocated the use of this approach, called direct instruction, in teaching mathematics to students with LD (e.g., Hallahan & Kauffman, 1986). In a 1998 meta-analysis, Swanson and Hoskyn (1998) concluded that the combination of direct instruction and strategy instruction was an effective approach for teaching students with LD in all academic areas.

In the 1980s, direct instruction approaches began to incorporate principles gained from cognitive psychology and were increasingly referred to by the terms explicit instruction or explicit strategy instruction. In some cases, strategies were rather broad heuristics meant to teach students how to approach any type of mathematical problem (e.g., Montague, 1992). In other cases, the approach was heavily scripted and detailed precise steps students should take to solve a particular problem type. This latter approach has been criticized for its failure to help students understand underlying concepts and build flexible thinking (e.g., flexible use of a mental number line for estimation, and fluency with number properties such as the commutative and distributive laws; Woodward & Montague, 2002). However, others have argued that high degrees of explicitness and highly systematic instruction are critical for students with LD (e.g., Owen & Fuchs, 2002; Jitendra et al., 1998).

Although the nature of explicit strategic instruction has evolved over time and can vary widely from study to study, there are a number of common features that define this approach. Generally, clear consistent modeling of step-by-step strategies through teacher explanation, modeling and demonstration; careful control of task difficulty; planful sequencing of teaching and practice examples; and specified procedures for providing corrective feedback characterize explicit systematic instruction. The studies reviewed here represent a range of approaches to providing explicit systematic instruction. However, all of them include most of the instructional features described above. In addition, this set of studies also demonstrates how explicit instruction has evolved over time to incorporate more innovative instructional features that support and encourage student interaction, flexibility, and generalization.

In Owen and Fuchs' (2002) research on teaching problems involving fraction concepts, students were shown transfer problems in a careful sequence. Transfer problems referred to those with extraneous information, differing terminology from the practice items (e.g., the terms "one third" and "divide equally into three pieces"), and multistep problems that included one step involving manipulation of fractions. Students received clear feedback on their attempts to apply what they had learned in their practice sets onto the broader spectrum of problems and when they experienced problems, teachers demonstrated the underlying similarities to the previously taught problems.

Van Luit and colleagues (1999) have developed a line of research for teaching students with LD that attempts to synthesize principles of explicit strategy instruction with advances in the understanding of the underlying nature of LD in mathematics (e.g., Geary, 2005; Brown & Campione, 1990; Fuchs & Fuchs, 1998). The approach differs from earlier versions of direct instruction in several important ways. As in traditional models of direct or explicit instruction, students are taught in a quite explicit fashion one problem solving strategy at a time. As with Owen and Fuchs (2002) and Jitendra et al. (1998), teachers explicitly present a series of problem-solving steps to students and model several problems of this type for a small group of students. Students are taught multiple problem solving strategies and practice with an array of problems that use different types of syntax and different types of situations. Teachers actively encourage students to think aloud, to either walk through the steps in their strategy or articulate a reason for their decision to, for example, divide rather than multiply. Most of the intensive instruction is conducted in small groups. Teachers in Van Luit and Naglieri (1999), Jitendra et al. (1998), and Owen and Fuchs (2002) also used visual representations to teach problem solving.

Tournaki (2003) used explicit instruction to help students with LD learn more sophisticated counting strategies. She capitalized on the important insight made by Siegler (1987) that a key milestone in beginning mathematics proficiency for children is the insight that to most effectively solve a simple addition problem, it is invariably easier to start counting from the larger number, rather than the first number. For 3 + 8, it is far more efficient to count 3 up from 8 than to begin with the 3 and "count up" 8. This insight requires students to have some grasp of the commutative law and also a reasonable sense of magnitude comparison—two essential components of number sense.

The goal of this study was to examine whether the counting on strategy could be successfully taught to elementary students with LD. Instruction was quite clear and explicit. An example of the Strategy Instruction condition from Tournaki (2003, p. 458) is as follows:

> *Q (Teacher): When I get a problem, what do I do?*
> *A (Desired student response, i.e., repeat of the rule): I read the problem: 5 plus 3 equals how many. Then I find the smaller number.*
>
> *Teacher points to the smaller number and says, 3. Now I count the fingers.*
> *Q (Teacher): So how many fingers am I going to count?*
> *A (Desired student response): 3.*

After a few problems, the teacher had students solve problems while thinking aloud, i.e., repeating the steps and asking themselves the questions. Teachers always provided clear, immediate feedback when students made errors.

Note how closely this approach aligns to the depiction of explicit instruction presented earlier. Yet note how the target goal is to intentionally propel students into use of a more sophisticated counting strategy than just adding two numbers together, based on the finding from cognitive psychology (Siegler & Shrager, 1984; Geary, 1993) that students with LD tend to solve a problem such as 3 + 8 by starting at 3 and counting "up" 8 objects, whereas nondisabled students quickly learn that since 3 + 8 is the same as 8 + 3, it is much more efficient to start with 8 and count up 3 more objects.

An interesting pattern emerges in the research of Tournaki (2003) on explicitly teaching students to use the counting on strategy. There is a significant impact on the immediate computation posttest (ES = 1.612). In other words, students with LD do better when taught a strategy than when they are simply given a set of addition problems and told to do them as fast as they can. However, the significant effect measured by the transfer test (ES = 0.801) indicates that strategy-based approaches that teach students about number families and number bonds pay dividends in terms of other important areas of mathematics such as estimation.

2. Contemporary Adjustments to Explicit Strategy Instruction

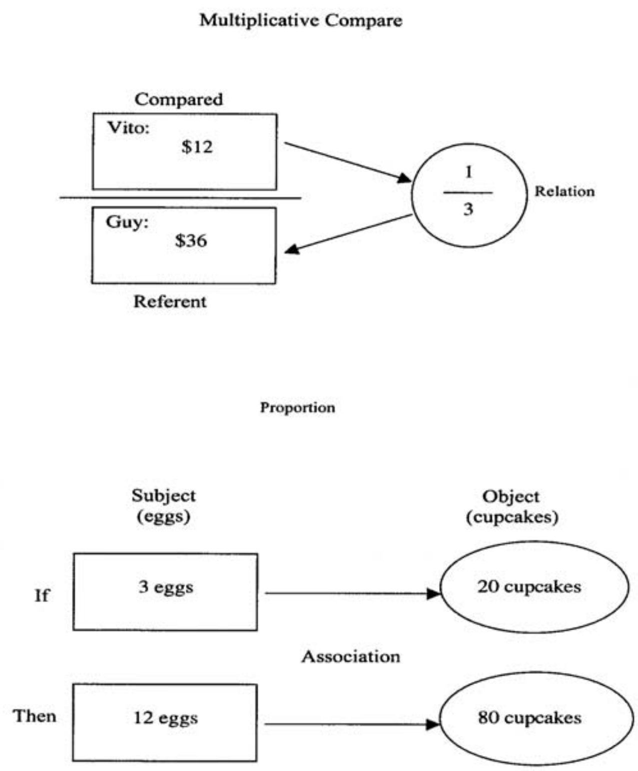
There are several additional important characteristics of most contemporary approaches to explicit strategy instruction. Van Luit and Naglieri (1999) provide a concise description. In their view, strategy instruction is when "students are taught to flexibly apply a small repertoire of strategies that reflect the processes most frequently evidenced by skilled students" (p. 99). They also stress the importance of a good deal of small group interaction in which students are encouraged and prompted to think aloud as they do mathematics, and peers provide feedback on their strategy selection and execution.

Van Luit and Naglieri (1999) began instruction with use of concrete objects but then expeditiously moved into mental solutions that entailed a good deal of thinking aloud. The final phase of each instructional cycle included a "phase of control, shortening, automization and generalization" (p. 101). What is similar between these two methods is that transfer and practice for automaticity are not assumed. Nor are students expected to develop these proficiencies by doing homework problems or by informal discussions with peers. Significant blocks of instructional time are dedicated to these goals, and teachers closely monitor student progress toward independent performance. Whereas the goal of automaticity and clear, explicit modeling remains central, teaching students how to transfer the knowledge they obtained is a major focus, and is characteristic of the more contemporary explicit strategy instruction studies.

Another strand in this research seems particularly relevant for students with LD who struggle with story problems. In related streams of research, Hutchinson (1993), Jitendra et al. (1998) and Xin et al. (2005) taught students in a systematic fashion a graphic representation to help them analyze the contents of a story problem. The three studies have addressed a) simple arithmetic word problems involving addition and subtraction ("change, combine, compare" following the Riley, Greeno, and Heller (1983) representational system) b) comparative problems involving multiplication (Xin et al., 2005), and c) word problems that typically are taught in beginning algebra.

The goal is to help students grasp the nature of word problems that involve an operation (either addition or multiplication of whole numbers) and its inverse operation. Rather than focusing on tricks such as "key words" students learn to use a visual representation to analyze the question and then discern how to handle relevant information. Exposure to all aspects of each of the problem types is deliberate and explicit. Practice is extensive, including opportunities for students to think aloud as they complete their graphic organizers. The instructor carefully highlights the key aspects of each problem type and provides a good deal of discrimination practice. Figure 1 below is an example of the graphic representation used to teach students a way to analyze multiplication problems involving comparisons (Xin, Jitendra, & Deatline-Buchman, 2005, p. 185).

Figure 2: General Problem-Solving Steps Employed in the Schema-Based Instruction and General Strategy Instruction Conditions



Source: Xin et al., 2005, p. 185.

Upon examining the full array of studies, one is struck by several features. The first is that all these studies address topics that are particularly problematic for students with LD, particularly those with difficulties in both mathematics and reading (Jordan, Hanich, & Kaplan, 2003). The second is that the pooled effects on word problems, computation, and transfer outcomes are all significant. The third is that the instructional strategies in the interventions do borrow from both the mathematics education research and the cognitive development research in mathematics. This seems an advance over the very general heuristics that comprised much of the mathematics intervention research in the special education literature 15 to 20 years ago. Those generic strategies were often borrowed from the research on reading comprehension or writing, and failed to capitalize on the advances made in research on the teaching and learning of mathematics. In fact, Xin et al. (2005) intentionally used the older, generic problem solving approach as the control group condition and found large effects favoring the more innovative approach for helping students understand the mathematical nature of the story problem.

Because in explicit strategy instruction students are invariably taught how to approach the problem type or types and are usually given precise wording to use as they think aloud, development of mathematical insight rarely plays a role in the design of the interventions (with the possible exception of Van Luit and Naglieri, 1999, in which multiple strategies are highlighted, potentially allowing such insights to develop). Therefore, there is not much known about the extent to which explicit instruction helps support students in developing such insights or understandings since proficiency and conceptual knowledge are always related in an integral fashion (Rittle-Johnson, Siegler, & Alibali, 2001).

In summary, this body of research on explicit instruction suggests that the field has made reasonable strides in understanding at least one type of intensive mathematics instruction that will help students with LD become more proficient in solving relatively basic grade level word problems and at least make some gains toward understanding how to translate stories or written problems into appropriate symbols, representations, and mathematical expressions.

This approach for providing explicit systematic instruction should also help inform the development and implementation of the type of preventative small group interventions that are increasingly used to help students who are struggling to acquire proficiency in mathematics in general classroom instruction. Preventative small group interventions provide students, who are identified as struggling in the Tier 1 core curriculum, with specific skill instruction in small groups. A major goal of preventative small group interventions is that they will reduce inappropriate referrals to special education, because if students benefit from a relatively low cost small group mathematics intervention in their general classroom, they are unlikely to require the intensive instruction that special education is intended to provide.

3. Studies Evaluating the Impact of Explicit Instruction for Low-Achieving Students

As described earlier, explicit instruction requires the teacher to be the provider of knowledge and to provide a great deal of structure and control concerning how content is learned, including the specific strategies or steps used by the children to solve the problems. Table 11 summarizes the results from the studies that investigated the effects of various strategies on the math achievement of low-achieving students. The four studies that investigated explicit instruction as a means for teaching low-achieving students are: Darch, Carnine, and Gersten (1984); Kroesbergen, Van Luit, and Maas (2004); Moore and Carnine (1989); and Woodward and Brown (2006). All but one of the effect sizes for the explicit instruction studies (i.e., Woodward & Brown, 2006) are significant.[2]

Table 11: Studies That Investigate the Effects of Various Instructional Strategies on Math Achievement for Low-Achieving Students

Study	Design	Sample	Duration/ Content	Contrast	Measure	Hedge's g	Standard Error	
Darch et al., 1984	RCT	73 low-achieving fourth-grade students in one school	Eleven 30-minute lessons/ Math story problems	Explicit Method with Fixed Time vs. Basal Instruction with Fixed Time	Researcher designed word problem test	1.914	***	0.408
Fuchs et al., 2005	RCT	139 first-grade students at risk for the development of math difficulty in 41 classrooms in 10 schools	48 sessions, 3 times weekly for 16 weeks/ Identifying numbers, more and less, addition and subtraction	Tutoring based on CRA vs. No Tutoring	Pooled computation measures (includes two tests)	0.441	**	0.179
					Pooled fact fluency measures (includes two tests)	0.180	(ns)	0.177
					Pooled conceptual and application measures (includes three tests)	0.414	*	0.179
Kroesbergen et al., 2004	RCT	265 students aged 8–11 years old from 13 general and 11 special elementary schools for students with learning and/or behavioral disorders in the Netherlands	Thirty 30-minute lessons, twice weekly, over 4 to 5 months/ Multiplication	Explicit vs. Traditional Instruction	Pooled Problem Solving Measures	0.569	*	0.262

Continued on p. 6-61

[2] An obvious outlier that was not included in the table was Cardelle-Elawar's (1995) study. The effect size was equivalent, for example, to the average control classroom being at the 3rd percentile and the average experimental classroom being at approximately the 84th percentile. The effect size for that study is extraordinarily large. This may, in part be due to the fact that the unit of analysis was the classroom, not the individual child. Effect sizes are larger when analysis is based on means of classrooms because individual differences among children within classrooms are minimized. However, there are likely to be other factors relating to the alignment of test to the intervention that lead to the study's extraordinary high effect size.

Table 11, continued

Study	Design	Sample	Duration/ Content	Contrast	Measure	Hedge's g	Standard Error	
Moore & Carnine, 1989	RCT	29 students in Grades 9–11 from three math classes for low-performing students in a high school in a medium-sized city in the Northwest	Twenty 50-minute lessons/ Ratio and proportion word problems	ATCD (Active teaching with empirically validated curriculum design) vs. ATB (Active training with basals)	Researcher designed criterion referenced test to assess student mastery of specific mathematics skills	0.994	**	0.386
Pasnak et al., 1991[a]	RCT	85 low-performing students from 17 Kindergarten classes in six neighboring Northern Virginia schools	3–4 sessions per week over three months/ Introductory mathematical concepts	Piacceleration vs. Control	SESAT math subtest	0.520	(ns)	0.348
Thackwray et al., 1985	RCT	60 third- and fourth-grade children with teacher perceived academic problems from three urban public schools	Four 45-minute sessions/Addition	Specific self instruction vs. Didactic	Pooled math quiz (ES = 0.501) and Peabody Individual Achievement Test (ES = 0.780)	0.641	*	0.325
Woodward & Baxter, 1997[b]	RCT	38 low-achieving third-grade students in nine classes in three schools located in the Pacific Northwest	One school year/ Third grade math	Everyday Mathematics vs. Heath Mathematics Program	ITBS including computation (ES = -0.176), concepts (ES =0.199), and problem solving skills (ES = -0.085) subtests	-0.223	(ns)	0.635
Woodward & Brown, 2006[b]	Quasi	53 students in two middle schools in nearby medium-sized suburban school districts. Students had been identified as low-achieving in mathematics by elementary school teachers. No student had an IEP for mathematics	One school year/ Both curricula emphasized core NCTM strands: numbers, operations, measurement, geometry, data analysis and probability	Transitional Math Curriculum vs. Connected Math Program	Pooled standardized (ES = 0.797) and researcher developed test (ES = 1.435)	1.116	(ns)	0.688

~ p < .10, * p < .05, ** p < .01, *** p < .001
[a] Data were adjusted for clustering that occurred within classrooms.
[b] Data were adjusted for clustering that occurred within schools.

Two of the studies (Darch et al., 1984; Moore & Carnine, 1989) used a highly explicit approach based on the Direct Instruction model articulated by Silbert, Carnine, and Stein (1989) and Engelmann and Carnine (1982). This is a traditional approach to explicit instruction, which has been widely used in the field of special education with students with LD, especially in the 1980s and early 1990s. With direct instruction, teachers model how to solve a specific problem type, and spell out the necessary steps. Students learn the steps and through careful sequences of examples, practice solving problems in the precise fashion that they were taught. Another one of the studies, Kroesbergen et al. (2004) also employed a highly explicit instructional approach. That is, students (8–11 years old) were instructed via directions and modeling by their teacher how and when to apply specific strategies for solving multiplication computation problems. Students were directed to only use the strategy taught by the teacher. Although highly explicit instruction has been shown to lead to enhanced academic outcomes for students with learning disabilities, and other students considered at risk for experiencing difficulties in mathematics (e.g., Gersten & Carnine, 1984; Baker, Gersten, & Lee, 2002) some have questioned the extent to which students actually learn the underlying rationale behind the strategies that are explicitly taught (e.g., Woodward & Howard, 1994; Woodward & Montague, 2002).

The two studies (Kroesbergen et al., 2004; Woodward & Brown, 2006) could be characterized as teaching students a variety of heuristics for problem solving but with significant segments of instruction following the highly explicit nature of classic direct instruction. In fact, one of the goals in the framing of some of the research studies we discuss below is an attempt to ponder and define the nature of explicit instruction for low-achieving students. Their thinking is helpful in beginning to unpack this construct.

Woodward and Brown (2006) found that despite statements by the National Council of Teachers of Mathematics (NCTM) (2000) indicating that students experiencing difficulties in mathematics benefit from a challenging curriculum, they could not locate any research to support this claim. They note, "In-depth examinations of this population indicate that without substantive modifications, these students do not exhibit high levels of success on either academic measures or everyday activities" (e.g., Baxter, Woodward, Wong & Voorhies, 2002; Woodward & Baxter, 1997, p. 151). Their analysis of the relevant research, with which we concur, notes that effective components of instruction for low-achieving students in mathematics supports the use of both concrete and visual representations of concepts, carefully orchestrated practice activities with feedback on all aspects of mathematics and high, but reasonable, expectations.

Woodward and Brown (2006) evaluated an intervention, written by Woodward, called Transitional Mathematics, and attempted to put these components into practice in six intensive, remedial middle school classrooms. The curriculum included numerous visual models for representing mathematical procedures in a meaningful way. They present, for example, difficult concepts such as place value in three-digit addition by both using a written algorithm and a visual model that depicts the algorithm. Regrouping was taught via systematic use of expanded algorithms as well as visual models of the expanded algorithm. Practice on relevant mathematics facts, and factoring was part of each daily lesson. The teacher explicitly introduced the concepts, and worked problems with the group that exemplified the concept before students broke into pairs. Because of the reading difficulties of many of the students, the teacher most often read the problem to the students. Guided practice consisted of approximately five problems worked on by students and reviewed with the teacher. This part of the lesson also included a good deal of checking for understanding (Good & Grouws, 1977) and attempts to explore any student misconceptions. Practice was typically done in pairs and included opportunities for students to explain their reasoning to each other and with the class. Students in the comparison classroom were taught using the Connected Mathematics Program, a commonly used middle school curriculum. Woodward and Brown characterize this as follows: The core emphasis of this program is problem solving, and students typically read descriptions of problems as part of each lesson. Connected Mathematics is much more contextualized in elaborate “real-world” problems, and has a more peripheral attention to skill development, in contrast with Transitional Mathematics that integrates the latter with distributed practice. This quasi-experimental study involved two schools, one the intervention school and one the comparison school—no mention was made concerning how schools were designated.

Regarding differences in achievement between groups, the effect size on the Terra Nova, a standardized mathematics achievement test was not significant but indicative that the Transitional Mathematics treatment is a promising approach (ES = 1.116). Note how this study, like the others in the explicit instruction set also includes an array of other practices deemed to be beneficial—use of guided practice, intensive use of visual models so that students can represent problems in multiple ways (Donovan & Bransford, 2005), clear and explicit instruction in use of the concepts and provision of heuristics for problem solving.

Kroesbergen et al. (2004) and Darch, Carnine, and Gersten (1984) used an approach that was even more explicit than the Woodward and Brown (2006) model. In these studies, teachers modeled an approach for solving problems and students were expected to follow the teachers’ model. Teachers did explain when the strategy was appropriate, and provided examples of occasions when it was not appropriate. In both cases, the degree of structure was higher than in Woodward and Brown (2006) and students were not given a chance to talk through their approach for solving the problem with a partner or the teacher. Both studies (Kroesbergen et al. and Darch, Carnine, & Gersten) demonstrated significant effects on researcher-developed measures that were aligned with curricula taught, favoring the explicit instruction groups. In the Kroesbergen et al. (2004) study, the effect size in the area of word problems involving multiplication was significant (ES = 0.569) and in Darch, Carnine, and Gersten (1984) there was also a significant effect size in the area of word problems (ES = 1.914) that cut across all four basic arithmetic operations.

Moore and Carnine (1989) explored the degree of explicitness within the context of highly interactive, teacher-directed instruction. In this study, high school students were taught how to solve ratio and proportion problems. Whereas control group students were taught to ask themselves “Is this information important?” students in the experimental condition were taught in a much more step-by-step fashion and were taught strategies for each of four types of problem sets. For this study, the effect size was significant (ES = 0.994) on a test of mastery of specific skills. However, one must take into account that this study measured only the topic covered, in contrast to Woodward and Brown (2006), which measured all aspects of mathematics covered in a typical school year.

F. Other Approaches for Teaching Students With Learning Disabilities

This next section addresses other instructional approaches for teaching mathematics to students with LD. The findings are organized by three major themes:

- Selection of examples to foster development of more sophisticated strategies for quick retrieval of basic arithmetic facts
- Emphasis on visual representation
- Emphasis on encouraging students to think aloud

It is interesting to note that virtually all of the studies in these categories also have at least a reasonably strong degree of explicitness in the design of their instruction—a feature that is consistent across the body of studies reviewed for this section.

1. Strategies for Quick Retrieval of Basic Arithmetic Facts

Quick retrieval of basic arithmetic facts or combinations has been assumed by virtually the entire mathematics education community as critical for success in more advanced mathematics. It is considered a necessary, though not a sufficient requirement for emerging mathematical competence. Researchers in the field of LD have found for several decades that slow and inaccurate retrieval of basic combinations is a clear, consistent early indicator of persistent serious difficulties in mathematics (Gersten, Jordan, & Flojo, 2005; Geary, 2005; Jordan, Hanich, & Kaplan, 2003; Goldman & Pellegrino, 1987; Hasselbring, Goin, & Bransford, 1988).

Two studies with students with LD were included in this classification, and are summarized in Table 12. Beirne-Smith (1991) attempted to examine whether sequencing of examples could enhance facility with basic addition combinations for students with LD. She used an array of facts developed by Carnine and Stein (1981) that was geared toward helping students see that to compute, for example, 8 + 2, all they needed to do was count up by 2. Examples of the array are 2 + 4, 2 + 5, and 2 + 6. The impact of the sequence did not lead to significant improvement over simple rote practice.

Table 12: Studies That Investigate the Use Of Strategies With Students With Learning Disabilities to Develop the Ability to Quickly Retrieve Arithmetic Facts

Study	Design	Sample	Duration/ Content	Contrast	Measure	Hedge's g	Standard Error	
Beirne-Smith, 1991	RCT	30 students with LD aged 6 to 10 years old from four schools in two adjacent southeastern school districts were tutored. 20 students with no learning disabilities in Grades 3–6 served as tutors	30-minute tutoring sessions for four weeks/ Single-digit addition facts	Counting-on procedure vs. Rote memorization	Oral test on addition facts	0.165	(ns)	0.448
Woodward, 2006	RCT	15 fourth-grade LD students from two "mainstreamed" classrooms in a school in a suburban school district in the Pacific Northwest	20 25-minute sessions daily over four consecutive weeks/ Multiplication facts	Strategy and timed practice vs. Time practice via direct instruction	Pooled researcher designed computation measures	0.377	(ns)	0.509

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

Woodward (2006) extended this line of research to much more complex multiplication combinations, which require many students to rely on some variant of multiplication tables and sheer rote practice. He developed an intervention containing two components. The first involved explicit instruction in an array of strategies that can help with quick retrieval of multiplication combinations. These included numerous shortcuts based on properties of numbers. One is counting backward for combinations of 9, i.e., knowing that 8×9 achieves the same answer as $8 \times 10 - 8$. Another is use of the distributive law, e.g., 37×5 equates to the same product 35×5 plus 2×5 An advantage of this strategy's approach is that students could not only learn more efficient ways to compute these multiplication facts but also develop their facility with using properties of numbers to solve problems.

However, Woodward (2006) noted that strategy instruction will not, in and of itself, promote quick retrieval of mathematical combinations for all students with LD (see Hasselbring, Goin, & Bransford, 1988). He therefore combined the strategy instruction and practice with timed practice drills. He compared students taught with a combined strategy instruction and timed practice approach to students taught only with timed practice.

Results were positive favoring the strategy group and nonsignificant. However, given the small sample size, and inconsistent findings across mathematics domains, one can only infer that this approach—or aspects of this approach—might be worth exploring in terms of development of more fluent retrieval as well as in helping students understand more about number families and increasing their ability to estimate. However, the set of studies on building fluency in computing mentally or retrieving arithmetic combinations indicates that there is a good deal more to be learned about how to improve students' proficiency in this critical area.

2. Use of Visual Representations, Visualization, and the Concrete-Representation-Abstract Approach

Adding It Up, the 2001 National Research Council report on the teaching of mathematics, eloquently describes the role of representations in the teaching and learning of mathematics, a role that has not always been adequately highlighted until recently in the instructional research on LD.

> Mathematics requires representations. In fact, because of the abstract nature of mathematics, people have access to mathematical ideas only through the representations of those ideas. … *Much of the real intellectual work in mathematics concerns the interpretation and use of representations of mathematical ideas* (pp. 94–95, emphasis added).

The authors explain that mathematical ideas are often metaphorical, and thus, a representation or multiple representations are excellent means for conveying mathematical ideas. This section summarizes a set of five recent studies on the role of visual representations as a key means for teaching mathematical ideas, strategies, and procedures to students with LD. Each researcher approaches the use of representations somewhat differently.

Table 13 presents information on each of the studies, as well as the outcomes of the studies. Because the instructional approaches are so different, the Task Group did not pool effect sizes across the set of studies. However, taken together, these approaches reflect a trend toward serious thinking about instructional uses of representations that include physical models using manipulatives, pictorial representations, abstract representations using geometric shapes, as well as increasingly abstract representations, such as number lines and graphs of functions and relationships. It should be noted, however, that the effect sizes across studies are quite variable.

Table 13: Studies That Investigate the Use of Concrete Instruction and Visual Representations Used for Students With Learning Disabilities

Study	Design	Sample	Duration/Content	Contrast	Measure	Hedge's g	Standard Error	
Butler et al., 2003[a]	RCT	50 students in Grades 6–8 with mild–moderate disabilities (42 students with specific learning disabilities in math and eight with other disabilities) from a public middle school located in a large urban area of the Southwest	Ten 45-minute lessons/ Fraction concepts and procedures	Concrete-representational-abstract (CRA) vs. Representational-abstract (RA)	Area Fractions, Quantity Fractions, and Improper Fractions subtests provided measures of conceptual understanding of fraction equivalency and Abstract Fractions and Word Problems subtest provided a measure of application	-0.095	(ns)	0.526
Manalo et al., 2000—Experiment 1	RCT	29 From three students (equivalent to eighth grade) with learning disabilities from two schools in the Palmerston North area of New Zealand	Five 25-minute sessions twice per week/ Addition and subtraction	Process mnemonics vs. Demonstration imitation	Immediate Posttests - Addition and subtraction computation skills tests	-0.043	(ns)	0.477
					1 week follow-up tests	0.076	(ns)	0.477
					6 week follow-up tests	0.956	~	0.506
Manalo et al., 2000—Experiment 2	RCT	28 From three students (equivalent to eighth grade) with learning disabilities from two schools in Auckland, New Zealand	Ten 25-minute sessions twice per week/ Addition, subtraction, multiplication, and division	Process mnemonics vs. Demonstration imitation	Immediate Posttests - Addition, subtraction, multiplication, and division computation skills tests	-0.153	(ns)	0.475
					1 week follow-up tests	0.180	(ns)	0.472
					8 week follow-up tests	1.876	**	0.579
Walker & Poteet, 1989[a]	RCT	70 sixth- and eighth-grade LD students receiving mathematics instruction in resource room programs in four Indiana school districts	Seventeen 30-minute lesson/ Problem solving strategies for simple word problems involving addition and subtraction	Instruction using diagrammatic representations vs. Traditional instruction	One-step story problem-solving test	0.349	(ns)	0.330
Witzel et al., 2003[a]	RCT	34 matched pairs of sixth- and seventh-grade students with learning disabilities or at-risk for difficulties in math (41 LD students and 27 at-risk) in 12 inclusive classrooms in an urban county in the Southeast	Nineteen 50-minute lessons/ Algebraic transformation equations	Concrete-representational-abstract (CRA) vs. Abstract	Algebra transformation equations test	0.826	*	0.346

~ p < .10, * p < .05, ** p < .01, *** p < .001
[a] Data were adjusted for clustering that occurred within classrooms.

The intriguing set of two experiments by Manalo and colleagues (2000) examines the use of easy-to-imagine visual stories and schema to help students remember rules, principles, and procedures. These studies address a potentially important issue in the practice of teaching mathematics: how to provide prompts or facilitators to help students create visual representations. In this study, explicit teacher-directed instruction (modeling followed by guided practice with clear feedback) is a constant. The variable is visualization.

Manalo et al. (2000) adapted an approach from a Japanese educator, Nakane. The goal of this approach is to "summarize the organization and the process of problem solving … using familiar metaphors expressed in familiar ways" (p. 138) and thus to teach mathematical operations in a clear, comprehensible fashion. The goal of the researcher was to present mathematical problems as interesting, easy to visualize narratives that would engage the students, and thus enhance their interest in the process, their memory of the procedures taught, and the questions that children must ask themselves before deciding on a strategy for solving a problem.

For Study 1, the topic was basic arithmetic computation problems; all participants were screened to ensure that they were not proficient in use of standard algorithms for multidigit operations involving regrouping, even though they had been taught this material before, often many times before. Numbers were presented as characters and operations as stories. For example, to teach subtraction, students were asked to visualize warriors with numbers on their uniforms, and to visualize that the bigger the number on the uniform, the stronger the warrior. The teacher used simple drawings to demonstrate the procedure or story. For subtraction, the top number represented the attackers and the bottom numbers the defenders, and students were told that the attackers weakened during the battle. The number on the uniform of a defender told a student how much strength was sapped from the warrior. In cases involving regrouping (e.g., 33-5), students were told that for example, a warrior with strength of 3 would not have adequate strength to sustain a battle with a defender with strength of 5. Thus, the army would need to regroup and borrow some strength from the warrior with strength of 30. The teacher used pictures to demonstrate the process of regrouping.

Similar stories were developed for multiplication and division. The approach used to teach students in both the experimental and control conditions was a combination of model-demonstration with guided practice and feedback. Two experiments were conducted. The first entailed the researcher as the teacher; the second used two different teachers. For both studies, the pattern of findings was similar. No significant effects were found on the immediate posttest or a test administered one week later. Yet, on the six-week and eight-week follow-up tests, the effect for Experiment 1 (in which the researcher did all the teaching) was 0.956, which bordered on statistical significance, and for Experiment 2, which used teachers other than the researcher, the effect was larger and statistically significant (effect size = 1.876).

Given the problems with maintenance of knowledge for many students with LD, these results seem worth noting. The use of consistent visual representations and stories to help students think through their decisions about appropriate computational processes is an important fact to note. One wonders about its impact on helping students with LD translate more complex mathematical problems and work with more complex mathematical concepts.

3. Visual Representations and Helping Students Understand Visual Representations by Use of the Concrete-Representational-Abstract (CRA) Method

Two of the studies in this section (Butler et al., 2003; Witzel et al., 2003) examine the use of concrete-representational-abstract (CRA) instruction for students with learning disabilities. This sequence of instruction begins at a concrete level, with students manipulating objects. Once students understand a topic concretely, they work with the topic using visual representations. Once the students are comfortable with how the topic can be represented in multiple ways, they work with the concepts at a more abstract level. The Walker and Poteet study (1989–1990) are also included in this section because their work can be seen as a precursor to the more complex CRA model.

In the earliest study in this subcategory, Walker and Poteet (1989–1990) compared a diagramming method of problem solving with a keyword approach. Subjects were middle school students with LD. In both conditions in this study, explicit instruction was a constant and not a variable. The experimental variable using a visual representation to help students organize information from one- and two-step story problems involving basic addition and subtraction, then to translating the pictures into numerical expressions, and ultimately to computing the answer.

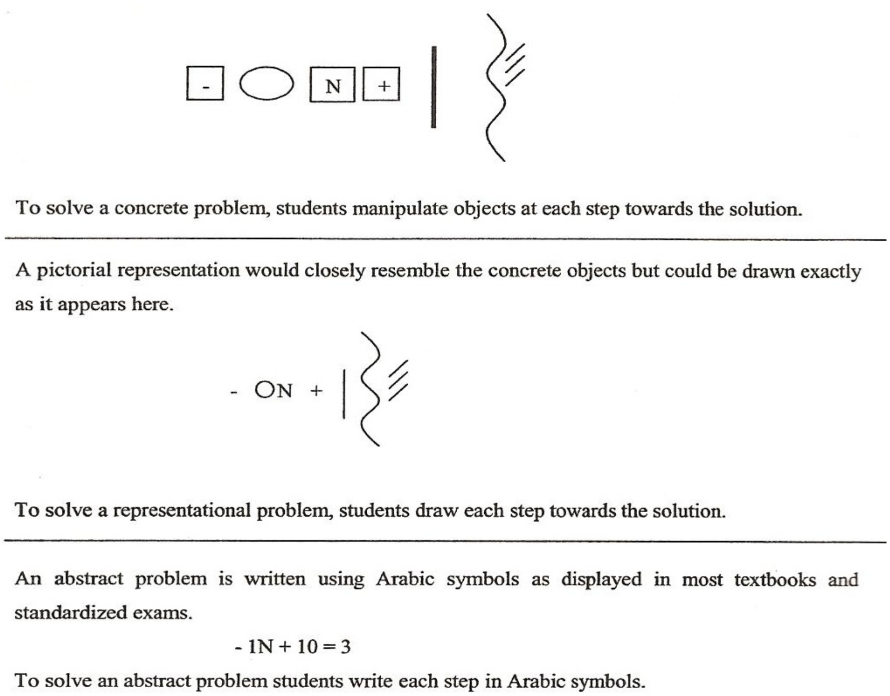
Although this skill seems exceptionally easy for middle school students, poor performance on word problems is prevalent with this group. In fact, on the pretest, the average score for students was equivalent to 16.43 correct (out of 32 possible problems, 51.34% correct). Students in the diagramming group were taught to create diagrams that bear similarity to those used by Xin et al. (2005) and Jitendra et al. (1998). Finally, students were asked to compute the actual solution. The comparison group was taught to identify keywords in the problem that could then be directly translated to specific numerical operations. Despite not reaching statistical significance, the effect size (0.349) suggests that, given the difficulty that students with LD have in developing proficiency in this area; the approach could well be labeled promising.

Butler et al. (2003) used the CRA to teach middle school students with LD basic concepts and procedures involving fractions. The topics included concepts and procedures related to equivalence of fractions and computations involving fractions. The authors note that researchers (e.g., Woodward & Montague, 2002) have suggested that many students with LD lack any real understanding of the concepts underlying various procedures that they can perform and that these problems truly surface once students begin to work with rational number concepts and operations. In this study, as in Walker and Poteet (1989–1990), the instructional methodology was similar for experimental and control students in that explicit instruction was used in both conditions. The major difference was that CRA students spent three days working with concrete objects, three days with visual representations, and only then moved on to abstract, symbolic notation. The control condition began with visual representations for three days. Major emphasis in both conditions was placed on fractions as part of a set, as opposed to fractions as area. As one can see in the first study in Table 13, the effect size for this method was nonsignificant. Thus, the CRA intervention implemented in this study was not more effective than the control condition in teaching fractions to middle school students with LD.

The other CRA study, Witzel et al. (2003), also conducted with middle school students, differs in several important ways from Butler et al. (2003). The first is that the topic was a more difficult one, algebraic transformation equations. The second is that the researchers used CRA quite differently. For example, Witzel et al. progressed more fluidly from concrete, to visual, to abstract. The third difference is that, in this case, a researcher-developed measure was used rather than the standardized measure used by Butler et al. Finally, as can be seen in the last study in Table 13, effect size (ES = 0.826) is statistically significant.

The authors note that because of the abstract nature of algebra, building a mathematically accurate concrete representation is much more of a struggle. Figure 3 presents an example of the instructional materials used and how the researchers grappled with representation of a variable (*x*) with concrete objects when x can represent any real number (Witzel et al., 2003, p. 127).

Figure 3: Concrete, Representational, and Abstract Examples of an Inverse Operation



Source: Witzel et al., p. 127.

G. Strategies That Encourage Students to Think Aloud

The Task Group identified two studies that examined strategies that encouraged students with LD to think aloud (Ross & Braydon, 1991; Schunk & Cox, 1986). Table 14 below summarizes characteristics for the two studies, and presents the effect sizes.

Table 14: Studies That Investigate the Impact of Think Aloud Strategies With Students with Learning Disabilities

Study	Design	Sample	Duration/ Content	Contrast	Measure	Hedge's g	Standard Error	
Ross & Braden, 1991[a]	RCT	94 elementary school students with LD in nine intact special education resource rooms classified as learning disabled in math	Nineteen 60-minute sessions over four weeks/ Addition and subtraction	Cognitive behavior therapy in which students are instructed to talk aloud vs. Direct instruction	Stanford Diagnostic Mathematics Test - computations	0.135	(ns)	0.434
Schunk & Cox, 1986	RCT	90 students classified with LD in math from six middle schools	Six 45-minute sessions/ Subtraction with regrouping	Continuous verbalization vs. No verbalization	Subtraction test	1.005	***	0.271

~ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$
[a] Data were adjusted for clustering that occurred within classrooms.

As previously mentioned, asking students to think aloud was a major component in many of the explicit instruction studies. What differentiates these two studies from those in the explicit instruction set (e.g., Manalo, 2000; Tournaki, 2003) is that in these studies, verbalization was the sole independent variable. In contrast, in the other explicit instruction studies, students thinking aloud was but one of several instructional components. Thus, these two studies suggest that: encouraging students to think aloud as they work on arithmetic problems shows promise as one component of a mathematics intervention.

Both of these studies were influenced, to some extent, by the research of Donald Meichenbaum (1985), which suggested that students with learning disabilities, behavior disorders and, in all likelihood attention deficit disorders, could be helped in many areas of both academic and social development by being taught to verbalize. The Meichenbaum approach targets one of the key characteristics of students with LD—Geary's (2005) concept of impulsivity and Kolligian and Sternberg's (1987) concept of lack of task persistence. By actively encouraging students to speak to themselves about the strategies they are using to solve a problem, the researchers felt that students would be inhibited from quickly, almost recklessly proceeding forward without serious thought. In addition, the focus on active encouragement of thinking aloud is integrally linked to Vygotsky's notion that thought is inner speech, and that students may well need to go through a period of actually thinking out loud, especially those students with learning difficulties.

Schunk and Cox (1986) examined the effectiveness of verbalizing the steps of problem solving with middle school students with LD working two- to six-column subtraction problems with and without regrouping. As in the Tournaki study, the teacher in the treatment group talked through the steps of solving the subtraction problem and then students worked several problems while verbalizing the steps.

Students in the comparison group learned the same procedures but were in no way encouraged to verbalize during problem solving. This study is different from both the Friedman (1992) and Lambert (1996) studies in that students were solving computation problems not word problems, although some of the computations were rather complex. A statistically significant effect size of 1.005 was found on a subtraction computation test that was closely aligned to the types of problems used during instruction. Effects were modest and not statistically significant for the Ross and Braden (1991) study (ES = 0.135), although this could be due to the fact that the measure lacked the tight alignment to the intervention of the Schunk and Cox study. In any case, one study (Schunk & Cox, 1986) but not the other (Ross & Baden, 1991) suggests that for students with LD, merely encouraging self-verbalization or thinking aloud can have beneficial effects in terms of learning mathematics.

H. Other Approaches for Teaching Low-Achieving Students

Unlike studies of other approaches to teaching students with LD, most of the other studies reviewed with low-achieving students can be characterized as providing primarily *implicit* instruction (with the exception of Fuchs et al., 2005). Implicit instruction refers to the teaching approaches that provide students with broad guidance in terms of general procedures for solving problems, including relatively broad questions to ask themselves. However, there is little in the way of specific guidance in how students construct knowledge, and these approaches do not necessarily include any mathematics in them. Students are provided strategies that are used to solve math problems, such as teaching students to think aloud, or use visual representations with strategic use of manipulatives. For example, Thackwray, Meyers, Schleser, and Cohen (1985) taught students five specific self-instructions to say out loud while solving addition word problems. Presumably, these self-instructions were intended to enhance student's ability to construct accurate representations of the problem features and solution strategies.

The four studies that are included in this category are Fuchs et al. (2005); Pasnak, McCutcheon, Holt, and Campbell (1991); Thackwray et al. (1985); and Woodward and Baxter (1997). Table 11 summarizes the characteristics of these four studies. It is important to note that these studies do provide various degrees of teacher direction, so we prefer the term "primarily implicit instruction" rather than implicit instruction because all studies reviewed below seem to provide instruction primarily, but not necessarily exclusively, via an implicit instructional approach. We treat each study separately as we did for studies considered in this section for evaluating the impact of explicit instruction for low-achieving students.

One study, Thackwray et al. (1985), examined the effects of encouraging students to think aloud as they worked using the cognitive behavioral model developed by Donald Meichenbaum (1985), which was a common special education technique in the 1970s and 1980s. This technique was based on the premise that thinking aloud consistently will increase students' ability to reflect on their actions and help dissipate some of the impulsivity that is typical of low-achieving students in mathematics. Students were taught five steps. The first step involved orienting students to solve the problem. The next two steps appear below:

Step 2: First, I have to look at the problem very slowly to determine if it is addition, subtraction, multiplication or division.

Step 3: This one is addition. I can tell by the sign. (Thackwray et al., 1985, p. 301).

First, the instructor (a graduate student) modeled the steps; gradually, the student performed the five steps independently with no prompting. This study exemplifies implicit instruction because teachers provided minimal control over how students solved the problems. Rather, students were allowed to verbalize as they wished. In the specific self-instruction condition, the experimenter modeled the self-instructions (verbalizations) while the teacher performed two, three, and four digit addition problems. Using Meichenbaum's (1975) five-step fading procedure, the experimenter gradually required the child to verbalize while performing each step toward solution alone while solving the math problem. In the didactic condition, children were simply provided instructions concerning what to verbalize during problem solving. However, no modeling of the verbalizing process was provided.

Thackwray et al. (1985) investigated the effectiveness of this approach in a study involving 60 third- and fourth-graders who were perceived as experiencing difficulties in mathematics by their teachers. Although teacher judgment is no substitute for a mathematics performance measure, it often is reasonably accurate (Hoge & Coladarci, 1989). The intervention was quite short: four 45-minute lessons. The content was problems involving whole number addition. The control group received typical lecture-demonstration-practice with feedback instruction. The effect size (ES = 0.641) was significant, suggesting evidence of efficacy for this approach. The outcome was a composite of a standardized test: the Peabody Individual Achievement Test (PIAT) and a 20-item addition test. Based on the one study, there appears to be evidence of the effectiveness of promise in this general approach for problem solving, though replication of these findings in other studies would seem important for further research.

1. Instruction in Piagetian Cognitive Operations (Classification, Seriation, and Number Conservation)

The writings of Jean Piaget have always played a role in instructional research in mathematics, most recently in Griffin, Case, and Siegler (1994). Pasnak et al. (1991) examined the impact of small group instruction on Piagetian cognitive operations on kindergartners' performance on the Stanford Early School Achievement Test (SESAT)-Mathematics. The SESAT is essentially a readiness test, as opposed to a mathematics achievement test.

The sample was also selected by the kindergarten teachers as students who were having difficulty learning the basics of mathematics in kindergarten. The researchers used the students' scores on the Otis Lennon School Ability Index to confirm that they were in the at risk category. On average, these students were at least .5 standard deviation units below the school mean.

Instruction focused on the three Piagetian concrete operations that many students acquire informally before kindergarten (classification, seriation, and conservation). Many types of manipulatives were used (bolts, cups, lima beans, dominoes etc.). The amount of time devoted to this instruction was appreciable, three months of 15–20 minute small group lessons, delivered three to four times a week. Control group students received typical kindergarten instruction in numbers and number concepts.

The effect size (0.520) was not statistically significant, when corrected for classroom level clustering. Nonetheless, the magnitude of the effect size, especially given the fact that a standardized achievement test was used which was not closely aligned to the specific content taught, suggests there may well be some promise to this approach.

2. Evaluation of the Effects of 'Reform' Curricula on Low-Achieving Students

Woodward and Baxter (1997) conducted a small, but oft-cited, quasi-experiment that examined the impact of *Everyday Mathematics*, one of the curricula assumed to be consistent with the 1989 NCTM *Curriculum and Evaluation Standards for School Mathematics*. The study involved 38 low-achieving third-grade students in nine classes. The researchers assessed the impact of the reform curriculum versus a more mainstream commonly used core mathematics text. Note that students in neither condition received any additional support in mathematics from either special education or Title I. Results were nonsignificant (when adjusted for within-school clustering) and favored the control group (ES = -0.223). The reform curriculum produced one positive effect on the Concepts section of the Iowa Test of Basic Skills (ITBS) and two negligible negative effects: on the ITBS subtests for Computation and Problem Solving. None of these effects was significant. One reasonable conclusion is that low-achieving students require additional support and intensive work on foundational skills and that use of an innovative curriculum will not lead to any serious benefit unless such support is provided above and beyond the students' classroom mathematics instruction. One notes that more recent research, including research by Woodward, adopts approaches that combine interest in teaching concepts along with procedures to build conceptual knowledge with the use of explicit instruction.

3. Response to Intervention: Evaluation of a Preventative Small Group Intervention for First-Graders at Risk for Experiencing Difficulties in Mathematics

We identified only one study that investigated the effects of tutoring using concrete-representational-abstract (CRA) instruction with low-achieving students.

Fuchs et al. (2005) screened first-grade students in 41 classrooms in 10 schools using a set of screening measures that are known to be valid and reliable (see for example Gersten, Jordan, & Flojo, 2005). These students received small group instruction three times per week, a typical procedure for preventative small group (Tier 2) interventions. Core components of the intervention included strategic use of manipulatives to ensure students understood more abstract visual representations and mathematical symbols, heavy emphasis on problem solving and discussion of solutions, and use of technology to provide individualized practice on basic addition and subtraction combinations to increase quick and fluent retrieval.

Fuchs et al. (2005) used a wide array of both researcher-developed and standardized measures, of computation and concepts, applications, or word problems, as well as addition and subtraction fact fluency. Effect sizes were 0.414 and significant, favoring the tutoring group for concepts or problem solving, 0.441 and significant for the combined computation measures and 0.180 but not significant for the two fact fluency measures. The effects were stronger for the computation and concepts measures than the fact fluency measure, indicating that the technology component appeared to be the weakest facet of the intervention. In interpreting the effect sizes, the reader should note that the control group students received no additional instruction. Thus the independent variable is receiving tutoring using a CRA-based instructional model versus receiving no additional support whatsoever.

In general, this appears to be an effective preventative small group early intervention for students who exhibit problems in mathematics at the beginning of the first grade. It also is a solid example of how both concepts, procedures, and problem solving can be taught and practiced in an intense, integrated fashion. It should be noted that beyond Bruner's concrete-pictorial-symbolic sequence, no information is provided about how the tutors interacted with the children about the mathematics.

I. Summary and Conclusions

The Task Group was able to locate a reasonable number of high-quality experimental and quasi-experimental studies that investigated the effectiveness of various mathematics interventions in teaching mathematics to students with LD and LA. These studies provide a great deal of guidance concerning some defining features of effective instructional approaches for students with learning disabilities as well as low-achieving students. These features, many of which are associated with explicit systematic instruction, can be roughly categorized as follows:

1) Concrete and visual representations (mathematical drawings)
2) Explanations by teachers
3) Explanations and math talk by students in whole class discussion
4) Students working together
5) Carefully orchestrated practice activities with feedback
6) High but reasonable expectations

Some additional features of this research are noteworthy beyond the generally consistent effectiveness of both explicit and primarily implicit instructional approaches (interestingly Kroesbergen et al. (2004) actually compared and found no differences in multiplication outcomes between these two approaches). The first is that studies varied widely in terms of mathematical skills that were targeted. Most included a focus on computation skills, while others included specific attention to word problem solving. This focus on problem solving in research on students with learning disabilities and low-achieving students is a relatively recent trend, and an important one, because students with LD and LA struggle, in particular, with word problems.

The second is that a small but important set of studies examined best methods for building quick retrieval of arithmetic combinations. Mathematics educators have long been aware of the importance of quick retrieval of basic combinations so that students can focus on the problem at hand. Retrieval is stressed in the NCTM's *Curriculum Focal Points for Prekindergarten through Grade 8 Mathematics: A Quest for Coherence* (2007). In addition, research on learning disabilities has consistently documented that inefficient and ineffective retrieval of combinations is typical of a student with learning disabilities in mathematics. Addressing this issue has been more of a struggle. Programs have been developed that orchestrate practice sets for each student and try to teach similar combinations together. However, Hasselbring et al. (1988) noted that even these programs are not successful with many students with LD.

The studies by Tournaki (2003) and Woodward (2006) are important because they demonstrate that there is wisdom in teaching students strategies about computation as a means of increasing speed and accuracy of retrieval. If nothing else, this type of instruction is more interesting and potentially engaging for students and more likely to build a deeper understanding of the number system than pure rote memorization. Note that Woodward intentionally paired strategy instruction with 15 minutes of timed practice. This mixture is one that seems to show promise.

Additionally, many studies examined approaches to instruction that, based on the description, included coverage of conceptual understanding. In some studies, students were provided visual models so that students could use a visual representation to either compute or solve a word problem. Others used strategies that encouraged children to analyze word problem structure, so that meaningful patterns could emerge such as via explicit instruction or verbalizing. Mathematics educators have long been aware of the importance of developing an understanding of number operations, and patterns in problem solving, and this emphasis on meaningful understanding of operations is stressed in the NCTM *Focal Points*. In addition, research on mathematical learning in general, and mathematical disabilities and low-achievement, is associated with the nature of development of areas of number sense, including conceptual understanding of mathematical procedures and strategies for obtaining solution (Gersten & Chard, 1999; Hecht, Vagi, & Torgesen, 2007). Programs have been developed that orchestrate practice sets for each student and try to teach meaningful understandings of numbers and number operations. However, Fuchs et al. (2005) remind us that future work is needed to increase the power of classroom as well as tutorial treatments in low-achieving (at-risk) children.

1. Quality of Mathematics Taught in the Studies

In order to obtain an independent review of the quality of the actual mathematics taught in this set of studies, two research mathematicians involved in mathematics education, and one prominent mathematics educator were asked to examine the mathematical content and the nature of instruction (as opposed to the research design and technical details) of a small subset of studies. They looked at studies that described the mathematics content and instructional procedure with some amount of detail because we saw no benefit in, for example, asking a mathematician to evaluate a study in which it said, “Students learned the material in the third-grade mathematics state standards,” or studies that focused on very simple algorithms. Thus, we tended to choose the studies that bit off the most ambitious mathematical material and developed seemingly effective means for teaching the materials to students with learning problems and learning disabilities. We selectively summarize some results in this section.

The reader will recall that Woodward (2006) employed a combination of individualized fact practice with instruction that involved work with number families and applications of the distributive law to ease mental computation fluency. (For example, students learned that it is usually easier to calculate 8×9 by remembering that since 9 is the same as $10 - 1$, this problem has the same answer as $8 (10 - 1)$ or 8×10 minus 8×1. Or that 9×8 is the same as 8×9, so if you know one, you know the other because they are equivalent.)

The attempt to link computation to number properties is admirable, but several problems were noted. One mathematician observed that students should not be taught that 9 × 3 is the same as 3 × 9. They are, in fact two different problems with the same answer. One refers to 9 sets of 3 units, the other to 3 sets of 9. For students to succeed in algebra, they must understand this difference and remember that two things may look very different and represent very different type of problem types but still have the same answer. The work on multiplication combinations could have resulted in intense work on applications of the commutative, associative and distributive properties of numbers, but based on the text of the article, it did not appear to do so. The importance of doing so for students with LD and other students with learning problems is critical. In contrast, the treatment in Woodward and Brown (2006), developed by the same author, appeared to offer a much richer mathematical menu to students.

A similar concern was expressed about the pre-algebra material used in Witzel et al. (2003). Algebra was taught only on the procedural level. The importance of understanding the nature of a defining variable appeared to be underdeveloped, as did the potential richness of the concrete and visual representations and their link to sets of story problems. Similar concerns were raised about the CRA research of Butler et al. (2003), where numerous opportunities to explore rich mathematical ideas were lost.

These are among the more ambitious studies in the set reviewed, and among the few that really try to delve into complex mathematical topics and concepts. Each of the studies demonstrated some success in reaching the population. However, more intensive collaboration with research mathematicians who know the underlying mathematics in the K–8 curriculum can result in even richer, more effective intervention research for these students.

The research mathematicians also noted that although the two studies that attempted to teach story problems to students (Xin et al., 2005; Fuchs et al., 2005) did not really teach problem solving in the sense that NRC (2001) defined it. However, the studies seemed to be solid attempts to help students understand how to use the mathematics they already knew in an increasing array of applications.

2. Conclusions

On a positive note, many of the studies seriously address two areas of extreme difficulty for students with LD and low-achieving students: application of mathematics to word problems and building of quick retrieval of basic arithmetic combinations. These two areas are essential components of any serious mathematics intervention for these students and we now possess several evidence-based approaches for addressing these areas.

It becomes difficult to conclude easy generalizations about the set of studies. A terse summary would be that explicit instruction is effective (often highly effective) in both domains. In addition, more implicit instructional approaches such as strategic use of concrete objects and visual representations shows some promise, although the number of studies supporting this approach is small, and results are not consistent. Finally, approaches that encourage students to think aloud as they solve problems seem to produce significant positive effects. The drawback in these generalizations is that these terms mean different things to different people. Thus, in translating these findings into practice, effects may be highly dependent on how these instructional principles are conceptualized and how carefully they are incorporated into instruction.

One process that might ease the transition is that many of the interventions included specific scripts for teachers to use for their lessons (although they were usually told to use these as a guide rather than an ironclad script). These could serve as templates for lessons as districts develop various professional academies and training institutes.

Moreover, some other barriers to translating the research findings into classroom practice are as follows including the lack of sufficient specificity concerning the actual content of the mathematics instruction that was provided, which makes replication and extension of the current studies difficult. From a pragmatic standpoint, this is understandable given the need for authors to both describe the instructional sequence and content while leveraging page length. Also, the reviewed studies tend to use either a criterion-referenced test with items that are not presented or comprehensively defined or a standardized achievement test. A notable problem with standardized achievement tests is that they are composed of items from many domains of mathematics skill (e.g., basic computation, long division, fractions) and therefore tend to provide limited specificity concerning the actual mathematics content that students have mastered (Geary, 2005; Hecht, 1998). Finally, criteria for identifying and including low-achieving students examined in these studies were not consistent, which makes generalization of findings uncertain. Most of the studies utilize quite systematic instruction, with high degree of structure and a deliberate pace. This degree of explicitness and detail seems critical for this group of students. Our hope is that research and development efforts will continue to incorporate these elements into instructional materials that can be used with students with learning disabilities and low achievement in mathematics.

BIBLIOGRAPHY

Badian, N.A. (1983). Dyscalculia and nonverbal disorders of learning. In H.R. Myklebust (Ed.), *Progress in learning disabilities* (pp. 235–264). New York: Stratton.

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

Barbaresi, W.J., Katusic, S.K., Colligan, R.C., Weaver, A.L., & Jacobsen, S.J. (2005). Math learning disorder: Incidence in a population-based birth cohort, 1976–82, Rochester, Minn. *Ambulatory Pediatrics*, *5*, 281–289.

Baxter, J., Woodward, J., Wong, J., & Voorhies, J. (2002). We talk about it, but do they get it? *Learning Disabilities Research and Practice*, *17(3)*, 173–185.

Beirne-Smith, M. (1991). Peer tutoring in arithmetic for children with learning disabilities. *Exceptional Children*, *57(4)*, 330–337.

Bereiter, C., & Engelmann, S. (1966). *Teaching disadvantaged children in the preschool*. Englewood Cliffs, NJ: Prentice Hall.

Bradley, R., Danielson, R.L., & D.P. Hallahan, D.P. (2002). *Identification of learning disabilities: Research to practice.* Mahwah, NJ: Erlbaum.

Brown, A.L., & Campione, C. (1990). Communities of learning and thinking: A context by any other name. *Contributions to Human Development, 21*, 108–126.

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 & Practice, 18(20),* 99–111.

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

Carnine, D.W., & Stein, M. (1981). Organizational strategies and practice procedures for teaching basic facts. *Journal of Research in Mathematics Education, 12(1),* 65–69.

Darch, C., Carnine, D., Gersten, R. (1984). Explicit instruction in mathematics problem solving. *Journal of Educational Research, 77(6),* 351–359.

DiPerna, J.C., Lei, P., & Reid, E.E. (2007). Kindergarten predictors of mathematical growth in the primary grades: An investigation using the early childhood longitudinal study-kindergarten cohort. *Educational Psychology, 99(2),* 369–379.

Donovan, M.S., & Bransford, J.D. (2005). Introduction. In the National Research Council (Ed.), *How students learn: Mathematics in the classroom* (pp. 1–30). Washington, DC: The National Academy Press.

Engelmann, S., & Carnine, D.W. (1982). *Theory of instruction: Principles and applications*. 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 disorder. Unpublished doctoral dissertation, Columbia University Teachers College.

Fuchs, D., Fuchs, L.S., Mathes, P.G., & Lipsey, M.W. (2000). Reading differences between low-achieving students with and without learning disabilities: A meta-analysis. In R. Gersten & S. Vaughn (Eds.*), Issues and research in special education* (pp. 81–104). Hillsdale, NJ: Erlbaum.

Fuchs, L.S., Compton, D.L., Fuchs, D., Paulsen, K., Bryant, J.D., & Hamlett, C.L. (2005). Responsiveness to intervention: Preventing and identifying mathematics disability. *Teaching Exceptional Children, 37(4),* 60–63.

Fuchs, L.S., Compton, D.L., Fuchs, D., Paulsen, K., Bryant, J.D., & Hamlett, C.L. (in press). The prevention, identification, and cognitive determinants of math difficulty. *Journal of Educational Psychology*.

Fuchs, L.S., & Fuchs, D. (1998). Peer-assisted learning strategies in math for kindergarten through grade 6. Unpublished article available from Lynn Fuchs, 328 Peabody, Vanderbilt University, Nashville, TN 37220 (lynn.fuchs@vanderbilt.edu).

Geary, D.C. (1993). Mathematical disabilities: Cognitive, neuropsychological, and genetic components. *Psychological Bulletin, 114(2),* 345–362.

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

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

Geary, D.C., Hoard, M.K., Byrd-Craven, J., Nugent, L., & Numtee, C. (2007). Cognitive mechanisms underlying achievement deficits in children with mathematical learning disability. *Child Development, 78(4),* 1343–1359.

Gersten, R., & Carnine, D. (1984). Direct instruction mathematics: A longitudinal evaluation of low-income elementary school students. *Elementary School Journal, 84(4),* 395–407.

Gersten, R., & Chard, D. (1999). Number sense: Rethinking arithmetic instruction for students with mathematical disabilities. *Journal of Special Education, 33(1),* 18–29.

Gersten, R., Clarke, B., & Mazzocco, M. (2007). Historical and contemporary perspectives on mathematical learning disabilities. In D. B. Berch, & 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. Brookes Publishing Company.

Gersten, R., Jordan, N.C., & Flojo, J. (2005). Early identification and interventions for students with mathematics difficulties. *Journal of Learning Disabilities, 38(4),* 293–304.

Goldman, S.R., & Pellegrino, J.W. (1987). Information processing and educational microcomputer technology: Where do we go from here? *Journal of Learning Disabilities, 20(3),* 144–154.

Good, T., & Grouws, D. (1977). Teaching effects: A process-product study in fourth grade mathematics classrooms. *Journal of Teacher Education, 28(3),* 49–54.

Griffin, S., Case, R., & Siegler, R. (1994). Rightstart: Providing the central conceptual prerequisites for first formal learning of arithmetic to students at risk for school failure. In K. McFilly (Ed.), *Classroom Lessons: Integrating cognitive theory and classroom practices* (pp. 25–49). Cambridge: MIT/Bradford Books.

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

Hallahan, D.P., & Kauffman, J.M. (1986). *Exceptional children* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.

Hasselbring, T.S., Goin, L.I., & Bransford, J.D. (1988). Developing math automaticity in learning handicapped children: The role of computerized drill and practice. *Focus on Exceptional Children, 20(6),* 1–7.

Hasselbring, T.S., Lott, A., & Zydney, J. (2005). Technology supported math instruction for students with disabilities: Two decades of research and development. Retrieved on November 1, 2007 from http://www.citeducation.org/mathmatrix/default.asp.

Hecht, S.A. (1998). Toward an information processing account of individual differences in fraction skills. *Journal of Educational Psychology, 90(3),* 545–559.

Hecht, S.A., Vagi, K.J. & Torgesen, J.K. (2007). Fraction skills and proportional reasoning. In Daniel B. Berch & Michele M.M. Mazzocco (Eds.), *Why is math so hard for some children?* Baltimore, MD: Brookes Publishing Company.

Hoge, R.D., & Coladarci, T. (1989). Teacher-based judgments of academic achievement: A review of literature. *Review of Educational Research, 59(3),* 297–313.

Hutchinson, N.L. (1993). Effects of cognitive strategy instruction on algebra problem solving of adolescents with learning disabilities. *Learning Disability Quarterly, 16(1),* 34–63.

Jitendra, A.K., Griffin, C.C., McGoey, K., Gardill, M.C., Bhat, P., & Riley, T. (1998). Effects of mathematical word problem solving by students at risk or with mild disabilities. *The Journal of Educational Research, 91(6),* 345–355.

Jordan, N.C., Hanich, L.B., & Kaplan, D. (2003). A longitudinal study of mathematical competencies in children with mathematics difficulties versus children with co-morbid mathematics and reading difficulties. *Child Development, 74(3*), 834–850.

Kolligian, J., & Sternberg, R.J. (1987). Intelligence, information processing, and specific learning disabilities: A triarchic synthesis. *Journal of Learning Disabilities, 20(1),* 8–17.

Kroesbergen, E.H., Van Luit, J.E.H., & Maas, C.J.M. (2004). Effectiveness of explicit and constructivist mathematics instruction for low-achieving students in the Netherlands. *The Elementary School Journal, 104*(3), 233–251.

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.

Lewis, C., Hitch, G.J., & Walker, P. (1994). The prevalence of specific arithmetic difficulties and specific reading difficulties in 9- to 10-year-old boys and girls. *Journal of Child Psychology & Psychiatry & Allied Disciplines, 35(2),* 283–292.

Manalo, E., Bunnell, J.K., & Stillman, J.A. (2000). The use of process mnemonics in teaching students with mathematics learning disabilities. *Learning Disability Quarterly, 23*(2), 137–156.

Mazzocco, M.M.M. (2007). Defining and differentiating mathematical learning disabilities and difficulties. In D. Berch, & M.M.M. Mazzocco (Eds.), *Why is math so hard for some children: The nature and origin of mathematics learning difficulties and disabilities* (pp. 29–47). Baltimore: Brookes Publishing Co.

McLean, J.F., & Hitch, G.J. (1999). Working memory impairments in children with specific arithmetic learning difficulties. *Journal of Experimental Child Psychology, 74(3)*, 240–260.

Meichenbaum, D. (1985). Teaching thinking: A cognitive-behavioral perspective. In S. Chipman, J. Segal, & R. Glaser (Eds.), *Thinking and learning skills (Vol. 2) Research and open questions* (pp. 407–428). Hillsdale, NJ: Lawrence Erlbaum Associates.

Milo, B.F., Ruijssenaars, A.J.J.M., & Seegers, G. (2005). Math instruction for students with special education needs: Effects of guiding versus directing instruction. *Educational and Child Psychology, 22(4),* 68–80.

Montague, M. (1992). The effects of cognitive and metacognitive strategy instruction on the mathematical problem solving of middle school students with learning disabilities. *Journal of Learning Disabilities, 25(4),* 230–248.

Moore, L.J., & Carnine, D. (1989). Evaluating curriculum design in the content of active teaching. *Remedial and Special Education, 10(4),* 28–37.

Murphy, M.M., Mazzacco, M., Hanich, L.B., & Early, M.C. (2007). Cognitive characteristics of children with mathematics learning disability (MLD) vary as a function of the cutoff criterion used to define MLD. *Journal of Learning Disabilities, 40(5),* 458–478.

National Council of Teachers of Mathematics. (2006). *Curriculum focal points for prekindergarten through Grade 8 mathematics: A quest for coherence*. Reston, VA: National Council of Teachers of Mathematics.

National Council of Teachers of Mathematics. (2000). *Principles and NCTM standards for school mathematics.* Reston, VA: Author.

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: National Council of Teachers of Mathematics.

National Research Council. (2001). *Adding it up: Helping children learn mathematics*. J. Kilpatrick, J. Swafford, and B. Findell (Eds.). Mathematics learning study committee, center for education, division of behavioral and social sciences and education. Washington, DC: National Academy Press.

Owen, R.L. & Fuchs, L.S. (2002). Mathematical problem-solving strategy instruction for third-grade students with learning disabilities. *Remedial and Special Education, 23(5),* 268–278.

Pasnak, R., McCutcheon, L., Holt, R.W., & Campbell, J.W. (1991). Cognitive and achievement gains for kindergartners instructed in Piagetian operations. *Journal of Educational Research, 85(1)*, 5–13.

Riley, M.S., Greeno, J.G., & Heller, J.J. (1983). Development of children's problem-solving ability in arithmetic. In H.P. Ginsburg (Ed.), *The development of mathematical thinking* (pp. 153–196). New York: Academic Press.

Rittle-Johnson, B., Siegler, R.S., & Alibali, M. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of Educational Psychology, 93(2)*, 346–362.

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(3)*, 247–256.

Schopman, E.A.M., & Van Luit, J.E.H. (1996). Learning and transfer of preparatory arithmetic strategies among young children with a developmental lag. *Journal of Cognitive Education, 5*, 117–131.

Schunk, D.H., & Cox, P.D. (1986). Strategy training and attributional feedback with learning disabled students. *Journal of Educational Psychology, 78(3)*, 201–209.

Siegler, R.S. (1987). The perils of averaging data over strategies: An example from children's addition. *Journal of Experimental Psychology: General, 116(3)*, 250–264.

Siegler, R.S., & Shrager, J. (1984). Strategy choice in addition and subtraction: How do children know what to do? In C. Sophian (Ed.), *Origins of cognitive skills* (pp. 229–293). Hillsdale, NJ: Erlbaum.

Silbert, J., Carnine, D., & Stein, M. (1989). *Direct instruction mathematics* (2nd ed.). Columbus, OH: Merrill.

Sowder, J.T. Philipp, R.A., Armstrong, Barbara E., & Schappelle, B.P. (1998). *Middle-grade teachers' mathematical knowledge and its relationship to instruction: A research monograph.* Albany, NY: State University of New York Press.

Stebbins, L., St. Pierre, R.G., Proper, E.L., Anderson, R.B., & Cerva, T.R. (1977). *Education as experimentation: A planned variation model* (Vol. IV A–D). Cambridge, MA: Abt Associates.

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(3)*, 277–321.

Swanson, H.L., & Sachse-Lee, C. (2001). Mathematical problem solving and working memory in children with learning disabilities: Both executive and phonological processes are important. *Journal of Experimental Child Psychology, 79(3)*, 294–321.

Thackwray, D., Meyers, A., Schleser, R., & Cohen, R. (1985). Achieving generalization with general versus specific self-instructions: Effects on academically deficient children. *Cognitive Therapy and Research, 9(3)*, 297–308.

Torgeson, J. (1994). Issues in the assessment of executive function: An information processing perspective. In G. R. Lyon (Ed.), *Frames of reference in the assessment of learning disabilities: New views on measurement issues*. (pp. 143–162). Toronto: Brooks.

Tournaki, N. (2003). The differential effects of teaching addition through strategy instruction versus drill and practice to students with and without learning disabilities. *Journal of Learning Disabilities, 36(5)*, 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(2)*, 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.

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

Witzel, B.S., 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(2),* 121–131.

Woodward, J. (2006). Developing automaticity in multiplication facts: Integrating strategy instruction with timed practice drills. *Learning Disability Quarterly, 29(4),* 269–289.

Woodward, J., & Baxter, J. (1997). The effects of an innovative approach to mathematics on academically low achieving students in inclusive settings. *Exceptional Children, 63(3),* 373–388.

Woodward, J., & Brown, C. (2006). Meeting the curricular needs of academically low-achieving students in the middle grade mathematics. *Journal of Special Education, 40(3),* 151–159.

Woodward, J., & Howard, L. (1994). Misconceptions of youth. *Exceptional Children, 61(2),* 126–136.

Woodward, J., & Montague, M. (2002). Meeting the challenge of mathematics reform for students with learning disabilities. *Journal of Special Education, 36(2),* 89–101.

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(3),* 181–192.

IV. 'Real-World' Problem Solving

A. Introduction and Background

Discussion and debate about the place of "real-world" problems in mathematics instruction—both as a site for learning mathematics, and as an outcome—has been a central theme in U.S. mathematics education for more than a century. The earliest goals of mathematics instruction in this country related to practical uses of mathematics, in such areas as shopkeeping, commerce, surveying, banking, and carpentry (see Michalowicz & Howard, 2003). Early U.S. mathematics textbooks are filled with practical problems intended to prepare students to enter the workplace and to address the particular application needs of the growing nation and society. Over the decades this emphasis on practical applications has ebbed and flowed. In the "new math" years (1960s and 70s), the predominant curricular emphasis was on mathematical precision and the structure of mathematics, but, even then, some critics, such as the applied mathematician Morris Kline (1973), continued to call for applications in the school curriculum. Indeed, even amidst the abstract and logic-focused new mathematics materials developed during the post-Sputnik era, there was at least one applications-oriented curriculum, the Unified Science and Mathematics for Elementary Schools (USMES) project.[3]

A resurgence of calls for emphasis on "real-world"[4] problems came in the 1989 Curriculum and Evaluation Standards for School Mathematics, of the National Council of Teachers of Mathematics (NCTM), which argued that "instruction should be developed from problem situations" (NCTM, 1989, p. 11). The document recommends that in the early grades (K–4), most problems used in instruction should arise from "school and other everyday experiences" (p. 23). Progressing through the grades, there should be a balance between "problems that apply mathematics to the "real-world" and problems that arise from the investigation of mathematical ideas" (p. 75), and by high school, even more of the problems can arise from mathematics itself.

The Instructional Practices Task Group begins by summarizing the definitions and operational meanings that researchers and developers have given to the term "real-world" problems. In the next section the Task Group will highlight some of the rationale and justifications that researchers and developers have used when arguing for and against particular characteristics of "real-world" problems and their uses as a part of the school mathematics curriculum. These viewpoints are sometimes based in research, and sometimes are more directly tied to experience and expert judgments. They sometimes relate to the question of how important it is for students to be able to apply their mathematical knowledge to particular types of problems as an outcome of schooling is. Finally, a synthesis of the

[3] Launched in 1970 through the Education Development Center, with funding from the National Science Foundation; described as a project designed so "students could carry out long-range investigations of real and practical problems based in their local environment" (http://www.coe.ufl.edu/esh/Projects/usmes.htm).
[4] We will include "real world" in quotation marks throughout because of the ambiguity of definition of this phrase in the current literature.

findings from 21 studies examining the impact of "real-world" problem-based instruction on mathematics learning outcomes is provided, followed by commentary about issues in this body of research, and recommendations.

B. What Do Researchers and Developers Mean by 'Real-World' Problems?

Here the Task Group draws on the conceptualizations of "real-world" problems assembled from both developers of instructional materials and researchers who study the impact of such problems. A serious problem in synthesizing the research in this area is that there is no clear, agreed-upon meaning for "real-world" problems. One characteristic mentioned frequently is the meaningfulness and relevance of the problem to the student audience. The Realistic Mathematics Education (RME) movement, which originated in the Netherlands through the work of the mathematician Hans Freudenthal (1973, 1991), has been influential in school mathematics in some countries. RME emphasizes relevance and the activity of doing mathematics; there are several lines of research that have a basis in the RME movement. For instance, researchers De Bock, Verschaffel, Janssens, Van Dooren, and Claes (2003) discuss "organizing mathematical activities around rich, attractive, and realistic contexts … [not only] aspects of the 'real' social or physical world; they can also refer to imaginary, fairy-like worlds as long as they are meaningful, familiar, and appealing to the students. It is not the amount of realism in the literal sense … but rather the extent to which it succeeds in getting students involved in the problem and engage them in situationally meaningful thinking and interaction" (p. 445). In the same tradition, van Dijk and others (2003) describe problems that "bring pupils into situations that make sense to them and provide them with opportunities to experience mathematics as it was developed in cultural history" (p. 164). Other scholars feel that "real-world" problems should be similar to problems that are encountered in applications beyond school, and that are authentic, for instance problems that are "…embedded in a rich narrative structure" and that may require students to make both mathematical and nonmathematical (e.g., ethical) decisions.

In contrast, the problems typically found in algebra textbooks that are sometimes called "story problems" or "word problems" also are sometimes studied in efforts to look at "real-world" problem solving. Jonassen (2003), for example, defines story problems as those that "typically present a quantitative solution problem embedded within a shallow story context" (p. 267).

A "real-world" oriented curriculum that has been studied by a number of the researchers cited in this paper, and whose authors provide detail about their conceptualization of "real-world" problems, is the Adventures of Jasper Woodbury video series (http://peabody.vanderbilt.edu/projects/funded/jasper/Jasperhome.html). This technology-based series is designed to motivate students by engaging them in the solution of complex, multistep problems. The goals of the materials are to promote problem-finding and to develop problem-solving skills. Each 15–20 minute video segment presents an adventure story which involves solving a challenge. The Cognition and Technology Group at Vanderbilt (CTGV) identifies "real-world" problems as being complex, which means having multiple steps, requiring integration of mathematical concepts, involving identification of relevant data, and demanding generation of appropriate questions (see CTGV, 1992; Hickey et al., 2001, pp. 613–14). The goal of this type of problem is to allow students to experience some of the ambiguity and complexity, as well as the intellectual excitement, that adults experience when solving actual problems involving mathematics, be it in engineering, business, accounting, architecture, transportation planning, etc. The hope is that, by anchoring mathematical procedures and concepts in an array of actual situations, students will see the value of knowing the procedures and will more likely be able to transfer what they learn in mathematics to actual problems. The terms "anchored instruction," "situated cognition," and "teaching for transfer" often recur in this literature.

In summary, note how diverse these meanings of "real-world" problem solving are in the literature. This creates challenges and opportunities for researchers, who in general could make progress on some of the fundamental "real-world" problem solving questions with more clarity and focus in the operationalization of the terminology.

What are the purported advantages and disadvantages of using various types of "real-world" problems in school mathematics instruction?

There are several related but distinct reasons advanced in both research and other educational rhetoric for including "real-world" problems in the school mathematics curriculum. Those who believe that students' ability to solve "real-world" problems should be an important outcome of school mathematics argue for the inclusion of such problems in the curriculum as preparation. Others contend that "real-world" problems should be in the curriculum because of their potential to engage and motivate students by engaging them in something they see as meaningful and important (see Bransford, Sherwood, Hasselbring, Kinser, & Williams, 1990; Bransford, Vye, Kinser, & Risko, 1990; CTGV, 1991; DeBock et al., 2003). Hiebert et al. (1996) comment that the "mathematics acquired in these realistic situations, proponents argue, will be perceived by students as being useful" (p. 14).

Another reason sometimes given is that such problems, especially when assigned to be done in groups, provide students with opportunities to learn some of the social problem-solving skills they will need to use later in the workplace (see Resnick, 1987b). The Vanderbilt group (CTVG) contends that "'anchor[ing]' or 'situat[ing]' instruction in the context of meaningful problem solving environments … allow[s] teachers to simulate in the classroom some of the advantages of 'in-context' apprenticeship training" (CTGV, 1992, p. 294; also citing Brown et al., 1989).

Finally, there is a view that students' learning and ability to make mathematical connections in the process of applying their knowledge to a wider range of "real-world" problems might be enhanced (see Hiebert et al., 1996). The notions of both near and far transfer in problem solving appear later in the literature synthesis.

Common to these views seems to be the assumption that, by teaching students mathematics through "real-world" problems, and by teaching students to solve such problems in school, students will become better solvers of the types of problems that they might encounter in everyday life or the workplace (see Verschaffel & De Corte, 1997), and that they will develop a genuine disposition to, and interest in, solving such problems. Some research has looked specifically at whether it is indeed the case that the use of "real-world" problems in instruction promotes such outcomes.

Anderson, Reder, and Simon (1996) have countered some of the claims attributed to the proponents of the use of "real-world" problems.[5] They claim that research has indicated that transfer can happen even if students learn in a situation that is not specific to the site of application. They argue further that using such problems can be inefficient: "Often real-world problems involve a great deal of busy work and offer little opportunity to learn the target competencies" (Anderson et al., 1996, p. 9). They also note that research indicates that workplace skills can be learned separately from the social context, and that in some cases they should be.

There is also the question of whether the contexts that developers imagine as being motivational and engaging for students actually are. Some researchers (e.g. Geary, 1995) have suggested that the contexts in which problems are offered may not be that intrinsically motivating to students. Geary emphasizes that making the mathematics interesting and also ensuring that adequate mathematics is learned may require "degrading" the mathematical content in ways that are not satisfactory. Hiebert et al. (1996) advocate the importance of students' "problematizing" mathematics and suggest that the particular context chosen for a problem is not necessarily as important as the way the teacher engages the students: "Given a different culture [valuing reflective inquiry and problematizing], even large-scale real-life situations can be drained of their problematic possibilities. Outside-of-school problems can provide contexts for important mathematical work, but the packaging of the task is not the primary determinant of the engagement" (pp. 16–18).

"Real-world" problems are often considered to be "open-ended," a term that is equally ill-specified in its meaning. Pehkonen (1997) provides some history of the idea of "openness" in mathematics education, citing work initiated in Japan in the 1970s that helped to launch international focus. He defines open problems in contrast to "closed problems" in mathematics, in which in "open problems," the starting situation or the goal is not explained exactly. Such problems then encompass problem situations in which the student must "find" the problem, "what if" problems, and problems in which multiple solution processes are possible. In a 1994 essay, Hung-Hsi Wu uses examples of problems from K–12 mathematics curricula to highlight that in the case of some open-ended problems, teachers are unlikely to know the required mathematics deeply and at the same time provide a suitably simplified explanation to students. Others have raised concerns about the adequacy of teachers' knowledge of the nonmathematical contexts—in which some of these problems are embedded—to assess the reasonableness of the problem's assumptions, and about the efficiency of using elaborate "real-world" problems in covering mathematics content.

In summary, even without a consistent definition of the notion of "real-world" problems, there are strongly held and argued positions, founded on a variety of bases, in support of, and critical of, the use of various types of "real-world" problems in school mathematics instruction.

[5] For replies and additional commentary concerning the Anderson et al. (1996) paper, see Greeno (1997) and Cobb and Bower (1999).

C. Research Studies Examining the Impact of 'Real-World' Problems in Mathematics Instruction

Despite the variety of reasons that have been advanced supporting the inclusion of "real-world" problems in mathematics instruction, the available research on the topic that met the Instructional Practices Task Group standards for inclusion addresses only two rather focused types of questions:

- Does the use of "real-world" problems in mathematics instruction, in comparison to typical instructional practice, lead to improved understanding of mathematical ideas, or improved computational performance, or improved mathematics performance? (Using "Real-World" Problems to Teach Mathematical Ideas).

- Does the use of particular instructional strategies to help students learn to solve "real-world" problems, in comparison to other strategies and to typical instructional practice, lead to improved performance on assessments that involve solving "real-world" problems; i.e., can near and far transfer be achieved? (Using Specific Strategies to Improve "Real-World" Problem Solving).[6]

Researchers from cognitive science, psychology, and mathematics education have undertaken a range of studies that examine phenomena related to "real-world" problems in mathematics teaching and learning. Most of this work has been descriptive and is not included in the meta-analytic discussion to follow. However, it could serve as an important basis for clarifying and disentangling the meanings of "real-world" problems as an instructional approach and as an outcome of schooling, and provide insights into the design of interventions and assessments that are focused on "real-world" problems. Ethnographic studies have looked, for instance, at the problem-solving strategies used in practices such as candy-selling, tailoring, carpentry, gardening, etc., and the relationship of such craft knowledge to performance on school-based problem tasks (see Presmeg, 2007). International studies such as the Programme for International Student Assessment (PISA) study provide a snapshot of U.S. students' performance on problem solving. This Organisation for Economic Co-operation and Development (OECD) initiative is a collaborative effort of the OECD member countries to "measure how well students at age 15, and therefore approaching the end of compulsory schooling, are prepared to meet the challenges of today's societies… moving beyond the school based approach towards the use of knowledge in everyday tasks and challenges" (Programme for International Student Assessment, 2003, p. 9). PISA is unique as an international assessment in its explicit effort to assess students' ability to "apply their knowledge and experience to real-world issues" (Programme for International Student Assessment, 2003, p. 9). The 2003 administration of PISA examined mathematical literacy and problem solving, and the performance of U.S. students was lower than the average performance for students from OECD countries (see Lemke et al., 2004). Thus for those

[6] The initial search for and screening of literature was initially only for the use of "real world" problems in instruction (question 1), not as the outcome (question 2). The literature identified for the second question resulted from the first, but we did not go back and systematically search for other studies that might fit this second question.

concerned that ability of U.S. students to solve "real-world" problems is an important outcome of schooling, studies such as PISA indicate that U.S. performance has substantial room for improvement.

Here the Task Group draws on the 21 studies that qualified as Category 1 or Category 2 studies and that examined, with random assignment or quasi-experimental methods, the impact of some type of "real-world" problem instructional intervention on student mathematics learning outcomes. There were 13 studies (five Category 1 and eight Category 2) that examined the effect of using "real-world" problems as the means of instruction on mathematics achievement. An additional eight studies (five Category 1 and three Category 2) examined specific strategies to solve "real-world" problems. The second group did not examine a "real-world" problem instructional intervention. Although many have argued that a major reason for using "real-world" problems in mathematics instruction is to increase interest and motivation, the Task Group did not search for studies that looked at motivation only as an outcome. For those studies of mathematics achievement that did include a motivation outcome, those outcomes are not discussed. The studies are presented according to the two categories mentioned earlier. See Tables 15 and 16 for a summary of the Category 1 studies, including effect size calculations.

D. Using 'Real-World' Problems to Teach Mathematical Ideas

The Task Group located 13 studies that introduced some version of a "real-world" problem instructional treatment, and that compared outcomes on student performance in mathematics. Five of these can be considered Category 1 studies for which effect sizes could be computed. Four of these studies contrast some type of "real-world" problem-based instruction with more typical mathematics instruction (although even this varies to some degree). The fifth is concerned with contrasting two different approaches to using "real-world" problems as an instructional approach. All employ outcome measures that assess mathematics performance on what might be considered "typical" types of school mathematics outcomes. In addition, some include outcome measures for transfer, or involving contextualized problems.

Three of the Category 1 studies in this area focus on computations with fractions. Anand and Ross (1987) developed three versions of an intervention aimed at teaching fifth- and sixth-graders how to divide fractions. The operationalization of "real-world" problem solving involved two ways of contextualizing problems: by adding such student-specific information as name, favorite candy bar, etc. into problems, or by simply providing a concrete context for a computational problem. The intervention was a CAI unit that included a "review of prerequisite mathematics facts… introduced the rule for dividing fractions and demonstrated its application to an example problem by using the following four-step solution.… This rule application was repeated for four additional problems" (p. 73). The treatments varied by changing the contexts for the learning material; there were abstract contexts provided, concrete contexts, and personalized contexts based on a biographical questionnaire for the students. The posttest involved context problems similar to those presented in the practice examples, transfer problems, and recognition memory of the rule definition and steps problems. Ninety-six students were randomly assigned to the four treatment conditions (control; concrete; personalized; abstract).

Table 15: Studies That Examine Use of "Real-World" Problems in Mathematics Instruction

Study	Design	Sample	Duration/Content	Contrast	Measure	Hedge's g	Standard Error	
Contextualized Mathematics Outcomes								
Anand & Ross, 1987	RCT	96 students in fifth or sixth grade attending a university affiliated elementary school that emphasized individual learning and progression	One lesson/ Division of fractions	Concrete and Personalized vs. Abstract	Transfer subtest	0.379	(ns)	0.249
Bottge & Hasselbring, 1993	RCT	36 students in two ninth-grade remedial mathematics classes in one Midwest high school	5 days/ Adding and subtracting fractions in relation to money and linear measurement	Contextualized problems vs. Word problems	Contextualized problem test	1.009	**	0.385
Bottge, 1999[a]	RCT	49 middle school average-achieving students in two intact pre-algebra classes	10 school days/ Story problems and transfer problems involving fraction computation	Contextualized problems vs. Word problems	Contextualized problem test	1.131	(ns)	0.693
Brenner et al., 1997[a]	RCT	128 seventh- and eighth-grade students in six intact pre-algebra classes at three junior high schools in a small urban area in Southern California	1 month/ Meaningful thematic contexts and other features	Anchored instruction vs. Traditional textbook	Pooled word problem representation (ES = -0.281), function word problem representation (ES = 0.877), and function word problem solution (ES = 0.393) tests	0.631	(ns)	0.402
Heterogeneity								
Q-value	*df (Q)*	*P-value*	*I-squared*	*Pooled ES (4 studies, 4 effect sizes)*		*Hedge's g*	*Standard Error*	
2.499	*3*	*0.475*	*0.000*			*0.616*	*****	*0.179*
Standard Mathematics Outcomes								
Anand & Ross, 1987	RCT	96 students in fifth or sixth grade attending a university affiliated elementary school that emphasized individual learning and progression	One lesson/ Division of fractions	Concrete and Personalized vs. Abstract	Pooled context (ES = 0.931***) and recognition (ES = 0.727**) subtests	0.828	**	0.257
Bottge & Hasselbring, 1993	RCT	36 students in two ninth-grade remedial mathematics classes in one Midwest high school	5 days/ Adding and subtracting fractions in relation to money and linear measurement	Contextualized problems vs. Word problems	Word problem test	-0.553	(ns)	0.368
Bottge, 1999[a]	RCT	49 middle school average-achieving students in two intact pre-algebra classes	10 school days/ Story problems and transfer problems	Contextualized problems vs. Word problems	Pooled computation (ES = 0.049) and word problem tests (ES = -0.198)	-0.124	(ns)	0.683
Brenner et al., 1997[a]	RCT	128 seventh- and eighth-grade students in six intact pre-algebra classes at three junior high schools in a small urban area in Southern California	1 month/ Meaningful thematic contexts and other features	Anchored instruction vs. Traditional textbook	Pooled equation solving (ES = -0.281) and word problem solving (ES = 0.110) tests	-0.086	(ns)	0.399
Heterogeneity								
Q-value	*df (Q)*	*P-value*	*I-squared*	*Pooled ES (four studies, four effect sizes)*		*Hedge's g*	*Standard Error*	
10.835	*3*	*0.013*	*72.312*			*0.066*	*(ns)*	*0.374*
A Study that Examined Two Different Approaches to "Real-world" Problem-based Instruction								
van Dijk et al., 2003[a]	RCT	238 fifth-grade students in 10 classes in the Netherlands	13 lessons in 3 weeks/ "Real-world" problems that entail division with a remainder	Student vs. Teacher constructed models	Curriculum specific posttest	0.402	(ns)	0.307

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

[a] Data were adjusted for clustering that occurred within classrooms.

Several contrasts (comparisons of results for two different treatment conditions) resulted in statistically significant and meaningful effect sizes: most striking was the significantly better performance of the personalized group than the abstract group[7] on the context and recognition measures (effect sizes = 1.434 and 1.116, respectively). The concrete and personalized treatment group also performed significantly better than the abstract treatment group on the context and recognition measures (effect sizes = 0.931 and 0.727, respectively). Of the nine effects computed for this study,[8] five produced effect sizes significant at the .05 level or better. All favored the personalized or personalized and concrete treatments over the abstract, with the strongest differences on the context and recognition outcome measures. Effect sizes on the transfer measure for the personalized group in comparison with the abstract group also was significant, with an effect size of 0.630, and combining the concrete and personalized treatments and comparing to the abstract also yielded an appreciable though non-significant effect size (0.379) on the transfer measure. For all of these measures the combined effect size is 0.679, which is significant.

The Task Group notes that the use of "contextualized" in the Anand and Ross study is a narrowly focused operationalization of "real world." "Contextualization" occurred by personalizing the problems through such means as using the students' names or interests within the problems. This would not fit most of the operational meanings for "real-world" problems discussed earlier. Nonetheless, the strong effects on context and recognition problems is interesting, and suggest that a very specifically focused type of contextualization can be more effective on context and recognition outcomes than abstract presentation of problems, and can have some effect on transfer.

Bottge and his colleagues have published two studies that met the Category 1 criteria (Bottge & Hasselbring, 1993; Bottge, 1999). Both studies pursue questions about the effect of "contextualized mathematics instruction" on the problem-solving performance of middle school and ninth grade students. The instructional interventions are video-based problem solving materials based on the principles that guided the Cognition and Technology Group at Vanderbilt (CTGV) in the design of the *Adventures of Jasper Woodbury series*. These include commitment to "guidance by an effective teacher; a rich, realistic source of information; and a meaningful problem-solving context" (Bottge & Hasselbring, 1993, p. 5).

In the 1993 study, 36 students in two ninth-grade remedial mathematics classes were assigned to treatment and control conditions, where the instruction was focused on problem solving in the area of fraction addition and subtraction. The students had experienced behavioral or academic difficulties. Students were compared on their ability to solve a contextualized problem following instruction. All students received review in fraction computation skills for five days prior to the intervention. The intervention was then an additional five days of problem solving that employed, for the "contextualized problem" (CP) group, an 8-minute contextualized problem presented via videodisc called *Bart's Pet Project*. The "word problems" (WP) condition received a series of standard word problems in instruction. In both conditions the students were guided to solve the problems by their teachers.

[7] Contrasts not included in table because of the need to use only independent contrasts in pooling effect sizes from multiple contrasts.
[8] There were three contrasts, each with three measures. Only one is included in the table. See above footnote.

Effect sizes were computed on two outcome measures; a word problem test and a contextualized problem-solving test administered via video. The effect of the CP condition in contrast with the WP condition on the contextualized problems was significant (ES = 1.009). The effect size on the word problem measure, perhaps a more typical school mathematics outcome measure, was not significant and favored the control group (ES = -0.553). It is possible that the video-based contextual outcome measure was overly aligned with the treatment and the word problem measure was overly aligned with the control, which makes the results unsurprising. Nonetheless, the study seems to demonstrate that with this particular group of at-risk students, the "real-world" treatment can make a difference on the transfer task.

In 1999, Bottge again looked at the effect of contextualized mathematics instruction on the problem-solving performance of middle school students. The topic was fraction computation, and the interventions involved two video-based contextualized mathematics problems, both in the spirit of the CTGV-designed *Adventures of Jasper Woodbury*. The control treatment was a more standard presentation of word problem instruction, using problems parallel to those in the video materials. Outcome measures included computation, word problems, a contextualized problem, and an applied transfer task. There was a noteworthy but nonsignificant effect size (1.131), favoring the treatment groups, on the contextualized problem. It is worth noting that on the computational outcome there was a slight but not statistically significant advantage for the control group (ES = -0.049) and similarly, on the word problem outcome measure, the nonsignificant effect size (-0.198) slightly favored the control group. Note also that the mathematical content of this instruction is not standard ninth-grade content and that students were provided with review on the procedural aspects of fraction computation prior to the intervention.

The fourth study in this group (Brenner et al., 1997), focused on student understanding of key pre-algebra ideas such as the functional relationship between two variables, and contextual translation and application. A unit emphasizing meaningful thematic contexts and other features (thereby possibly confounding the "real-world" emphasis with other characteristics) was developed and used in three pre-algebra classes, and the control condition was three pre-algebra classes using a traditional algebra textbook. The effect size for the anchored instruction treatment in contrast with the traditional condition on solution to the function word-problem test was appreciable, though not significant (ES = 0.393). The effects on the word problem and equation solving measures were notably nonsignificant and slightly favored the control group. So, in this case, the influence of the treatment on the mathematical content that was especially aligned with the treatment was the strongest.

The final study in this group is of a different type. A group of researchers in the Netherlands (van Dijk, van Oers, Terwel, & van den Eeden, 2003) undertook a study with fifth-grade students to compare two different approaches to "real-world" problem-based instruction, in the spirit of the Dutch Realistic Mathematics Education (RME) movement of teaching through problems. Two different approaches to mathematization, or modeling (a type of "real-world" problem instruction) were used. The experimental treatment was called "guided co-construction," where during instruction on the topic of percentages and graphs, students were guided by teachers to create their own models of the problems that were serving as the foundation of instruction. This was compared to a more traditional (within RME) expository approach, in which teachers provide students with models for the instructional problems. The posttest "measured the pupils' achievement regarding percentages and graphs in a quantitative way" (p. 177) which the Task Group takes to be a measure of what are typical school mathematics outcomes for the Dutch context, rather than a transfer task. The effect size favoring the guided co-construction group was not significant but encouraging (ES = 0.402).

In some of these studies the medium for introducing contextualized problems is video-based material, and the outcome measure is also video-based, causing over-alignment of the treatment with the outcome measure. Although the novelty of using video is not mentioned as a possible confound for studies of this type, it is worth considering how this particular instructional approach may affect students' interest and engagement.

The Task Group also calculated a pooled effect size across the four studies that are most similar (Anand & Ross, 1987; Bottge, 1999; Bottge, & Hasselbring, 1993; and Brenner et al., 1997) on the contextualized mathematics outcomes; see Table 15. Using the pooled measures in each of these studies, the pooled effect size was 0.616 and statistically significant. Thus the meta-analysis suggests that the impact of using "real-world" contexts in mathematics instruction on mathematics performance on similar "real-world" problems is significant. And, the impact on performance on other areas of mathematics, including computation, simple word problems, and equation solving, is not, at least when using this small set of studies as evidence. There are a number of caveats to be considered here; only four studies, all of them somewhat different, were included. And, the outcome measures are a mix of what might be thought of as "typical" mathematics measures, as well as more specialized transfer measures of contextualized or "real-world" problem solving.

In summary, the findings from these five studies, taken together, indicate that under certain conditions, the effect of treatments that employ contextualized problems in instruction on performance on contextual problems involving particular areas of mathematics can be significant. The results of these studies cannot be considered conclusive in providing direction on the general question of the use of "real-world" problem solving as a strategy for improving mathematics learning. However, they do suggest that certain well-defined "real-world" problem solving approaches can lead to improved performance on specific outcome measures, both for typical school mathematics performance, and more strongly, for transfer to "real-world" problem solving.

There were eight additional studies[9] that were classified as Category 2 but which will be discussed here because they provide additional insight into what has been learned from the Category 1 studies, or because they raise other interesting research issues. There were various, distinct flaws in these studies. For instance, in some, the use of "real-world" problems in instruction is confounded by concomitant instructional interventions, such as use of small groups, or emphasis on exploration in the curriculum, or inclusion of student writing as an instructional strategy. Other flaws also occurred, including the use of volunteer teachers in the treatment conditions, lack of matched control groups, lack of evidence of testing of

[9] Ben-Chaim, Fey, Fitzgerald, Benedetto, & Miller, 1998; Bottge Rueda, Serlin, Hung, & Kwon, 2007; CTGV, 1992; DeBock, Verschaffel, Janssens, Van Dooren, & Claes, 2003; Hickey et al., 2001; Henderson & Landesman, 1995; Irwin, 2001; Klein, Beishuizen, & Treffers, 1998.

group equivalence, outcome measures overly aligned with the treatment or control conditions, or only one unit assigned to each experimental condition. Thus no conclusions about impact of "real-world" problem solving as an instructional approach can be drawn from this set of studies.

Seven of the studies (all but Bottge et al., 2007) are designed to compare some type of "real-world" problem solving treatment to a control condition of typical mathematics instruction. Several use specific curricular interventions as the "real-world" problem intervention. In the case of Ben-Chaim et al. (1998) the treatment involved a full curriculum at the middle grades, the Connected Mathematics Program. A problem in using such a study to examine the specific impact of "real-world" problems is that full curricula such as this employ a range of principles and instructional approaches, and so findings cannot be clearly attributed to any particular component of the intervention. Henderson and Landesman (1995) used a similarly broad intervention (thematically integrated instruction) in their study, thereby making it difficult to interpret the impact of "real-world" problems. Given this, finding appropriate designs and measures that would allow a more focused look at the place of "real-world" problems in curricula that also include other interventions seems a worthwhile direction to pursue.

Klein et al. (1998) is another study in the Dutch context, comparing effects of two different approaches to teaching with realistic problems, where the introduction of flexible solution is handled differently in the treatment and control. This study is noteworthy because of the fine-grained detail in explaining the difference in these two approaches to working with "real-world" problems. The Task Group mentions the CTGV 1992 study because there is a creative outcome measure that has to do with problem solving planning. In the DeBock et al. (2003) research study, there is a very strong initial focus on the mathematical topic (applying linear models), raising the possibility that "real-world" instructional approaches may be better used for the teaching of some specific mathematical ideas rather than others. This study is also interesting because the assessments are varied according to the treatments, in an attempt to compare the impact of different treatments on assessments.

The Task Group also notes a final study that does raise some ideas that are worthy of consideration. In Bottge et al. (2007), the performance of different groups of students who were instructed using Enhanced Anchored Instruction (EAI) is compared. EAI is an instructional approach based on the concept of anchored instruction as advanced by the CTGV, which involves having students solve a problem in a multimedia format and then apply what they have learned in hands-on problem settings, such as building skateboard ramps (p. 32). The mathematical topics in this case involved rates, construction of graphs, lines of best fit, and fraction calculation. In this study, the emphasis is on the possibly differential impact of the "real-world" oriented curriculum on different groups of students (students with learning disabilities, and high- and average-achieving students). This is classified as a Category 2 study because of design issues, but it is an interesting example for consideration. The authors report that, following treatment, students in the inclusive classes (which include learning disabled students) outscore the students in the typical classes. This, together with other studies by Bottge, as well as the study by Henderson and Landesman (1995) that is concerned with bilingual instruction as well as thematic integration, suggests that more systematic research on the impact of "real-world" problem based instruction on particular subgroups of students who have been traditionally underserved in mathematics, may be worthwhile.

Three of the studies examine the impact of video-based instruction that involves the presentation of mathematical problems through "real-world," contextual settings (Cognition and Technology Group at Vanderbilt [CTGV], 1992; DeBock et al., 2003; and Hickey, Moore, & Pellegrino, 2001). Two (CTGV, 1992; Hickey, Moore, & Pellegrino, 2001) report on the impact of the implementation of the *Jasper Woodbury* series. DeBock et al. (2003) use video material based on Gulliver's Travels. There are other innovations in the use of video-based instruction, including involvement of students in cooperative groups, for instance, which can cause confounding; in addition, the types of outcome measures used in these studies vary in terms of their closeness to the focus of the intervention.

Remaining mindful that all of these studies have flaws that prevent their inclusion in Category 1, five of them report significantly better performance of treatment groups (some kind of "real-world" instruction) than of control groups, on particular measures that tend to emphasize "real-world" problems in one way or another (Ben–Chaim et al. 1998; CTGV, 1992; Hickey et al., 2001; Henderson & Landesman, 1995; and Irwin, 2001). In contrast, Klein et al. (1998) report no difference on procedural competence between the control and treatment groups, and DeBock et al. (2003) report a negative result, where students using the video instructional treatment performed worse than those who solved non-embedded problems on the outcome measure.

No conclusions about impact can be drawn from these studies. Instead, they highlight the complexities of these research issues, and point toward interesting questions, designs, and measures that could help form a foundation for subsequent research.

E. Using Specific Strategies to Improve 'Real-World' Problem Solving

The search for studies that examined the use of "real-world" problems as an instructional strategy led to a small number of Category 1 studies that concerned the impact of different instructional strategies for teaching students to solve "real-world" problems. Note that this is not necessarily all of the studies that have examined strategies for improving "real-world" problem solving. This work is distinguished from what is included in the prior section in part by a particularly strong focus on the primary goals of both near and far transfer outcomes. Lynn Fuchs and her research group have undertaken a series of studies in this vein, several of which met our criteria. A study by Fuchs, Fuchs, Hamlett, and Appleton (2002) was aimed at enhancing mathematical problem-solving performance of fourth-graders with mathematical disabilities. All students participated in their regular classroom mathematics instruction using a basal text, and a six-lesson base treatment on approaching mathematical problem solving. One experimental group received 24 sessions of problem-solving tutoring; one received 24 sessions of computer-assisted practice; a third received both the tutoring and computer-assisted practice sessions.[10] Small group tutoring was provided on problem-solving rules and on transfer. The computer-assisted practice emphasized tasks intended to lead to far-transfer. There were three types of outcome measures: ability to solve story problems, transfer story problems, and “real-world” problems. Significant effects favoring the problem-solving tutorial group were found on the story problem and transfer story problem measure (effect sizes of 1.340 and 0.982, respectively). The effect size on the “real-world” problem solving measure was not significant (-0.041) and slightly favored the computer-assisted practice group, indicating no significant differences between treatments on the primary outcome measure of far transfer.

Table 16: Studies That Examine Strategies to Improve “Real-World” Problem Solving

Study	Design	Sample	Duration/ Content	Contrast	Measure	Hedge’s g	Standard Error	
Barron, 2000	RCT	96 sixth-grade students in a public magnet school for academically talented children	Four 1-hour sessions/ Contextual problem solving	Problem solving collaboratively in triads vs. Problem solving individually	Pooled transfer measures	0.287	(ns)	0.204
Fuchs et al., 2002	RCT	40 fourth-grade students with mathematical disabilities in six classrooms in three schools	24 sessions/ Mathematical problem solving	Problem-solving tutoring vs. Computer-assisted practice	Pooled story problems (ES = 1.340**), transfer story problems (ES = 0.982**) and “real-world” problem-solving measures (ES = -0.041)	0.760	~	0.454
Fuchs et al., 2004[a]	RCT	351 third-grade students in 24 classrooms in seven schools in an urban district	34 lessons over 16 weeks/ Mathematical problem solving	SBTI vs. Control	Transfer-4 measure (a measure of far transfer that approximated real life problem solving)	1.123	*	0.513
				SBTI expanded vs. Control		2.087	***	0.600
Fuchs et al., 2006[a]	RCT	445 third-grade students in 30 classrooms in seven schools in an urban district	16 weeks/ Mathematical problem solving strategies	SBI vs. Control	Pooled transfer measures	0.545	(ns)	0.439
				SBI-RL vs. Control		1.077	*	0.464
Rudnitsky et al., 1995	RCT	401 third- and fourth-grade students in 21 classrooms in six schools	18 days/ Addition and subtraction word problems	Writing and discussion vs. Practice and explicit heuristics	Near transfer posttest	0.190	~	0.115

~ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$
[a] These studies use classroom-level analyses.

[10] Contrasts with this condition not included in table.

A second study (Fuchs, Fuchs, Finelli, Courey, & Hamlett, 2004) built on these findings and implemented schema-based transfer instruction (SBTI), which explicitly teaches students about transfer features of problems in an effort to improve their near- and far-transfer performance. Twenty-four third-grade teacher volunteers in seven urban schools were randomly assigned to one of three conditions: control, SBTI, and SBTI expanded (this included focus on additional and challenging superficial problem features such as irrelevant information, and the concept of "real-life" situations that introduce more information than problems typically used in school). The 16-week treatments were compared using four outcome measures: Transfer-1 (novel problems structured in the same was as those in the instruction); Transfer-2 (novel problems that varied in the three transfer features taught in SBTI); Transfer-3 (novel problems that varied in transfer features taught in both SBTI and expanded SBTI); and Transfer-4 (measure of far transfer that varied from the problems used in instruction in six major ways). Calculation of effect sizes for Transfer-4 (measure of far transfer that approximated "real-life" problem solving) yielded significant effects for the SBTI expanded vs. the control condition (ES = 2.087), and for the SBTI vs. the control group (ES = 1.123). The Task Group can conclude that this particular, highly specific instructional approach can result in stronger performance on a "real-world" problem outcome.

This group of studies led to a randomized controlled study published in 2006 (Fuchs, Fuchs, Finelli, Courey, Hamlett, Sones et al., 2006). Three treatment conditions were implemented: the "teacher-designed" condition, which was the control, with teachers using the district curriculum; and two schema-broadening instruction (SBI) conditions. One SBI condition was the problem-solving instruction used in earlier studies, emphasizing superficial problem features, problem structures, and problem types. The second SBI condition is expanded to include "explicit instruction in strategies for tackling the complexities involved in real-life problems" (p. 296). The Task Group found a significant effect size (ES = 1.077) for the enhanced schema-broadening instruction aimed at preparation for solving real-life problems in comparison to the control group members on pooled far transfer measures. In addition, the effect of the SBI treatment in comparison to the control on the far transfer outcome measures was encouraging, though not significant (ES = 0.545). The results of these two studies (Fuchs et al. 2004, Fuchs et al. 2006), in contrast to the Fuchs, Hamlett, & Appleton et al. (2002) suggest that the enhanced schema-broadening instruction, which explicitly helps students to recognize and attend to irrelevant and extraneous features in real-life problems, is effective in enabling students to successfully solve real-life transfer problems.

Two additional studies met the Category 1 criteria and have been classified as being about promoting student performance on "real-world" problems through a particular instructional strategy. Barron (2000) used *Jasper Woodbury*-style video-based microworlds as the instructional treatment being tested, in comparison to control conditions, for its effect on a student problem solving performance measure. Two different grouping strategies for students using the *Jasper Woodbury* materials were compared. In one condition, the sixth-grade students worked in triads. In the other, they worked individually. The effect size calculation yielded an encouraging though not significant effect of the triad arrangement (ES = 0.287). Rudnitsky, Etheridge, Freeman, and Gilbert (1995) focused on helping third- and fourth-grade students solve arithmetic word problems through two different treatments: a "writing-to-learn" approach, in which students created their own mathematical stories and problems, and a control condition involving practice and explicit heuristics. On a near-transfer problem-solving posttest, there was no significant effect size (ES = 0.190).

In summary, these five studies examine several very different types of instruction intended to improve near or far transfer performance on "real-world" problems. The strategies were: student grouping, computer-assisted instruction, problem-solving tutoring, schema-based instruction, enhanced schema-based instruction, and problem writing. It is not reasonable to calculate pooled effect size for these five studies, given the differences in the instructional interventions. It is important to note that, of all of these strategies, the only one that shows promise on an empirical basis is the enhanced schema-based instruction in both Fuchs et al. (2004) and Fuchs et al. (2006). Note too that the mathematical domain is narrow (whole number arithmetic) and this was undertaken only at the third grade. At the same time, the heart of the intervention—a focus on extraneous and irrelevant information—is a feature that some would surely say is a defining feature of "real-world" problems; these are messy problems. Fuchs and her colleagues seem to have demonstrated that, under very specific conditions, in a very narrow area of mathematics, it is possible to teach students how to address these issues and be effective problem solvers.

There were three studies classified as Category 2 that also examine instructional strategies for improving performance on "real-world" problems. All of them have design flaws which exclude them as studies from which the Task Group can draw conclusions about impact. However, these studies are instructive because they provide ideas about various kinds of instructional interventions that have been attempted, and about interesting outcome measures. Serafino and Cicchelli (2003) contrast two instructional approaches within the anchored instruction model on which the Jasper materials are based. The Structured Problem Solving model includes a more teacher-dominated, structured approach, with more focused teacher questions and summaries. The Guided Generation model casts the teacher in more of a facilitator role. Shyu (1999) investigated the effects of a video-based anchored instruction program based on the Jasper Series. There were three instructional treatments—the video-based instruction, the printed, story-book version, and regular instruction. Both of these studies designed alternative instructional approaches for use with problem-based curricula.

Verschaffel and DeCorte (1997) conducted an experiment with 10–12 year olds in which the treatment involved a sequenced introduction of "real-world" problems and discussion of the information available and the approach to the problem. What is of particular interest in this study is the outcome measure, "disposition toward realistic modeling," which is intended to assess students' tendency to use "real-world" knowledge and realistic considerations in their problem solving. The authors report a significant difference favoring the treatment group, but these results are not robust given that the treatment condition had only one unit. Nonetheless, it is worth noting that the instructional interventions in this study seem to share the principles in the Fuchs et al. approaches, in which there is explicit focus on features characteristic of "real-world" problems. Further development of such approaches for use in wider contexts might be fruitful.

Based on an examination of studies primarily concerned with testing different approaches to teaching that enable students to solve “real-world” problems, the Task Group concludes that some instructional approaches will lead to better student performance on “real-world” problems than others. However, these instructional practices are highly specified, and the studies only demonstrate their effectiveness for relatively narrow classes of problems. For those who view performance on “real-world” problems as an important outcome of K–12 mathematics education, there are still far more open questions about what will lead to far transfer and which instructional methods are best than there are conclusions.

F. Conclusion

It is difficult to draw conclusions from the set of studies that examine the impact of the use of “real-world” problems and related instructional strategies in instruction on student mathematics performance, including performance on “real-world” problems. The body of studies is small; the outcome measures are often designed by the researchers and information is not available on psychometric characteristics of these measures; and, confounding variables that are difficult to measure reliably, such as fidelity of implementation and other contextual features, are not always included in the study reports.

The set of studies also has a certain homogeneity. Of the 21 studies discussed here, 10 of them are focused on instructional materials that introduce “real-world” problems through the Jasper Woodbury series, or similar video materials. Researchers have not undertaken the necessary rigorous examination of print instructional materials that have as their primary goal the introduction of mathematical ideas through “real-world” problems. Nor has there been adequate attention to the possibility that different mathematical ideas, topics, and procedures might best be learned through particular instructional approaches; perhaps using “real-world” problems is good for some mathematical topics and not for others. The Task Group found very few studies that started from any clear hypotheses about why a particular intervention would be likely to help with a particular area of mathematics.

Debates about the place of “real-world” problems in the mathematics classroom are complicated by a number of issues; the operationalization of the term “real-world” problems varies by mathematician, researcher, developer, and teacher; fidelity of the teachers’ implementation of the instructional materials or instructional strategy is difficult to assess; contextual features, such as SES, or the school’s orientation toward reform matter; and most likely, although not addressed in the studies the Task Group examined, teachers’ knowledge and capacity to use such problems effectively varies greatly.

A particularly relevant issue to focus on in this domain may be the degree to which students’ ability to apply mathematical knowledge to “real-world” or “authentic” problem situations is a valued and agreed-upon outcome of school mathematics. If “real-world” problem solving is not seen as an essential outcome of K–12 mathematics education, then the modest accumulation of research available (meeting our screening criteria) on the topic would suggest that there is no great value in using “real-world” problems as a main element of mathematics instruction nor is there great value investing significant time to design effective instructional strategies that rely on “real world” contexts. However, if ensuring that students know the needed mathematics, and can apply that mathematics to the more complex and open kinds of problems that can be encountered in the "real world" is important, then the studies reviewed here offer some promise.

The Task Group concludes that, under certain conditions, for specific domains of mathematics, instruction that features the use of "real-world" contexts shows potential promise for having a positive impact on student achievement. However, these results are not yet sufficient as a basis for widespread policy recommendations.

If the goal of application of the mathematical knowledge in contexts is considered important, then these studies would suggest that continued investment in research and development that is coordinated with state standards may be worthwhile, with several caveats. More studies should use standardized outcome measures in place of the researcher or developer-designed instruments, so that the results can accumulate in a more useful way. If such measures are not used, then the design of outcome goals and measures needs more integrated involvement of psychometricians and mathematicians, who can watch for the difficulties of overly confounding the outcome assessment with the intervention, or of assessing mathematics too narrowly. Studies that look beyond special populations of students (e.g., remedial students, special education students) are needed. Randomized control experiments are necessary for generalization and clarity about the scale-up potential and outcomes of specific interventions. And, more attention is needed to the specific kinds of mathematical outcomes that are obtained by specific types of "real-world" problem interventions. For instance, "real-world" approaches may be especially useful for introducing particular mathematical concepts and processes, and less useful or inefficient for the introduction of other topics. Thus far the research has made little systematic progress on this matter.

BIBLIOGRAPHY

Anand, P.G., & Ross, S.M. (1987). Using computer-assisted instruction to personalize arithmetic materials for elementary school children. *Journal of Educational Psychology, 79(1),* 72–78.

Anderson, J., Reder, L.M., & Simon, H.A. (1996). Situated learning and education. *Educational Researcher*, *25*, 5–11.

Barron, B. (2000). Problem solving in video-based microworlds: Collaborative and individual outcomes of high-achieving sixth-grade students. *Journal of Educational Psychology, 92(2),* 391–398.

Ben-Chaim, D., Fey, J.T., Fitzgerald, W.M., Benedetto, C., & Miller, J. (1998). Proportional reasoning among 7th grade students with different curricular experiences. *Educational Studies in Mathematics, 36(3),* 247–273.

Bottge, B.A (1999). Effects of contextualized math instruction on problem solving of average and below average achieving students. *The Journal of Special Education, 33*(2), 81–92.

Bottge, B. & Hasselbring, T. (1993). A comparison of two approaches for teaching complex, authentic mathematics problems to adolescents in remedial math classes. *Exceptional Children, 59(6),* 556–566.

Bottge, B.A., Rueda, E., Serlin, R.C., Hung, Y., & Kwon, J.M. (2007). Shrinking achievement differences with anchored math problems: Challenges and possibilities. *Journal of Special Education, 41(1),* 31–49.

Bransford, J.D., Sherwood, R., Hasselbring, T., Kinzer, C., & Williams, S. (1990). Anchored instruction: Why we need it and how technology can help. In D. Nix & R. Spiro (Eds.), *Cognition, education, and multimedia: Exploring ideas in high technology* (pp. 115–141). Hillsdale, NJ: Lawrence Erlbaum Associates.

Bransford, J.D., Vye, N., Kinzer, C., & Risko, V. (1990). Teaching thinking and content knowledge: Toward an integrated approach. In B. Jones, & L. Idol (Eds.), *Dimensions of Thinking and Cognitive Instruction* (pp. 381–413). Hillsdale, NJ: Lawrence Erlbaum Associates.

Brenner, M., Mayer, R., Moseley, B., Brar, T., Duran, R., Reed, B., & Webb, D. (1997). Learning by understanding: The role of multiple representations in learning algebra. *American Educational Research Journal, 34(4),* 663–689.

Cobb, P., & Bower, J. (1999). Cognitive and situated learning perspectives in theory and practice. *Educational Researcher, 28(2),* 4–15.

Cognition and Technology Group at Vanderbilt. (1991). Technology and the design of generative learning environments. *Educational Technology, 31(5),* 34–40.

Cognition and Technology Group at Vanderbilt. (1992). The Jasper Series as an example of anchored instruction: Theory, program description, and assessment data. *Educational Psychologist, 27(3),* 291–315.

De Bock, D., Verschaffel, L., Janssens, D., Van Dooren, W., & Claes, K. (2003). Do realistic contexts and graphical representations always have a beneficial impact on students' performance? Negative evidence from a study on modeling non-linear geometry problems. *Learning and Instruction, 13(4),* 441–463.

Freudenthal, H. (1973). *Mathematics as an educational task.* Dordrecht, The Netherlands: Reidel.

Freudenthal, H. (1991). *Revisiting mathematics education*. Dordrecht, The Netherlands: Reidel.

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 & Practice, 17*(2), 90–106.

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

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

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

Greeno, J.G. (1997). On claims that answer the wrong questions. *Educational Researcher, 26(1),* 5–17.

Henderson, R. & Landesman, E. (1995). Effects of thematically integrated mathematics instruction on students of Mexican descent. *Journal of Educational Research, 88(5),* 290–300.

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

Hiebert, J., Carpenter, T., Fennema, E., Fuson, K., Human, P., Murray, H., et al. (1996). Problem solving as a basis for reform in curriculum and instruction: The case of mathematics. *Educational Researcher, 25(4),* 12–21.

Irwin, K. (2001). Using everyday knowledge of decimals to enhance understanding. *Journal for Research in Mathematics Education, 32(4),* 399–420.

Jonassen, D.H. (2003). Designing instruction for story problems. *Educational Psychology Review,* 15*(3),* 267–296.

Klein, A.S., Beishuizen, M., & Treffers, A. (1998). The empty number line in Dutch second grades: Realistic versus graduate program design. *Journal for Research in Mathematics Education, 29(4),* 443–464.

Kline, M. (1973). *Why Johnny can't add: The failure of the new math*. New York, NY: Random House, Inc.

Lemke, M., Sen, A., Pahlke, E., Partelow, L., Miller, D., Williams, T., et al. (2004). *International outcomes of learning in mathematics literacy and problem solving: PISA 2003 results from the U.S. perspective* (No. NCES 2005-003). Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Michalowicz, K.D., & Howard, A.C. (2003). Pedagogy in text: An analysis of mathematics texts from the nineteenth century. In G. Stanic and J. Kilpatrick (Ed.), *A history of school mathematics* (pp. 77–109). Reston, VA: National Council of Teachers of Mathematics.

National Council of Teachers of Mathematics (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.

Pehkonen, E. (1997). Introduction to the concept of “open-ended problem.” In E. Pehkonen (Ed.), *Use of open-ended problems in the mathematics classroom*. Research Report 176. Helsinki: ERIC. ED 419714.

Presmeg, N. (2007). The role of culture in teaching and learning mathematics. In F. Lester (Ed.), *Second handbook of research on mathematics teaching and learning*. Charlotte, NC: Information Age Publishing.

Programme for International Student Assessment. (2003). *PISA 2003 assessment framework–Mathematics, reading, science and problem-solving knowledge and skills*. Paris: Organisation for Economic Co-operation and Development.

Resnick, L.B. (1987a). *Education and learning to think.* Washington DC: National Academy Press.

Resnick, L.B. (1987b). Learning in school and out. *Educational Researcher, 16*, 13–20.

Rudnitsky, A., Etheredge, S., Freeman, S., & Gilbert, T. (1995). Learning to solve addition and subtraction word problems through a structure-plus-writing approach. *Journal for Research in Mathematics Education, 26(5),* 467–486.

Serafino, K., & Ciccelli, T. (2003). Cognitive theories, prior knowledge, and anchored instruction on mathematical problem solving and transfer. *Education and Urban Society, 36(1),* 79–93.

Shyu, H. (1999). Effects of media attributes in anchored instruction. *Journal for Research in Mathematics Education,* 32(4), 399–420.

The Adventures of Jasper Woodbury. Retrieved November 1, 2007, from http://peabody.vanderbilt.edu/projects/funded/jasper/Jasperhome.html.

van Dijk, I.M.A.W., van Oers, B., Terwel, J., & van den Eeden, P. (2003). Strategic learning in primary mathematics education: Effects of an experimental program in modeling. *Educational Research and Evaluation, 9(2),* 161–187.

Verschaffel, L., & De Corte, E. (1997). Teaching realistic mathematical modeling in the elementary school: A teaching experiment with fifth graders. *Journal for Research in Mathematics Education, 28(5),* 577–601.

Wu, H.S. (1994). The role of open-ended problems in mathematics education. *Journal of Mathematical Behavior, 13(1),* 115–128.

V. The Role of Technology in Mathematics Education

Although young in historic terms, computer technology has a strong presence in our lives and in the research literature. This report synthesizes what is known from high-quality research about the effectiveness of a variety of approaches to applying computer technology to the solution of educational problems in mathematics instruction. The report begins with a brief overview of the categories of computer applications that mathematics educators have used. Next, using the prior reviews, syntheses, and meta-analyses as context and background, the Instructional Practices Task Group's own original meta-analyses of rigorous studies for those categories that included an adequate body of studies that fit the Task Group's criteria are presented. These included drill and practice, tutorials, calculators, and computer programming. The Task Group's basic question is: What is the role of technology including computer software, calculators and graphing calculators in mathematics instruction and learning? The last section summarizes answers to this question on the basis of the Task Group's review of high-quality research.

A. Categories of Instructional Software

As an all-purpose device, a computer can take a variety of forms and play a variety of instructional roles. The term *computer-based instruction* (CBI) will refer to all these applications of computer technology to education. As an interactive device, the computer can be programmed to provide opportunities for active learning and reflective thinking on the one hand, or to provide drills on the other. It might manage and individualize instruction. It can perform tedious calculations, potentially having positive effects (if it thus allowed engagement in topics otherwise impossible or difficult for students to approach) or unintended consequences (students become overly dependent on calculators). This paper uses the following categories to classify instructional software (with the caveat that software programs can combine pedagogical categories):

- Drill and practice;
- Tutorials;
- Tools (including calculators) and problem solving;
- Computer programming;
- Simulations;
- Games;
- Internet;
- Tools for teachers.

A brief description of each of these categories is provided below, and Table 17 provides complementary descriptions. Table 17 lists several features that may distinguish more effective from less effective computer-based practice (including unique features—those that can not easily be duplicated in noncomputer environments).

Drill and practice software provides practice on skills and knowledge to help students remember and use that which they have been taught. A main goal is to achieve automaticity, or fast, accurate, and effortless performance, freeing working memory so that attention can be directed to the more complicated aspects of complex tasks. As with most drill in any medium, drill and practice computer programs present tasks or exercises and give feedback to students.

Tutorials attempt to introduce and teach new subject-matter content, by presenting information and often by attempting to engage students in one-to-one Socratic dialog (e.g., tutorials using artificial intelligence to engage in dialogues). These are usually developed in situations in which a well-defined set of information must be acquired.

The term *computer-assisted instruction* (CAI) is commonly used to refer to drill and practice programs, tutorials, or their combination. A specific type of CAI is the integrated learning system (ILS), a large suite of programs, mainly tutorial, but with drill and practice included, that provides sequenced instruction across several grade levels, tracking students' progress and branching as necessary, and maintaining extensive records of student progress (using *computer-managed instruction*, or CMI, which is discussed in a following section).

Simulations are models of some part of the world (such as the noncomputer board games "Life" or "Monopoly"), and computer simulations are often more complex mathematical models that respond in relatively realistic ways to input based on "real-world" data. Most simulations present situations with components and interactions among those components and generate data about them in response to student input that mirrors relationships in those physical-world or mathematical situations. Thus, students play a role of an active member of a system, making decisions and analyzing the results of those decisions. Goals often are to motivate engagement, develop intuition about a problematic situation, facilitate acquisition of skills and knowledge, and enhance transfer of mathematical skills as students perform activities reflecting those in the "real world."

Games may share characteristics with simulations, as the term "simulation games" suggests. This category is broader, encompassing games that are no more than drill and practice with game-like elements used as rewards, to those in which mathematics is intrinsic to the goals, rules, and tasks of the game. Games of the former type, in which mathematics is extrinsic, often have goals similar to those of drill and practice software. The latter are often designed to promote acquisition of mathematical knowledge and skill, as well as problem solving.

Tools include a wide variety of software programs that perform specific sets of functions, such as calculation, statistics and graphing, computer-based laboratories (CBL, e.g., sensors, including statistical analysis and display of the resulting data), or manipulation of mathematical expressions in symbolic form. Pedagogical goals may include allowing more complex problem solving by transferring routine aspects of tasks to the technological tool and encouraging students to solve problems in practical, applied settings. *Calculators*, including graphing calculators, are widely available tools that have generated a large amount of interest and research. They have been used for many purposes, from facilitating problem solving by allowing students to perform far more, and more complex, arithmetic operations than would have been possible without technology, to serving as simple fact checkers.

Problem solving applications may be one or more tools as above, but may also include the presentation of problems and feedback (similar to the feedback of CAI).

Computer programming involves the provision of computer languages or environments to facilitate students' creation of procedures that solve mathematical problems. Goals may include students' learning and reflecting on algorithms (arithmetic or algebraic), as expressed in the computer language, gaining specific knowledge and skills (e.g., in geometry), and learning certain problem-solving strategies, such as problem determination and explication, problem decomposition, and construction and evaluation of procedures. Some environments are tuned for special purposes, such as the development of mathematical models for simulations, or providing a scripting language within a geometric construction program.

The *Internet* provides general information searching and retrieval functions. Educational applications include specifically organized inquiries (e.g., "WebQuests"). The Internet also offers myriad applications and features ("blogs," groups, etc.) that may be harnessed for the purposes of mathematics education. The Internet can also be the delivery medium for any of the other categories of software; those are considered within their specific category.

Tools for teachers include a variety of software programs designed to aid pedagogical tasks. For example, electronic blackboards ("smart boards") facilitate the display of information or demonstration of any type of software, and with "clickers," can aggregate students' responses; management systems help store, organize and analyze information, such as achievement data, and may include item, test, or practice generators; and hand-held devices facilitate classroom interaction (e.g., each student has a device, and responses or data entry are easily and quickly inputted and evaluated or aggregated).

Computer management systems include *computer-managed instruction*, or CMI, in which the computer analyzes assessments of students, directs their course through a curriculum, and provides reports at individual and aggregate levels. These can be stand-alone systems or can form the foundation for other categories of CBI, such as CAI.

Table 17: Categories of Educational Software

Category	Typical Pedagogy	Possible Features
Drill and Practice	Linear Repetitious Presentation of task, student response, feedback	Sequence[a] Management[b] Feedback[c] Controlled introduction of items Distributed practice Reinforcement schedules
Tutorials	Linear progression with various amounts of branching Didactic presentation and, sometimes, Socratic dialog, presentation of information, questioning, and feedback depending on the response; branching to explanations or review	Sequence Management Feedback Instructional events[d]
Tools and Problem Solving	Specific functions (calculator, graphing, computer-based laboratories, geometric construction, CAS) Problem Solving may include presentation of problems and feedback	Integration/data communication across tools (or with other software categories) Specific feature sets
Computer Programming	Specific language Specific educational environment, specific tasks	Mathematics emphasis Integration into curriculum
Simulations	Nonlinear; exploratory/inquiry-oriented Provides a model of “real-world” or mathematical situation in which students act; then responds to students input following that model	Integrated with tutorials or teaching tools Appropriate simulation[e]
Games	Provides a set of tools and/or miniature “world” as setting for attempting to achieve a goal within a framework of rules Provides clear goals, a set of artificial rules, and elements of competition	Mathematics emphasis Intrinsic mathematics Manipulation of concepts Motivational elements
Internet	Type: General information search/retrieval, “WebQuest,” other	
Tools for Teachers	Type: electronic blackboard, demonstration/display, management system (CMI; may include practice generator), item/test/practice generator, classroom interaction (each student has device)	Integration/data communication across tools (or with other software categories)

[a] Sequence: Consists of building a sequence of mathematical strategies/skills/concepts.
[b] Management: Computer management may consist of record keeping only (includes “picks up where left off”), more sophisticated formative assessment, or formative-assessment-with-branching [e.g., remediation].
[c] Feedback: Corrective feedback may be knowledge of correctness only, or also provide answer, or also provide remediation or explanation. May attend to speed of response.
[d] Features most of the events of instruction identified by cognitive psychology to correspond to learning processes (e.g., gaining attention, informing learner of objectives, stimulating recall of prior learning, presenting stimuli with distinctive features, guiding learning, eliciting performance, providing informative feedback, assign performance, enhancing retention and transfer).
[e] Appropriate abstraction or simplification of the problem situation vs. oversimplification or misrepresentation of the “real-world” situation or the mathematics.

B. Methods

1. Syntheses of Existing Reviews

Prior syntheses and meta-analyses related to the effects of different forms of instructional technology on student mathematics achievement were identified through keyword searches in PsycInfo and Web of Social Sciences Citation Index. Experts in the field of technology instruction and meta-analysis provided additional references. Finally, the reference lists from the identified syntheses and related original studies were reviewed to identify additional syntheses.

From among the group of reviews that were identified, 26 quantitative syntheses and meta-analyses were included in the Task Group's synthesis of existing reviews.[11] These were reviewed to ascertain the number of included studies that focused on the primary population of interest (elementary and junior high school students taking part in mathematics-related technology interventions), the nature of the technology, and the syntheses procedures. Results from these quantitative syntheses and meta-analyses that addressed the primary population of interest of the Panel and the technologies considered were then summarized. The pooled effect sizes from these meta-analyses are presented in Tables C-1 through C-7 of Appendix C.

2. The Task Group's Meta-Analyses

For the Task Group's original meta-analyses, studies were located using the Group's search procedures and the keywords listed in Appendix A. Original empirical studies on technology were categorized based on the category of software on which the intervention focused. Effect sizes were calculated for the Category 1 studies, and effect sizes were pooled when appropriate. All effect sizes have been adjusted for clustering, when appropriate. Study characteristics are provided for each of the Category 1 studies that were included in the meta-analyses.

A number of methodological decisions in preparing the data for analysis and in choosing which effect sizes to include in the pooled analyses were made. In particular, four key issues were confronted, as follows.

First, a number of studies evaluated the effects of more than one technology intervention and/or more than one comparison group. Specifically, three studies (Battista & Clements, 1986; Clements, 1986; and Emihovich & Miller, 1988) evaluated the effects of a programming

[11] The Task Group examined literature reviews, syntheses, and meta-analyses, and conducted syntheses reported here of prior *quantitative* syntheses and meta-analyses, as follows: Becker (1992); Burns & Bozeman (1981); Chambers (2002); Christmann, Badgett, & Lucking (1997); Ellington (2003); Ellington (2006); Fletcher-Flinn & Gravatt (1995); Gordon (1992); Hamilton (1995); Hartley (1978); Hembree (1984); Hembree & Dessart (1986); Hembree (1992); Khalili & Shashaani (1994); Khoju, Jaciw, & Miller (2005); Kuchler (1999); Kulik & Kulik (1991); Kulik (1994); Kulik (2003); Lee (1990); Lou, Abrimi, & d'Apollonia (2001); Niemiec & Walberg (1984); Ryan (1991); Slavin, Lake, & Groff (2007); Slavin & Lake (2007); Smith (1997).

treatment (using Logo) and a tutorial or drill and practice (i.e., CAI) treatment, compared to a no-treatment control. In two of these cases (Clements, 1986; and Emihovich & Miller, 1988), the CAI group was focused, at least in part, on mathematical content. In these cases, the programming versus control group contrast was included in the meta-analysis exploring the effects of programming interventions, and the CAI versus control group contrasts were included in the meta-analyses exploring the effects of drill and practice programs (Emihovich & Miller, 1988) or tutorial programs (Clements, 1986). The programming vs. CAI comparisons are noted in the programming section and presented in Appendix C. In other studies, two similar treatments were compared with a no-treatment comparison group. In these cases (for example, in Oprea, 1988), the treatment that was more similar to a typical intervention that schools would be likely to implement was included. Still, in other studies, a specific intervention was compared to multiple comparison groups. In this situation, the most relevant intervention versus control contrasts was chosen on a case-by-case basis.[12]

Second, studies often explored the effects of interventions on a range of outcomes. For the purposes of this meta-analysis, the focus was only on mathematics-related or problem-solving outcomes. In cases in which multiple outcomes within these domains were available, an average effect size across the multiple outcomes was calculated.

Third, studies often reported effects on a variety of independent samples of students. For example, studies sometimes reported results by race, gender, grade level, or disability status. In cases in which it is likely that the intervention experience was different for these subgroups multiple effect sizes for a study are presented; for example, separate effects by grade level and disability status. In addition, multiple effect sizes are reported for studies that present results from multiple trials exploring the same intervention and outcome (for example, across sites or across samples or cohorts).

Fourth, a number of studies met the criteria for being Category 1 studies but did not compare a technology intervention to a no-treatment control group. Instead, these studies compared two different versions of technology interventions. Although these studies are not appropriate to pool in a meta-analysis, some suggested findings from these comparison studies are presented.

Finally, for studies about calculators only, there were three additional methodological decisions that were made in preparing the data for analysis and in choosing which effect sizes to include in the pooled analyses. First, three studies (Szetela, 1980; Szetela, 1982; and Wheatley, 1980) presented outcomes based on assessments where the calculator treatment group was allowed to use calculators, while the no-calculator comparison group was not. In two of these studies, (Szetela, 1980; Wheatley, 1980) this was the only information available. In the one case in which both an assessment allowing calculator use and a standard paper and

[12] For example, in Johnson-Gentile et al., 1994, one comparison group received an intervention that used manipulatives that was almost identical to the curriculum of the programming intervention. In this case, the programming versus the no-treatment comparison group contrast was included. However, in Ortiz and MacGregor (1991), we chose to include the programming intervention versus a textbook-based intervention contrast, because the no-treatment comparison group did not receive any instruction on the outcome that was being evaluated ("the concept of variable").

pencil assessment that did not allow calculator use were available (Szetela, 1982), the effect size for the latter was included Table C-9 in Appendix C summarizes the effects for any additional sets of effect sizes in which students were allowed to use calculators.

Second, three studies (Duffy & Thompson, 1980; Standifer & Maples, 1981; and Standifer & Maples, 1982) evaluated the effects of two calculator interventions and a no-calculator control group. For the purposes of the meta-analysis, contrasts that are most similar to contrasts in other studies are included, thus attempting to compare the basic treatment of using a calculator during instruction versus not using a calculator.[13] The two Standifer and Maples studies compared the effects of using a standard hand-held calculator and a "programmed feedback" calculator. Focus was on the hand-held versus control group contrast, with the additional effect sizes presented in Table C-9 in Appendix C. The Duffy and Thompson study includes one condition that simply provides students with calculators in the classroom and does not provide guidance to teachers, a second condition that provides calculators plus instructional packages for teachers, and a no-calculator control condition. Again, focus was on the basic calculator versus no calculator contrast, with the effect sizes for the more enhanced treatment documented in Table C-9.

Third, there were three studies (the same three as noted in the previous paragraph) that also provided effect size information for Total Achievement scores. This information is also presented in Table C-9.

C. Categories of Instructional Software: Findings

This section summarizes findings from studies that examined specific categories of instructional software. For each category, results from prior syntheses and meta-analyses provide background information and, for each category for which the Task Group conducted its own meta-analyses, those results are presented.

1. Drill and Practice

a. Prior Syntheses and Meta-Analyses

Many of the studies in the prior CBI reviews probably included drill and practice software, so there is reason to expect similar results for reviews that delineated this category. The detailed effect size information from prior quantitative syntheses and meta-analyses are presented in Table C-1 in Appendix C. Prior syntheses and meta-analyses (see Table 18) suggest that CAI drill and practice generally improves students' performance compared to conventional instruction, with the greatest effects on computation, and less effect on concepts and applications (recall the caveats expressed previously, and note the discussion in the following section on tutorials). Prior reviews have found that drill and practice positively affects attitudes toward mathematics and instruction in mathematics. They suggest that drill and practice is equally effective at all grade levels and may be more effective for males. Drill

[13] In some studies, however, the only treatment or comparison contrasts were a calculator plus additional materials treatment versus a no-calculator control group (for example, Szetela, 1982). In these cases, these contrasts were included.

and practice is the only category of instructional software that shows stronger effects for serving as a substitute for conventional instruction, rather than as a supplement to it. It may be that such programs address students' instructional needs for practice adequately and efficiently, making substantial teacher intervention less important.

Variance in the findings of these reviews, and even wider variance in the individual studies, suggests that general conclusions should be made cautiously. Probably at work here are critical variables, including the quality of the particular software, but also contextual variables (e.g., settings, such as urban, suburban, or rural and student or family characteristics) and implementation variables (e.g., duration, use of the intervention as a supplement or substitution for conventional instruction, and fidelity of implementation; support and availability of resources, funds, and time; setting within the school) (Clements, 2007). One implication is that we should examine the influence of these variables when possible. The scarcity of information in this regard suggests a second implication, to which we will return: The field needs more comprehensive and nuanced reporting and analysis.

Table 18: What Prior Reviews Say About Drill and Practice

- General findings
  - — Generally improves students' performance when compared to conventional classroom instruction (median ES[a] = 0 .345)
  - — Greatest effects on computation, less on concepts and applications
  - — Positive effects on attitudes toward mathematics and instruction in mathematics
- Contextual variables
  - — No consistent differences by grade level
  - — No consistent differences by ability level
  - — Differences favored males
- Implementation variables
  - — Differences favored programs that substitute for other mathematics instruction
  - — Differences favored experimenter or teacher developer-designed (vs. commercially designed) software
  - — No consistent differences by program duration

[a] The median effect size is the median across meta-analyses that reported a pooled effect size. Pooled effect sizes for individual meta-analyses are provided in Appendix C; Drill and Practice is Table C-1.

b. The Task Group's Meta-Analysis of Drill and Practice Software

Table 19 presents the studies in the Task Group's meta-analysis of high-quality experimental and quasi-experimental studies on the effects of drill and practice software on achievement. From all the studies reviewed, only 12 met the criteria for inclusion. These 12 studies yielded a total of 18 effect sizes. Of these, 16 were positive (4 of which were statistically significant) and 2 negative (neither statistically significant), with a mean pooled effect size of 0.320, which was statistically significant. Although this is conjectural, there seems to be a trend for greater effects of interventions that were shorter and focused on developing the automaticity of specific skills. If this is indeed the case, this would be consistent with reports from other reviews.

The Task Group extended each meta-analysis to ascertain whether effect sizes were mediated by particular contextual and implementation variables. Results are presented in Table 20. A between-group p-value was calculated using CMA software to determine if the effect of a particular contextual or implementation variable was significant, and results of these analyses are shown in the same row as the name of the variable.

Table 19: Studies That Examine Effects of Drill and Practice Technology on Mathematics Achievement

Study	Design	Sample	Duration/Content	Contrast	Measure	Hedge's g	Standard Error	
Drill and Practice								
Ball, 1988[a]	QED	91 fourth-grade students in five classes in two schools	12 computer lessons over 6–8 weeks/ Fractions	Computer lessons on fractions vs. traditional fraction instruction	Posttest—Fractions	0.815	~	0.457
Campbell et al., 1987	RCT	48 third-grade students in a middle-class suburban school in the Southeast	20 minutes daily of D&P for 5 weeks (+30 min instruction for both T &C)/ Division of whole numbers/Milliken Mathematics Sequences program	Milliken Mathematics D&P vs. worksheets	Posttest—Division	0.445	(ns)	0.288
Carrier et al., 1985	RCT	144 fourth-grade students in six classrooms in a metropolitan school district	10–15 minutes per lesson over 14 weeks/ Multiplica-tion and division	Three different D&P vs. Worksheets	Post: Symbolic algorithms, mult. & division	0.228	(ns)	0.167
Emihovich & Miller, 1988	RCT	24 first-grade students in five classrooms in an elementary school in the Southeast	20, 30-min sessions (3 months)/ Addition/ subtraction, basic mathematics skills	Series of CAI software vs. regular reading and mathematics instruction	CTBS—Mathematics	0.407	(ns)	0.399
Fletcher et al., 1990[a]	RCT	41 third-grade students in rural Saskatchewan, Canada	Spring semester/ 3rd-grade mathematics	Milliken Mathematics Sequence vs. Control (traditional instruction + worksheets)	Canadian Tests of Basic Skills (CTBS)	0.412	(ns)	0.693
		38 fifth-grade students in rural Saskatchewan, Canada	Spring semester/ 5th-grade mathematics			0.338	(ns)	0.697
Fuchs et al., 2006	RCT	33 first-grade learning disabled students in nine classrooms in three Title 1 schools in a metropolitan school system	50, ten-minute sessions over 18 weeks/ Addition and subtraction	mathematics FLASH vs. spelling FLASH	Post: addition, subtraction, and story problems	0.177	(ns)	0.349
Kraus, 1981	RCT	19 second-grade students in one school in a southwestern Ohio city	5 sessions over a two week period (average 64 minutes)/ Fish Chase game: addition	Fish Chase vs. Hangman	Addition speed test	1.454	**	0.523

Continued on p. 6-118

Table 19, continued

Study	Design	Sample	Duration/Content	Contrast	Measure	Hedge's g	Standard Error	
Drill and Practice								
McCollister et al., 1986	RCT	15 Kindergarten students in achievement level 1 in one public school in a large Southern city	6 sessions/ Numeral recognition and cardinal counting	How Many Squares computer program vs. Milton Bradley Flannel board	Pine & Burts (1984) numeral recognition and cardinal counting	-0.548	(ns)	0.529
		13 Kindergarten students in achievement level 2 in one public school in a large Southern city				0.344	(ns)	0.575
		25 Kindergarten students in achievement level 3 in one public school in a large Southern city				0.429	(ns)	0.405
Podell et al., 1992: Study 1	RCT	24 second-grade students in New York City public schools	Up to 10 15-minute sessions, three times per week/ Addition	Mathematics Blaster - Addition vs. Worksheets	Accuracy rate: mean trials to criterion	0.150	(ns)	0.408
		28 learning disabled students in Grades 2–4 in New York City public schools				0.783	*	0.397
Podell et al., 1992: Study 2	RCT	20 students in New York City public schools, ages 6–9	Up to 10 15-minute sessions, three times per week/ Subtraction	Mathematics Blaster - Subtraction vs. Worksheets	Accuracy rate: mean trials to criterion	0.627	(ns)	0.478
		22 learning disabled students in New York City public schools, ages 6–11				0.568	(ns)	0.435
Saracho, 1982[a]	QED	256 Spanish speaking migrant children attending third through sixth grade	3 hours a week, 60 hours for the academic year/ Elementary mathematics	D&P vs. regular classroom instruction	CTBS—Grades 3–6	-0.118	(ns)	0.304
Saunders & Bell, 1980	RCT	101 advanced Algebra students in four classes in one public high school	<1/2 hr per week for the school year/ Algebra II	Algebra problems using BASIC vs. regular Instruction	Cooperative Mathematics Test: Algebra II	0.136	(ns)	0.201
Watkins, 1986	RCT	82 first-grade students from a suburban Southwestern school	3, 15 min sessions per week (October through June)/ Mathematics Machine D&P	Mathematics D&P vs. Reading D&P	California Achievement Test	0.432	~	0.221

Heterogeneity							
Q-value	*Df (Q)*	*P-value*	*I-squared*	*Pooled ES (12 studies, 18 effect sizes)*	*Hedge's g*	*Standard Error*	
14.678	*17*	*0.619*	*0.000*		*0.320*	***	*0.078*

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

[a] Data were adjusted for clustering that occurred within classrooms.

Contextual variables. The studies in the Task Group's meta-analysis yielded the following results regarding contextual variables:

- *Age or grade.* The effect of drill and practice software was confirmed as significantly effective at the elementary level, but there are not enough studies at the other levels to make any comparisons or other conclusions.
- *Ability.* There is no evidence that children with and without learning disabilities benefit differently from use of drill and practice software.

Implementation variables. These studies yielded the following results regarding implementation variables:

- *Duration.* Results of the Task Group meta-analysis indicate that duration of the intervention is not a significant moderator. This is consistent with findings from prior syntheses and meta-analyses.

- *Substitute versus supplement.* Consistent with other review findings, the effect sizes calculated in the Task Group meta-analysis were higher for drill and practice interventions that substituted for, rather than supplemented, classroom practice. Effect sizes were significant only for substitution implementations, but the difference between the two did not reach statistical significance.

- *Experimenter or teacher vs. commercial developer.* Effects sizes were larger for experimenter or teacher-developed, compared to commercial, drill and practice software, with no significant difference between them. This is consistent with prior review findings, both yielded significant effects.

In summary, the Task Group's meta-analysis of rigorous studies about the effects of drill and practice software produced a mean pooled effect size of 0.320 that is statistically significant. This finding is consistent with the conclusions of prior syntheses and meta-analyses. There is no solid evidence that students of different ability levels or disability status benefit differently. Results suggest higher effect sizes when drill and practice software is used as a substitute, rather than supplement, to instruction (although comparisons were not significant).

Table 20: Subgroup Analysis

	Drill and practice				Tutorials				Programming			
	N studies/ ES	Hedges g		Se	N studies/ ES	Hedges g		se	N studies/ ES	Hedges g		Se
Contextual variables												
Grade level		ns [a]				ns				*** [a]		
Elementary	11 / 17	0.352	***	0.084	2 / 4	0.235		0.214	**9 / 22**	**0.854**	***	**0.134**
Middle School	0 / 0	na		na	3 / 4	0.138		0.088	**7 / 8**	**0.218**		**0.151**
High School	1 / 1	0.136		0.201	4 / 5	0.480	~	0.246	**0 / 0**	**na**		**na**
Mixed	0 / 0	na		na	1 / 1	0.379		0.441	**0 / 0**	**na**		**na**
Ability		ns				ns				na		
Learning disabled	3 / 4	0.303		0.258	4 / 5	0.238		0.143	0 / 0	na		na
Non-LD	10 / 13	0.356	***	0.087	6 / 9	0.356	**	0.136	14 / 30	0.674	***	0.115
Migrant (span speaking)	1 / 1	-0.118		0.698	0 / 0	na		na	0 / 0	na		na
Implementation variables												
Duration		ns				** [a]				ns		
Less than 4 weeks	3 / 8	0.492	**	0.184	**3 / 5**	**0.642**	***	**0.181**	3 / 3	0.974	~	0.530
4 to 8 weeks	2 / 2	0.550	*	0.243	**0 / 0**	**na**		**na**	2 / 2	0.910	*	0.441
Greater than 8 weeks	7 / 8	0.223	*	0.095	**6 / 9**	**0.141**	~	**0.075**	9 / 25	0.625	***	0.124
Supplementation vs. substitution[a]		ns				ns				ns		
Supplement	4 / 4	0.290		0.250	2 / 3	0.425	*	0.175	7 / 9	0.721	**	0.208
Substitute	8 / 14	0.370	***	0.092	7 / 11	0.288	*	0.112	7 / 21	0.655	***	0.141
Curricular Integration[b]		ns				*** [a]				ns		
Low	2 / 2	0.766		0.637	**0 / 0**	**na**		**na**	1 / 1	-0.065		0.444
Medium	7 / 11	0.319	**	0.100	**3 / 6**	**0.037**		**0.074**	9 / 12	0.739	***	0.159
High	3 / 5	0.259	~	0.139	**6 / 8**	**0.503**	***	**0.108**	4 / 17	0.682	***	0.168
Commercial vs. researcher		ns				** [a]				na		
Commercial	7 / 13	0.268	*	0.104	**5 / 8**	**0.092**		**0.068**	14 / 30	0.674	***	0.115
Researcher-designed	5 / 5	0.441	**	0.172	**4 / 6**	**0.516**	**	**0.165**	0 / 0	na		na
Total	12 / 18	0.320	***	0.078	9 / 14	0.302	**	0.099	14 / 30	0.674	***	0.115

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

[a] Between group test of significance p-value (ns = not significant; na = not applicable).

[b] Supplementation categories are defined as follows: "supplement" = the technology treatment served as an addition to regular class time in mathematics; "substitute" = time spent on treatment technology substituted for at least some portion of math instruction/class time.

[c] Curricular integration characterizes the level of integration with the regular math curriculum. "Low" is categorized as little to no integration with math curricula; "Moderate" is defined as covering topics related to the regular math curricula and possibly coordinating instruction with technology; "High" is defined as curricula that was designed around the specific technology intervention.

2. Tutorials

a. Prior Syntheses and Meta-Analyses

Prior syntheses and meta-analyses (see summary Table 21) suggest that CAI tutorials improve students' performance compared to conventional instruction, with slightly greater effects on concepts and applications than on computation. The detailed effect size information from prior quantitative syntheses and meta-analyses are presented in Table C-2 in Appendix C. These syntheses most frequently identify tutorials as the most effective software category, when compared to drill and practice, simulations and games, and tools. They suggest that tutorials appear to be effective at all grade levels, particularly the higher grades and that tutorials are more effective when they supplement, rather than replace, conventional instruction, when they involve experimenter or teacher-developed, rather than commercially-developed, software, and when they are developed for a specific audience rather than a general audience. These findings come from syntheses and meta-analyses with different inclusion criteria than those used by the Instructional Practices Task Group.

Table 21: What Prior Reviews Say About Tutorials

- General findings
  - Generally improves students' performance when compared to conventional classroom instruction, with a median pooled effect size of 0.38 (Table C-2)
  - More researchers have claimed that tutorials are more effective than drill and practice (Burns & Bozeman, 1981; Khalili & Shashaani, 1994; Lee, 1990)
  - Somewhat higher effect sizes for concepts and applications than computation
  - No effects on attitudes
  - Often low fidelity of program implementation
- Contextual variables
  - Slight advantage for higher grade levels
  - No consistent differences by ability level
  - No consistent differences by gender
- Implementation variables
  - Differences favoring programs that supplement instruction versus substitute
  - Differences favoring experimenter or teacher developed programs vs. commercially developed software
  - Differences favoring specific vs. a general audience
  - No consistent differences by program duration

b. The Task Group's Meta-Analysis of Tutorial Software

Table 22 presents the studies in the Task Group's meta-analysis of high-quality experimental and quasi-experimental studies on the effects of tutorial and mixed tutorial and drill and practice software on achievement. From all the studies reviewed, only nine met the criteria for inclusion. These studies yielded a total of 14 effect sizes. Of these, 10 were positive (two of which were statistically significant), one negative, and three near zero, with a significant mean pooled effect size of 0.302. Those studies assessing mathematics achievement only had a mean pooled effect size of 0.288, which was statistically significant. Those that assessed problem-solving ability had a mean pooled effect size of 0.425, which also was statistically significant. Several contextual and implementation variables were examined.

Contextual variables. These studies yielded the following results regarding contexts (see Table 20).

- *Age or grade.* Similar to the results of the prior syntheses and meta-analyses, the IP meta-analysis indicates that tutorials have a slight advantage for high school students, but there are no significant differences between those effects and effects for other grade levels.

- *Ability.* Similar to the results of the prior syntheses and meta-analyses and the results for drill and practice, there was no significant difference between effects for students with and without learning disabilities, although the tendency for effects to be lower in schools with lower achievement needs further study.

Implementation variables. The Task Group's meta-analysis indicates the following regarding implementation variables.

- *Duration.* Tutorials were significantly more effective in studies in which interventions were less than 4 weeks in duration than those in which interventions were greater than 8 weeks. This must be interpreted with caution: Some treatments took place over many weeks, but the time students used the software remained limited. Thus, this finding may have more to do with limiting the confounding effects of other factors.

- *Substitute vs. supplement.* Tutorials were more effective when they supplement, rather than replace, conventional instruction, but the difference was not significant. However, they are significantly more effective when they are highly integrated with the regular mathematics curriculum (compared to medium integration, which had near-zero effects).

- *Experimenter or teacher vs. commercial developer.* Consistent with the prior syntheses and meta-analyses, there are stronger effects for experimenter or teacher-developed, compared to commercial, software.

The Task Group's meta-analysis of rigorous studies is consistent with the conclusions from the prior syntheses and meta-analyses. The Task Group analysis suggests that there is a suggestion that high school students may benefit more from tutorials than students at other grade levels (although comparisons were not significant). Tutorials were significantly more effective if they were highly integrated with the regular mathematics curriculum than when they were less integrated. Tutorials developed by a researcher or teacher had significantly greater positive effects than commercial software.

One of the studies in the Task Group set examining tutorials, Dynarski et al. (2007), includes two recent large randomized trial evaluations, and warrants particular attention because of the scale of these two studies (3,136 students in one study, 1,402 in a second study). The results suggest caution. The near-zero effect sizes in Dynarski et al. (0.071, -0.064) suggest that results of using tutorials are not guaranteed to be superior to standard instruction. Moreover, the results suggest additional questions that must be addressed in future research. Scaling up software interventions may be particularly difficult, and the more encouraging results from earlier and smaller studies (e.g., Fuchs et al., 2002, nonsignificant effect size of 0.586; Henderson et al., 1985, significant ES of 0.976; Thompson & Rickhuss, 1992, nonsignificant effect size of 0.774; or Wheeler & Regian, 1999, significant effect size of 0.517) may reflect efficacy under advantageous (i.e., closer to "ideal") conditions more than effectiveness at scale.

The Task Group findings indicate that tutorials are more effective if they are highly integrated into the curriculum (see Table 20), which requires that such integration be done either by the curriculum or software developers or by teachers. This is an extensive task that may demand additional time for both professional development and for work with colleagues on curriculum. A final issue is amount of use of tutorial software in classrooms. In the Dynarski et al. (2007) study, it considers teacher reports of tutorial software usage in the classroom. But when the study considered software recorded usage, usage in the classroom was much lower; compare teachers' report of 51 hours of usage to the products' reports of 17 hours for sixth grade; or teachers' report of 46 hours of usage to the products' reports of 15 hours for ninth grade. Even the teacher data are substantially lower than publishers' recommendations. This is consistent with the Panel's National Survey of Algebra Teachers that indicated low frequency of the use of technology (averaging "less than once a week;" (Hoffer, Venkataraman, Hedberg, & Shagle, 2007). These are issues in scaling up software use and suggest important questions for future research.

The direct implications of the Dynarski study are serious cautions to anyone who believes merely introducing technology will raise students' scores. This was a rigorous randomized control trials design conducted in 33 districts and 1,232 schools. The products being evaluated had been identified as being effective and widely used. Teachers were trained. There were no significant effects. Thus, educators must consider not only empirical evidence of effectiveness of a particular software package but also issues of scale-up, including integration with the extant curriculum, fidelity of implementation, including amount of use, and technological and pedagogical support.

To return to the software per se, studies also show that fine-tuning the mathematics and pedagogy in software can make a significant difference in learning. For example, in a study of another cognitive tutor (geometry), holding time-of-instruction constant, one group discussed why and how they used the strategy they used, and the other practiced more problems. The authors report that the former group had significantly greater understanding and showed greater transfer (Aleven & Koedinger, 2002).

Table 22: Studies That Examine Effects of Tutorials or Tutorials Plus Drill and Practice on Mathematics Achievement

Study	Design	Sample	Duration/Content	Contrast	Measure	Hedge's g	Standard Error	
Tutorial + Drill & Practice								
Clements, 1986	RCT	24 first-grade students from a middle-class midwestern school system	44 sessions (22 weeks)/ Elementary mathematics using CAI drill, tutorial, and problem-solving software (mathematics and reading)	CAI drill, tutorial, and problem-solving software (mathematics and reading) vs. traditional instruction	WRAT Mathematics score	0.397	(ns)	0.398
		24 third-grade students from a middle-class midwestern school system				0.142	(ns)	0.395
Dalton & Hannafin, 1988	RCT	22 low-achieving eighth-grade students from five sections in same school	Two lessons/ Geometry & area of circle	Computer initial and remedial instruction vs. traditional initial and remedial instruction	Mastery quiz: area of circle	-0.001	(ns)	0.426
		25 high-achieving eighth-grade students from five sections in same school				0.571	(ns)	0.395
Dynarski et al., 2007: Study 1[a]	RCT	3,136 sixth-grade students in 10 different districts across U.S., focused on lower achievement districts	One academic year, wide variation, but overall average use of the CAI was 17hrs/yr/ General mathematics	CAI vs. Control (standard instruction)	SAT-10 mathematics battery	0.071	(ns)	0.106
Dynarski et al., 2007: Study 2[a]	RCT	1,404 algebra students in 10 different districts across U.S., focused on lower achievement districts	One academic year, average use of the CAI was 15hrs/yr/ Algebra	CAI vs. Control (standard instruction)	ETS End-of-Course Algebra Assessment	-0.064	(ns)	0.117
Fuchs et al., 2002	RCT	18 fourth-grade students with mathematics disabilities in three schools in a southeastern city	24 sessions (twice per week for 12 weeks)/ Problem-solving	Computer vs. Control	Pooled problem-solving score (three subtests)	0.586	(ns)	0.486
		20 fourth-grade students with mathematics disabilities in three schools in a southeastern city	48 sessions (four times per week for 12 weeks)/ Problem-solving	Tutor + Computer vs. Tutor only (ES of Tutor + Computer vs. Control was 1.281)		-0.147	(ns)	0.448
Henderson et al., 1985	RCT	81 students attending five general mathematics or intro to algebra classes in one high school with large proportion of Latino students	Three modules (Three sessions)/ Factors and prime numbers	Computer-Video vs. Control	Combined Recognition and Constructed scores on Factors and Prime numbers test	0.976	***	0.234
Moore, 1988	RCT	117 seventh- and eighth-grade students in the lowest level of remedial mathematics in four middle schools	School year (Sept - May)/ Middle school mathematics instruction	Milliken Mathematics Sequences + written assignments vs. Direct Instruction using Mathematics for Individual Achievement	District Mathematics Placement Test	0.273	(ns)	0.185

Continued on p. 6-125

Table 22, continued

Study	Design	Sample	Duration/Content	Contrast	Measure	Hedge's g	Standard Error	
Tutorial + Drill & Practice								
Thomas & Rickhuss, 1992	RCT	17 high school students (average age 15) in one algebra class	1 week/ Algebra—solving equations	CAI MuMath/Solving Equations vs. noncomputer instruction	Solving equations	0.364	(ns)	0.465
			1 week/ Algebra—factorization	CAI MuMath/ Factorization vs. noncomputer instruction	Factorization	0.774	(ns)	0.480
Triffiletti et al., 1984	RCT	20 learning disabled students (ages 9–15) in a private school in Jacksonville, FL	school year (Sept - May)/ SPARK-80 Computerized Mathematics System	SPARK-80 vs. Resource Room	Key Mathematics Diagnostic Arithmetic Test (grade equiv)	0.379	(ns)	0.441
Wheeler & Regian, 1999[a]	RCT	493 ninth-grade students in 40 traditional mathematics instruction classes in Texas, New Mexico, and Ohio	one session per week, for the school year/ Word Problem Solving (WPS) Tutor	WPS vs. Control	Word Problem solving combo (concrete & abstract)	0.517	*	0.206
Heterogeneity								
Q-value	*df (Q)*	*P-value*	*I-squared*	*Pooled ES (9 studies, 14 effect sizes)*		*Hedge's g*	*Standard Error*	
24.385	*13*	*0.028*	*46.688*			*0.302*	**	*0.099*
Mathematics outcomes only								
Heterogeneity								
Q-value	*df (Q)*	*P-value*	*I-squared*	*Pooled ES (7 studies, 11 effect sizes)*		*Hedge's g*	*Standard Error*	
20.519	*10*	*0.025*	*51.264*			*0.288*	*	*0.112*
Problem-solving outcomes only								
Heterogeneity								
Q-value	*df (Q)*	*P-value*	*I-squared*	*Pooled ES (two studies, three effect sizes)*		*Hedge's g*	*Standard Error*	
1.939	*2*	*0.379*	*0.000*			*0.425*	*	*0.175*

Note: The 2 studies with problem-solving outcomes are Fuchs et al. (2002) and Wheeler & Regian (1999), all others have mathematics outcomes.
~ p < .10, * p < .05, ** p < .01, *** p < .001
[a] Data were adjusted for clustering that occurred within classrooms.

3. Tools: Calculators and Graphing Calculators

Among the many categories of technology, calculators, including graphing calculators, have probably generated the greatest amount of debate. Some have championed their use in developing problem-solving ability by allowing students to perform far more, and more complex, arithmetic operations than would have been possible without technology. Others have bemoaned their misuse as simple fact checkers. A concern is that calculators have an insidious effect on paper-and-pencil arithmetic and algebraic skills.

Calculators have been used in mathematics education for 70 years, since Emmett Betts engaged students with calculating machines in 1937. They metamorphosed from the original bulky and expensive machines to the electronic calculators of the 1960s, the inexpensive handheld, four-function calculators of the 1970s, and the wide variety of basic, scientific, and graphing calculators available today.

The usefulness of calculators in homes and businesses may seem clear, but their use in education, at first blush, seems equally problematic—students should *learn* to compute without calculators. The Panel's survey of the nation's algebra teachers indicated that the use of calculators in prior grades was one of their concerns (Hoffer et al., 2007).

a. Prior Syntheses and Meta-Analyses

Previous reviews (see summary in Table 23) have suggested that calculators of all types, basic, scientific and graphing, may benefit students' achievement in and attitudes toward mathematics (see the detailed effect size information from prior quantitative syntheses and meta-analyses in Tables C-3 and C-4 in Appendix C). Effects are usually more positive when students are allowed to use calculators during testing. Effects on concepts, contrary to perhaps the most common concern, are near zero but positive, and effects on problem solving were positive.

Table 23: What Prior Reviews Say About Calculators

- General findings
  - — Generally improve students achievement and attitudes (median pooled effect size on computation, 0.41, Table C-3)
  - — Generally improve mathematical problem solving (median pooled effect size, 0.19) but little or no effect on conceptual development (median pooled effect size near zero)
  - — Most effective in facilitating learning of operational skills ("operational" indicating that the report was unclear as to whether the instrument assessed the computational, conceptual, or both domains)
  - — Graphing calculators particularly effective for conceptual skills (Table C-4)
  - — All the effects mentioned are lower when testing without calculators
- Contextual variables
  - — No consistent differences by grade levels
  - — Some differences by ability level
- Implementation variables
  - — Special calculator instruction may have more positive effects (pedagogical uses vs. merely providing calculators) on computational, operational, and problem solving competencies

b. The Task Group's Meta-Analysis of Calculators

Turning to the Task Group's meta-analysis of rigorous studies, Tables 24, 25, and 26 provide individual study and pooled effect sizes for each of the three focal outcomes: computation (Table 24), problem solving (Table 25), and concepts (Table 26).[14] The tables disaggregate studies analyzed at the student level from studies analyzed at the classroom level.

[14] Many of the studies had a concern that calculators may impede computational achievement, and thus were testing to see whether use of calculators had a positive or negative effect on computation. The studies that assessed the effects on problem solving and concepts often hypothesized that calculators would improve these outcomes.

Meta-analytic pooled effect sizes and accompanying statistics are based on pooling of similarly aggregated effect sizes. In other words, studies that analyzed data at the student-level are pooled together, and similarly, studies that analyzed at the classroom level are pooled together.

It was the Task Group's hope to discern through this meta-analysis any differences that might exist between the effects of graphing calculators and non-graphing calculators. However, nearly all of the peer-reviewed published studies using graphing calculators examine the effects on students in advanced mathematics courses (such as Algebra 2, Trigonometry, Precalculus, and Calculus). As a result, only one of the included studies (Graham & Thomas, 2000) used a graphing calculator.

Table 24 presents studies that contrast treatment condition using calculators with a non-calculator control condition on computational outcomes. Seven of the studies (one with effects at four different grade levels) analyzed data at the individual student level, and the remaining three (including nine comparisons) used the classroom or teacher as the unit of analysis. For the student-level set, then, ten effect sizes were calculated. The pooled effect size is 0.319, which borders on statistical significance. In only one of the included studies (Wheatley, 1980), students were allowed to use calculators during assessment. Once that study is removed from the analysis, the pooled effect size is 0.307, and is not statistically significant. For the classroom-level set, nine effect sizes were calculated. The mean effect size is -0.085, which is not statistically significant.

Outcomes of studies that examine the effects of calculator use on problem solving are presented in Table 25. The seven comparisons in the student-level set (note that four were from a single study, Szetela (1982)) yielded a mean pooled effect size of 0.304, which borders on statistical significance. The four comparisons from the classroom-level set yielded a mean pooled effect size of -0.063, which was not statistically significant.

Regarding outcomes on measures of conceptual development, presented in Table 26, the four comparisons in the student-level set yielded a mean pooled effect size of 0.278, which is not statistically significant. The three comparisons from the classroom-level set yielded a mean pooled effect size of 0.128, which was not statistically significant.

Several contextual and implementation variables were examined. These are summarized in Table C-8 in Appendix C.

Contextual variables. These studies yielded the following results regarding contexts.

- *Age or grade.* There were no statistically significant differences among the effect sizes for elementary school-aged students (ES = 0.367, ns, five effect sizes within four studies) versus secondary school-aged students (ES = 0.113, ns, four effect sizes within three studies) for computation, nor for applications or concepts. However, this is based on a small sample of studies and thus there may not be sufficient power to detect differences in effect sizes (e.g., differences in effect sizes for applications, which were larger for secondary than for elementary, and statistically significant only for secondary, should be evaluated in future research).

Implementation variables. The Task Group's meta-analysis indicates the following regarding implementation variables.

- *Duration.* Studies that provided interventions for shorter periods (less than 3 months) had stronger effects on computation than studies that extended over longer time periods (3 months or longer). Specifically, the pooled effect size, under random effects assumptions, for interventions taking place for less than three months was a statistically significant 0.503 ($p < .05$, seven effect sizes within five studies); while the effect for studies taking place for longer than 3 months was not statistically significant (ES = -0.134, three effect sizes within three studies).[15] Although this would be an interesting finding if valid, it is based on a small sample, and thus it is likely that other factors unrelated to program duration may have led to this result. There were no such significant differences for applications or concepts.

- *Special calculator instruction.* Using alternative interventions or enhancing the intervention did not, as a whole, yield significantly higher effect sizes.

In summary, effect sizes of the Task Group's meta-analysis to examine the effects of calculator use on computation skills are smaller than those reported in prior syntheses and meta-analyses.

Concerning the impact of calculator use on problem-solving competencies, the Task Group's meta-analysis at the student level yielded a borderline significant, positive effect, but classroom-level analyses were near zero. The results in Table 25 are mainly for the outcomes in which students were not allowed to use calculators to solve problems on the assessments, Wheatley (1980) is the only study that includes outcomes where calculators were allowed. When looking at outcomes in which calculators were permitted on the assessments, effects were more positive (e.g., two of the four contrasts examined from Szetela (1982) reached statistical significance, see Table C-9 in Appendix C). Assessing proficiency with the same tools available as were available during instruction may be viewed as constituting a valid comparison, perhaps especially for problem-solving outcomes. Comparing these conclusions to those in the syntheses of previous reviews, the pattern is similar to what was found for computation: The effect sizes in the present meta-analyses are smaller.

Effect sizes on conceptual development tended to be positive, favoring the calculator treatments, but generally small and all nonsignificant (see Table 26). This is consistent with the prior syntheses and meta-analyses, which reported near-zero pooled effect sizes.

[15] The five studies taking place for less than three months include: Schnur & Lang (1976), Standifer & Maples (1981), Szetela (1980), the Grade 3, 5, and 7 sample of Szetela (1982), and Wheatley (1980). The three studies taking place for three or more months include: Campbell & Virgin (1976), Standifer & Maples (1982), and the Grade 8 sample of Szetela (1982).

The two specific alternative interventions or enhancements to calculators (programmed feedback calculators and supplementary materials) used in studies identified as high quality by the Task Group (Standifer & Maples, 1981, 1982; Duffy & Thompson, 1980) did not yield any significant effect sizes. This is in contrast to the findings of the prior syntheses and meta-analyses (and to the findings for formative assessment discussed in the Task Group report), which include a wider range of enhancements, including more recent interventions. More research needs to be conducted, for example, on essential distinctions such as between functional and pedagogical use.

Finally, there are several important caveats. Effects of calculator use, especially appropriate versus inappropriate pedagogical use in the early grades, have not been adequately researched. Similarly, long-term effects of inappropriate calculator use may be negative (Wilson & Naiman, 2004); there is no reliable evidence. The Task Group's meta-analysis could not include adequate research on graphing calculators; high-quality research is needed regarding this type of calculator.

Table 24: Studies That Investigate the Effects of Calculators on Computation Outcomes

Study	Design	Sample	Duration/ Content	Contrast	Measure	Subgroup	Hedge's g	Standard Error	
Computation Outcomes									
Student-level analyses									
Campbell & Virgin, 1976[b]	Quasi	252 fifth- and sixth-graders in two North York elementary schools (Canada).	7 months/ Basic computation	Calculators to check work vs. No calculators	Metropolitan Achievement Test computation score	Overall	0.022	(ns)	0.642
Schnur & Lang, 1976[a]	RCT	60 youths ages 9 to 14 in four summer compensatory education classes in rural Iowa.	26 weeks/ Computation	Calculators to check work and compute subset of problems vs. Compensatory education program	Computational Skills Program Computational Test	Overall	0.855	~	0.512
Standifer & Maples, 1981[a]	RCT	141 students in 6 third-grade classrooms in Monroe, Louisiana	11 weeks/ Computation	Hand-held, four function calculator vs. No calculator in regular mathematics curriculum (see Table C-9 for effects of programmed feedback calculator vs. No calculator in regular mathematics curriculum)	Science Research Associates Assessment: computation score	Overall	0.635	(ns)	0.398
Standifer & Maples, 1982[a]	RCT	113 students in 10 third- and fourth-grade Title I compensatory mathematics classrooms in Monroe, Louisiana	5 months/ Computation	Hand-held, four function calculator vs. General remedial mathematics curriculum (see Table C-9 for effects of experimental group 2 using programmed-feedback calculators + regular remedial curriculum)	Science Research Associates Assessment: computation score	Overall	0.023	(ns)	0.329
Szetela, 1980	RCT	39 students in two seventh-grade classes in a middle class elementary school (likely in Canada)	3 weeks/ Focus on learning the concept of ratios	Calculator-based instruction with four-function calculator vs. Instruction without calculators	Researcher-designed test on ratios	Overall	0.322	(ns)	0.316

Continued on p. 6-130

Table 24, continued

Study	Design	Sample	Duration/ Content	Contrast	Measure	Subgroup	Hedge's g	Standard Error	
Szetela, 1982	RCT for Grades 3, 5 and 7 and Quasi for Grade 8	46 third-grade students in a middle income school in Richmond, British Columbia	8 weeks/ All grades focused on problem solving. Grade specific foci included: Grade 3: whole number operations in multiplication, basic division; Grade 5: introduction to decimals, operations with decimals; Grades 7 and 8: decimals, ratios, and percents.	Regular instruction plus calculator-specific materials vs. Regular instructional activities	Researcher-designed computational skills (16 items); tailored to grade level	Third grade	1.337	***	0.323
		33 fourth-grade students in a middle income school in Richmond, British Columbia				Fifth grade	-0.307	(ns)	0.342
		47 seventh-grade students in a middle income school in Richmond, British Columbia				Seventh grade	0.279	(ns)	0.288
		54 eighth-grade students in a middle income school in Richmond, British Columbia				Eighth grade	-0.267	(ns)	0.270
Wheatley, 1980[c]	Quasi	44 sixth-grade students in two classes (same teacher) in an elementary school in a Midwestern university town	6 weeks/ Problem solving	Problem solving with calculators vs. Problem solving intervention without calculators	Measure of computational errors (reverse coded) on five researcher-designed problems	Overall	0.573	(ns)	0.691
Heterogeneity									
Q-value	*df (Q)*	*P-value*	*I-squared*	*Pooled ES: student level (7 studies, 10 effect sizes)*			*Hedge's g*	*Standard Error*	
20.822	*9*	*.013*	*56.776*				*0.319*	*~*	*0.077*
Classroom-level analyses									
Duffy & Thompson, 1980	RCT	Approx. 135 students in 20 fourth-grade classrooms in Columbus, Ohio	26 weeks/ Application problems, decimals, rounding, estimation	Calculators only plus regular mathematics program vs. Regular mathematics curriculum (see Table C-9 for effects of calculator plus instructional packages for teachers, plus regular mathematics program)	CTBS computation score	Fourth grade	0.037	(ns)	0.428
		Approx. 122 students in 18 fifth-grade classrooms in Columbus, Ohio				Fifth grade	0.325	(ns)	0.452
		Approx. 129 students in 19 sixth-grade classrooms in Columbus, Ohio				Sixth grade	-0.395	(ns)	0.444
Szetela & Super, 1987	Quasi	Approx. 424 students in 21 seventh-grade classrooms in an urban-rural district in Canada	One school year/ Problem solving	Problem solving with calculators vs. Problem solving intervention without calculators	Rational Numbers test—40 item test used in British Columbia	Overall	-0.076	(ns)	0.423

Continued on p. 6-131

Table 24, continued

Study	Design	Sample	Duration/ Content	Contrast	Measure	Subgroup	Hedge's g	Standard Error	
Wheatley & Shumway, 1979	RCT	Students in 50 classrooms in second through sixth grade in five Midwestern states. Ten classrooms in each grade level.	7 months/ Basic four-function calculator/ computation	General calculator use (teachers trained but determine how they will implement) vs. No calculators/regular mathematics program	Stanford Achievement Test - Computation score	Second grade	-0.603	(ns)	0.587
						Third grade	0.352	(ns)	0.577
						Fourth grade	-0.460	(ns)	0.580
						Fifth grade	-0.434	(ns)	0.579
						Sixth grade	0.315	(ns)	0.576
Heterogeneity									
Q-value	*df (Q)*	*P-value*	*I-squared*	*Pooled ES: student level (three studies, nine effect sizes)*			*Hedge's g*	*Standard Error*	
4.010	*8*	*0.856*	*0.000*				*-0.085*	*(ns)*	*0.167*

~ p < .10, * p < .05, ** p < .01, *** p < .001
[a] Data were adjusted for clustering that occurred within classrooms.
[b] Data were adjusted for clustering that occurred within schools.
[c] The treatment group was allowed to use a calculator during assessment for this outcome.

Table 25: Studies That Investigate the Effects of Calculators on Problem Solving Outcomes

Study	Design	Sample	Duration/ Content	Contrast	Measure	Subgroup	Hedge's g	Standard Error	
Problem Solving Outcomes									
Student-level analyses									
Campbell & Virgin, 1976[b]	Quasi	150 fifth- and sixth-graders in two North York elementary schools (Canada)	7 months/ Basic computation	Calculators to check work vs. No calculators	MAT computation score	Overall	*0.238*	*(ns)*	*0.642*
Szetela, 1980	RCT	39 students in two seventh grade classes in a middle class elementary school (likely in Canada)	3 weeks/ Focus on learning the concept of ratios	Calculator-based instruction with four-function calculator vs. Instruction without calculators	Researcher-designed ratio problems test	Overall	0.869	**	0.329
Szetela, 1982	RCT for Grades 3, 5 and 7 and Quasi for Grade 8	46 third-grade students in a middle income school in Richmond, British Columbia	8 weeks/ All grades focused on problem solving. Grade specific foci included: Grade 3: whole number operations in multiplication, basic division; Grade 5: introduction to decimals, operations with decimals; Grades 7 and 8: decimals, ratios, and percents.	Regular instruction plus calculator-specific materials vs. Regular instructional activities	Researcher designed problem-solving post-test (10 items)—correct answer measure was used to calculate effect sizes (other measure available were problems attempted and correct operation used)	Third grade	0.522	~	0.296
		33 fourth-grade students in a middle income school in Richmond, British Columbia				Fifth grade	-0.507	(ns)	0.345
		47 seventh-grade students in a middle income school in Richmond, British Columbia				Seventh grade	0.227	(ns)	0.288
		54 eighth-grade students in a middle income school in Richmond, British Columbia				Eighth grade	0.344	(ns)	0.270
Wheatley, 1980[c]	Quasi	44 sixth-grade students in two classes (same teacher) in an elementary school in a Midwestern university town	6 weeks/ Problem solving	Problem-solving with calculators vs. Problem-solving intervention without calculators	Process score (processes used to solve problems)	Overall	0.353	(ns)	0.689
Heterogeneity									
Q-value	*df (Q)*	*P-value*	*I-squared*	*Pooled ES: student level (four studies, seven effect sizes)*			*Hedge's g*	*Standard Error*	
9.105	*6*	*0.168*	*34.101*				*0.304*	*~*	*0.167*

Continued on p. 6-133

Table 25, continued

Study	Design	Sample	Duration/ Content	Contrast	Measure	Subgroup	Hedge's g	Standard Error	
Classroom-level analyses									
Duffy & Thompson, 1980	RCT	Approx. 135 students in 20 fourth-grade classrooms in Columbus, OH	26 weeks/ Application problems, decimals, rounding, estimation	Calculators only plus regular mathematics program vs. Regular mathematics curriculum (see Table C-9 for effects of calculator plus instructional packages for teachers, plus regular mathematics program)	CTBS applications score	Fourth grade	-0.413	(ns)	0.433
		Approx. 122 students in 18 fifth-grade classrooms in Columbus, OH				Fifth grade	-0.161	(ns)	0.440
		Approx. 129 students in 19 sixth-grade classrooms in Columbus, OH				Sixth grade	0.178	(ns)	0.440
Szetela & Super, 1987	Quasi	Approx. 424 students in 21 seventh-grade classrooms in an urban-rural district in Canada	One school year/ Problem solving	Problem solving with calculators vs. Problem-solving intervention without calculators (see Table C-9 for larger positive effects where calculator group was able to use calculators)	Combination of two researcher designed problem solving measures: translation problems (20 items) and process problems (20 items)	Overall	0.140	(ns)	0.424
Heterogeneity									
Q-value	*df (Q)*	*P-value*	*I-squared*	*Pooled ES: student level (two studies, four effect sizes)*			*Hedge's g*	*Standard Error*	
1.230	*3*	*0.746*	*0.000*				*-0.063*	*(ns)*	*0.217*

~ p < .10, * p < .05, ** p < .01, *** p < .001
[a] Data were adjusted for clustering that occurred within classrooms.
[b] Data were adjusted for clustering that occurred within schools.
[c] The treatment group was allowed to use a calculator during assessment for this outcome.

Table 26: Studies That Investigate the Effects of Calculators on Concept Outcomes

Study	Design	Sample	Duration/ Content	Contrast	Measure	Subgroup	Hedge's g	Standard Error	
Concepts Outcomes									
Student-level analyses									
Campbell & Virgin, 1976[b]	Quasi	252 fifth- and sixth-graders in two North York elementary schools (Canada).	7 months/ Basic computation	Calculators to check work vs. No calculators	MAT concepts score	Overall	0.129	(ns)	0.642
Graham & Thomas, 2000[a]	Quasi	84 students in Grades 9 and 10 in two schools in New Zealand	3 weeks/Algebra	Graphic calculator used to learn algebraic variables vs. Standard algebra instruction	Kuchmann (1981) designed to measure algebraic understanding	Overall	0.328	(ns)	0.489
Standifer & Maples, 1981[a]	RCT	141 students in six third-grade classrooms in Monroe, LA	11 weeks/ Computation	Hand-held, four function calculator vs. No calculator in regular mathematics curriculum (see Table C-9 for effects of programmed feedback calculator vs. No calculator in regular mathematics curriculum)	Science Research Associates Assessment: computation score	Overall	-0.076	(ns)	0.395
Standifer & Maples, 1982[a]	RCT	113 students in 10 third- and fourth-grade classrooms in Monroe, LA	5 months/ Computation	Hand-held, four function calculator vs. General remedial mathematics curriculum (see Table C-9 for effects of experimental group 2 using programmed-feedback calculators + regular remedial curriculum)	Science Research Associates Assessment: computation score	Overall	0.546	(ns)	0.332
Heterogeneity									
Q-value	*df (Q)*	*P-value*	*I-squared*	*Pooled ES: student level (4 studies, 4 effect sizes)*			*Hedge's g*	*Standard Error*	
1.519	*3*	*0.678*	*0.000*				*0.278*	*(ns)*	*0.213*
Classroom-level analyses									
Duffy & Thompson, 1980	RCT	Approx. 135 students in 20 fourth-grade classrooms in Columbus, OH	26 weeks/ Application problems, decimals, rounding, estimation	Calculators only plus regular mathematics program vs. Regular mathematics curriculum (see Table C-9 for effects of calculator plus instructional packages for teachers, plus regular mathematics program)	CTBS concepts score	Fourth grade	0.063	(ns)	0.428
		Approx. 122 students in 18 fifth-grade classrooms in Columbus, OH				Fifth grade	0.221	(ns)	0.440
		Approx. 129 students in 19 sixth-grade classrooms in Columbus, OH				Sixth grade	0.103	(ns)	0.439
Heterogeneity									
Q-value	*df (Q)*	*P-value*	*I-squared*	*Pooled ES: student level (one study, three effect sizes)*			*Hedge's g*	*Standard Error*	
0.071	*2*	*0.965*	*0.000*				*0.128*	*(ns)*	*0.252*

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

[a] Data were adjusted for clustering that occurred within classrooms.

[b] Data were adjusted for clustering that occurred within schools.

[c] The treatment group was allowed to use a calculator during assessment for this outcome.

4. Computer Programming

One of the early uses of educational technology in education was engaging students in programming computers as a way to explore, learn, or apply and practice mathematical ideas. For example, the original developers of Logo developed this programming language to serve as a conceptual framework for learning mathematics (Feurzeig & Lukas, 1971; Papert, 1980). Classroom observations suggested that children use certain mathematical concepts in Logo programming. As an illustration, first-graders use such mathematical notions as number, arithmetic, estimation, measure, patterning, proportion, symmetry, inversion, and compensation (Kull, 1986). Similar observations of intermediate graders indicated that Logo may make it possible to explore certain mathematical concepts, such as angle measure or recursion, earlier than is currently believed (Carmichael, Burnett, Higginson, Moore, & Pollard, 1985; Papert, Watt, diSessa, & Weir, 1979). Here the Task Group investigates whether engaging students in computer programming has significant effects on their mathematics achievement and problem-solving ability.

a. Prior Syntheses and Meta-Analyses

Detailed effect size information from prior quantitative syntheses and meta-analyses are presented in Table C-5 in Appendix C. Previous reviews (see summary Table 27) indicate that programming improves students' performance compared to conventional instruction, with the greatest effects on concepts and applications, especially geometric concepts, and weaker effects on computation. They also have indicated that programming positively affects problem solving, as well as attitudes toward mathematics and instruction in mathematics, more so than other software categories. On the basis of prior syntheses and meta-analyses, computer programming appears to have the same effectiveness at various grade levels. There is some evidence it is more effective for students of average, rather than low or high socioeconomic status (SES). Earlier syntheses and meta-analyses have argued that programming is somewhat more effective when it supplements, rather than replaces, conventional instruction, consistent with suggestions for mediated instruction of programming. Certain computer languages, especially the Logo computer language, have stronger positive effects than other computer languages. Other syntheses have similarly concluded that direct teacher involvement and better designed languages result in better instruction, and provide more guidance for instruction (Clements & Sarama, 1997). As with other types of software, Logo programming can be particularly effective when embedded in a curriculum and then in a context that includes professional development for teachers.

Table 27: What Prior Reviews Say About Programming Interventions

- General Findings
  - — Logo programming can increase students' mathematical achievement, especially if it is integrated into a coherent curriculum with teacher mediation (Clements & Sarama, 1997)
  - — The median pooled effect size for mathematics achievement across the meta-analyses is 0.35; for problem solving, the median is 0.285 (see Table C-5)
  - — Impacts more likely on concepts and applications as opposed to computation
  - — Positive effects on attitudes toward mathematics and instruction
- Contextual variables
  - — Differences favoring elementary school-age (vs. secondary) in achievement (similar for problem solving)
  - — Differences favoring average SES students vs. either high or low SES
  - — No consistent differences in ability level
- Implementation variables
  - — Differences favoring shorter duration programs (up to 18 weeks; based on only one meta-analysis)
  - — Differences favoring programs that supplement rather than substitute for other mathematics instruction for problem solving (substitution is slightly higher for achievement). Narrative reviews conclude that better outcomes result from curriculum integration and mediated teaching
  - — Differences favoring computer programs designed to support learning (such as Logo)

b. The Task Group's Meta-Analysis of Computer Programming Interventions

Table 28 presents the studies in the Task Group's meta-analysis of high-quality experimental and quasi-experimental studies on the effects of students' engaging in computer programming on their achievement. From all the studies reviewed, only 14 met the criteria for inclusion. These 14 studies yielded a total of 30 effect sizes. Of these, 24 were positive, 1 negative, and 5 near zero, with a mean pooled effect size on combined outcome measures of 0.674, which was statistically significant. Those assessing mathematics achievement only had a mean pooled effect size of 0.698, which also was statistically significant. (An important note is that some of these interventions involved changes in curriculum, using technology, but also altering content and teaching.) Those that assessed problem solving ability had a mean pooled effect size of 0.518, which was also statistically significant.

Although only two studies (Johnson-Gentile et al., 1994; Ortiz & MacGregor, 1991) reported effects on retention, both reported a larger effect size for the delayed, compared to the immediate, posttests (1.901 immediate, significant ES; 2.410 delayed for Johnson-Gentile et al.; 0.437 immediate, bordering on significant .898 for Ortiz & MacGregor). These findings suggest that computer programming, possibly due to the more extensive processing (due to the programming activity per se) over multiple modalities (e.g., numeric or symbolic and visual or graphic) or the ability to actively submit one's ideas for evaluation and feedback (e.g., did the program run as expected) facilitates students' development of higher level conceptual structures. That is, computer programming requires a complete, precise, and abstract explication, potentially leading to conceptually richer concepts. Students specify steps to a noninterpretive agent, with thorough specification and detail, then observe, reflect on, and correct. The computer serves as an explicative agent.

Several of these studies also compared computer programming to a CAI-based treatment, and so were not included in the basic meta-analysis in Table 28. Showing consistently higher scores for the computer programming than the CAI groups (but none reaching levels of statistical significance), these contrasts can be found in Appendix C in Table C-10. Several contextual and implementation variables may have contributed to the inconsistency.

Contextual variables. These studies about programming yielded the following results regarding contexts (Table 20).

- *Age or grade.* Effects were significantly higher when used with elementary school students than with middle school students, consistent with previous reviews.

Implementation variables. These studies yielded the following regarding implementation variables.

- *Duration.* There were no significant differences for interventions of different durations.

- *Substitute versus supplement.* Both substitution and supplementation programming treatments had statistically significant positive effects (ES 0.721 and ES 0.655), and the differences in effects between these two types of treatments were not significant.

- *Level of integration.* The differences across subcategories of curricular integration also did not reach statistical significance, but there is a clear pattern of effect sizes in which stronger effects are related to high (0.682, significant) or medium (0.739, significant), compared to low (-0.065, not significant) integration.

The Task Group's meta-analysis of rigorous studies on the effects of computer programming on mathematics achievement supports the conclusions of the previous syntheses, with a significant mean pooled effect size of 0.698 for mathematics achievement and 0.518 for problem solving (See Table 28) (compare to the median pooled effect sizes of 0.35 and 0.258, respectively, for the previous meta-analyses). Effects were higher for elementary school students than for older students. There is a suggestion that greater curricular integration yields stronger positive effects. Further, this meta-analysis suggested a result not previously revealed—that results for delayed posttests might be greater than those for immediate posttests.

Table 28: Studies That Examine Effects of Computer Programming on Mathematics Achievement

Study	Design	Sample	Duration/ Content	Contrast	Measure	Hedge's g	Standard Error	
Programming								
Battista & Clements, 1986	RCT	12 fourth-grade students in a midwestern middle school	42 sessions (two 40-min per week)/ LOGO	Logo vs. C (computer literacy)	Problem-solving Tests 1&2, Combined	0.660		0.609
		26 sixth-grade students in a midwestern middle school				0.049		0.392
Blume & Schoen, 1988	QED	50 eighth-graders in two midwestern junior high schools	A semester-long class/ BASIC	Basic vs. C	Combined problem solving and logic	-0.065		0.444
Clements, 1986	RCT	24 first-grade students from a middle-class midwestern school system	44 sessions (22 weeks)/ LOGO	Logo vs. C	WRAT Mathematics score	1.072	*	0.423
		24 third-grade students from a middle-class midwestern school system				0.636		0.405
Clements et al., 2001	QED	51 Kindergarten students in a school near Kent, OH	Incorporated into classes over entire academic year/ LOGO and geometry	Logo vs. Control	Geometry	2.842	***	0.609
		71 Kindergarten students in a school near Buffalo, NY				0.121		0.495
		87 first-grade students in a school near Kent, OH				0.938	~	0.495
		92 first-grade students in a school near Buffalo, NY				0.394		0.486
		103 second-grade students in a school near Kent, OH				0.009		0.481
		96 second-grade students in a school near Buffalo, NY				1.457	**	0.502
		56 third-grade students in a school near Kent, OH				0.571		0.511
		47 third-grade students in a school near Buffalo, NY				0.674		0.524
		158 fourth-grade students in a school near Kent, OH				0.184		0.470
		92 fourth-grade students in a school near Buffalo, NY				0.353		0.486
		103 fifth-grade students in a school near Kent, OH				0.093		0.481
		95 fifth-grade students in two schools, one near Kent, OH (Site 1) and one in Buffalo, NY (Site 2)				0.982	*	0.492
		141 sixth-grade students in a school near Kent, OH				0.526		0.474
		108 sixth-grade students in a school near Buffalo, NY				0.011		0.479
Clements & Battista, 1989	RCT	48 third-grade students of seven teachers from a middle class midwestern school	78 sessions, three 45-55 min per week (26 weeks)/ LOGO	Logo vs. C (computer composition/music + some Logo)	Combined posttest	1.495	***	0.328
Degelman et al., 1986	RCT	15 Kindergarten students attending a private day care center	15 minutes/day for five weeks/ LOGO	Logo vs. C	Problem solving (proportion correct)	1.284	*	0.576

Continued on p. 6-139

Table 28, continued

Study	Design	Sample	Duration/ Content	Contrast	Measure	Hedge's g	Standard Error	
Emihovich & Miller, 1988	RCT	24 first-grade students in five classrooms in an elementary school in the southeast	20, 30-min sessions (3 months)/ LOGO	Logo vs. C	CTBS - Mathematics	0.719	~	0.408
Johnson-Gentile et al., 1994	QED	150 fifth- and sixth-graders in six classrooms in two schools, one urban and one suburban	8 class days/ LOGO and geometry	Logo vs. control	Logo Geometry Motions Unit Posttest	1.901	***	0.420
Kapa, 1999	RCT	15 fifth-grade students from four classes in two elementary schools in Tel Aviv, Israel, working individually	twice per week for 45 min for semester/ LOGO-STAT-programming and graphing	individual: LOGO-STAT vs. C (Q-Text, linguistic problem-solving)	Problem solving (range 1–6)	0.520		0.496
Kapa, 1999[a]		30 fifth-grade students from four classes in two elementary schools in Tel Aviv, Israel, working in pairs		pairs: LOGO-STAT vs. C (Q-Text, linguistic problem-solving)		0.697	~	0.412
Lehrer & Randle, 1987	RCT	24 first-grade students in a low SES New York City school	35 sessions, twice per week 20–25 min, 5 months/ LOGO	Logo vs. C	TOH (avg of TOH 1–3)	1.254	**	0.434
Oprea, 1988[a]	QED	54 sixth-grade students in three schools in a small midwestern city	6 weeks/ BASIC applied to mathematics content	Wholistic BASIC vs. C	Mathematical Generalization Instrument	0.391		0.679
Ortiz & MacGregor, 1991	RCT	59 sixth-grade students from four classrooms in two metropolitan area public schools	5, 50-min sessions/ LOGO and the concept of a variable	Logo vs. textbook-based instruction on concept of variable	Concept of variable instrument	0.437	~	0.260
Thompson & Wang, 1988[a]	QED	40 sixth-grade students from two classrooms (taught by the same teacher) in one school	3, 45-min sessions/ LOGO and graphing skills	Logo vs. C	Posttest-Cartesian coordinates	0.538		0.696
Turner & Land, 1988[a]	QED	153 middle school students in seven classrooms in four inner-city midwestern public schools	1hr/week, 16 weeks/ LOGO and angles and distance, variables, rectangular coordinate systems, negative numbers, etc.	Logo vs. C	Mathematics Multiple Choice Posttest	-0.267		0.471
Heterogeneity								
Q-value	*df (Q)*	*P-value*	*I-squared*	*Pooled ES (14 studies, 30 effect sizes)*		*Hedge's g*	*Standard Error*	
53.503	*29*	*0.004*	*45.797*			*0.674*	***	*0.115*
Mathematics outcomes only								
Heterogeneity								
Q-value	*df (Q)*	*P-value*	*I-squared*	*Pooled ES (9 studies, 23 effect sizes)*		*Hedge's g*	*Standard Error*	
45.078	*22*	*0.003*	*51.196*			*0.698*	***	*0.138*
Problem-solving outcomes only								
Heterogeneity								
Q-value	*df (Q)*	*P-value*	*I-squared*	*Pooled ES (six studies, eight effect sizes)*		*Hedge's g*	*Standard Error*	
8.512	*7*	*0.290*	*17.765*			*0.518*	**	*0.169*

~ p < .10, * p < .05, ** p < .01, *** p < .001
[a] Data were adjusted for clustering that occurred within classrooms.

Note: The studies with mathematics outcomes are Clements (1986), Clements et al. (2001), Clements & Battista (1989), Emihovich & Miller (1988), Johnson-Gentile et al. (1994), Oprea (1988), Ortiz & MacGregor (1991), Thompson & Wang (1988), and Turner & Land (1988).

The studies with problem-solving outcomes are Battista & Clements (1986), Blume & Schoen (1988), Clements (1990) (not part of main pooled analysis; it was based on the same study or sample as Clements & Battista (1989), Degelman et al. (1986), Kapa (1999), and Lehrer & Randle (1987).

Based on the small number of studies in the subsequent categories, the Task Group did not conduct a meta-analysis of studies in these categories. The findings of prior syntheses and reviews are briefly presented to guide future research.

5. Tools: Computer—Existing Reviews

Studies in a broader and more ill-defined category of technology, software tools and exploratory environments (excluding calculators which were discussed above), were found to have inconsistent effects on student performance when compared to conventional classroom instruction in the synthesis of existing reviews. Detailed effect size information from prior quantitative syntheses and meta-analyses are presented in Table C-6 in Appendix C. One review reported that problem solving software appeared as effective as other categories of software (Edwards et al., 1975). However, pooled effect sizes reported in other meta-analyses have tended to be low, including 0.04 for tool and exploratory environments (Lou et al., 2001, who emphasize that commercial tests may underestimate effects), 0.10 for "computer-enhanced instruction" (a broad interpretation, Kulik & Kulik, 1991), and 0.24 for secondary students' use of problem-solving software (Kuchler, 1999).

In contrast to these limited effects, a recent randomized trials evaluation of a middle-school mathematics approach in which software is a key component reported a larger effect size (Roschelle et al., 2007). The approach focuses on proportionality, with software that connects different representational systems; for example, linking visual forms such as graphs and simulated motions to linguistic forms such as algebraic symbols and narrative stories of motion in an interactive and expressive context. The approach also embeds the software within a curriculum and includes professional development for the teachers, which may account for its success. Caveats include the short duration of the study (less than a month) and the participation of all volunteer teachers; for these reasons this study did not meet the Task Group's inclusion criteria.

Based on the small number of studies for any particular subcategory of tools and exploratory environments, the Task Group did not conduct a meta-analysis of this category except for one specific type of tool, the calculator, discussed previously.

6. Simulations and Games—Existing Reviews

Detailed effect size information from prior meta-analyses on simulation and games are presented in Table C-7 in Appendix C. The prior syntheses of the effects of simulation and game software revealed inconsistent effects on student performance when compared to conventional classroom instruction, with three previous meta-analyses providing a median pooled effect size of 0.23. All specific findings come from only one of these meta-analyses; thus, all results are tentative.

Table 29: What Prior Reviews Say About Simulations and Games

- General findings
  - — Larger effects for computation or a combination of goals than for concepts and applications (but based on one meta-analysis, with small numbers of effect sizes; see Table C-7)
  - — Arithmetic and "general" subjects showed higher effects than geometry and algebra
  - — Attitudes toward mathematics and instruction positively affected by use of simulation software
- Contextual variables
  - — Junior high students benefited more than elementary students
  - — Simulations appear more effective for males
- Implementation variables
  - — Effects were greater for studies of 1–18 weeks compared to those of 19–36 weeks duration
  - — Higher effects of supplemental use on achievement than substitution.
  - — Substitutions shows a negative effect on problem solving
  - — No differences between experimenter or teacher-developed and commercial software
  - — Higher gains in a context that combines guidance both with the subject matter content (e.g., other forms of instruction) and with students' interaction with the simulation

7. Internet

There is no consistent empirical research base on the many types of learning and teaching that can be delivered or supported over the Internet. Two categories of software that appear to have tentative support, based on previous syntheses, are online learning and Web-based inquiry (e.g., Fadel & Lemke, 2006). Possible negative effects of using the Internet for mathematics and mathematics instruction also need to be researched. There were an insufficient number of original empirical studies to conduct an original meta-analysis on the use of the Internet in mathematics instruction.

8. Tools for Teachers

Such tools as electronic blackboards and quick-response devices have mostly descriptive studies to support them (e.g., Fadel & Lemke, 2006). The application of computer-managed instruction (CMI) has already been discussed as a component of ILSs. In addition, direct studies of CMI show a pooled effect size of 0.14 in a previous meta-analysis (Kulik, 1994). There were an insufficient number of original empirical studies to conduct an original meta-analysis on this topic.

D. Conclusions and Implications

This review summarized what is known about the role of technology, including different categories of computer software and calculators, in mathematics instruction and learning. Before reviewing the findings, several general issues are discussed.

Some reviewers have decried that the literature on educational technology is too inconsistent and uneven to make “sweeping conclusions about the effectiveness of instructional technology” (Kulik, 2003). Both a conceptual analysis and empirical review concur that any such sweeping conclusions are not warranted, but also suggest that such conclusions should not be sought as guides for educational practice. “Technology” is not a single, monolithic entity (Clements & Sarama, 2003). This review has shown different effects for different categories of software, has identified contextual and implementation variables, and whenever possible has distinguished between different applications of computer technology. However, the present research corpus is weak in distinguishing the effects of specific features of software categories and specific software applications (such as in Table 17; this major gap in research will be discussed in the succeeding section, “Instructional Software: Features and Pedagogical Strategies”). There are too few studies on documented implementations of specific strategies for educational technology, and even fewer studies on particular educational technology programs. Longitudinal studies are also needed.

Although some previous meta-analyses identified their effects (e.g., 0.19 to 0.24) as “weak,” any such classification is dubitable, because the importance of any pooled effect size depends on a variety of factors (Lipsey & Wilson, 2001). For CBI, one particular issue is that students are often maximally engaged with the computer materials for 15–30 minutes two to three times per week. Pooled effect sizes must be interpreted in that context (Slavin & Lake, 2007).

Existing research, and the many available reviews of this body of research, suggests that specific categories and uses of educational technology can make a significant, positive contribution to students’ learning of mathematics. The Task Group conducted its own meta-analyses to evaluate those conclusions of previous reviews.

1. Drill and Practice

Prior syntheses and meta-analyses suggest that CAI drill and practice generally improves students’ performance compared to conventional instruction, with the greatest effects on computation, and more limited on concepts and applications. It is the only category of instructional software that shows, in previous reviews, higher effects for serving as a substitute for conventional instruction, rather than as a supplement to it. It may be that such programs address students’ instructional needs for practice adequately and efficiently, making substantial teacher intervention less important.

The Task Group’s meta-analysis of rigorous studies supports these conclusions. Drill and practice software had a significant positive effect on mathematics achievement. When analyzed for different ages and grades, positive effects were confirmed for the elementary level, but there were too few studies at other levels to make comparisons or conclusions. Effect sizes were higher for interventions that substituted for, rather than supplemented, classroom practice.

In summary, drill and practice through high-quality CAI, implemented with fidelity, can be considered a useful tool in developing students' automaticity, or fast, accurate, and effortless performance on computation, freeing working memory so that attention can be directed to the more complicated aspects of complex tasks. A caveat is that older studies may have used software better designed to use research-based strategies (and fewer "bells and whistles," graphics and sound not related to instruction) than many more recently published programs. Using such strategies to incorporate features such as those in Table 17 will likely maximize positive effects. The following section includes additional caveats relevant to drill and practice.

2. Tutorials

Prior syntheses and meta-analyses suggest that CAI tutorials improve students' performance compared to conventional instruction, with slightly greater effects on performance on concepts and applications measures than on computation measures. Based on these prior syntheses and meta-analyses, tutorials appear to be effective at all grade levels, particularly the higher grades. Reviews indicate that they are more effective when they supplement, rather than replace, conventional instruction, when they involve experimenter or teacher-developed, rather than commercially developed, software, and when they are developed for a specific audience rather than a general audience.

The Task Group's meta-analysis of rigorous studies similarly indicates that tutorials can increase mathematics performance, both overall achievement and, possibly more so, mathematical problem-solving ability. It supported the conclusion that tutorials are more effective as supplements, rather than replacements, for conventional instruction and when they are highly integrated with the regular mathematics curriculum. Finally, tutorial software developed by researchers or teachers was more effective than that developed by commercial companies. Findings of individual studies provide serious caveats, however, including the need to consider empirical evidence of effectiveness of a particular software package, and issues of scale-up, including integration with the extant curriculum, and fidelity of implementation, including amount of use, and technological and pedagogical support.

In summary, tutorials, as well as software packages that combine tutorials with drill and practice, that are well designed (e.g., including features in Table 17; see also Clements, 2007; Clements & Battista, 2000) and implemented can be considered as potentially useful tools in introducing and teaching specific subject-matter content to specific populations, especially at the junior and senior high school levels. Research suggests that tutorials be designed to develop specific educational goals for specific populations. Caveats are that results are not guaranteed, and care must be taken that there is evidence that the software increases learning and that the requisite support conditions to use the software effectively are in place.

3. Tools: Calculators and Graphing Calculators

Prior syntheses and meta-analyses suggest that calculators of all types, basic, scientific and graphing, may benefit students' achievement in (and attitudes toward) mathematics, and effects are more positive when calculators are used during testing. Previous reviews also indicate that effects of calculator use on calculation, contrary to perhaps the most common concern, are near zero but positive (even when calculators are not allowed on the assessments), and effects on problem solving were positive.

The Task Group's meta-analyses of 11 studies that met the Panel's rigorous criteria (only one study less than 20 years old) found limited to no impact of calculators on calculation skills, problem-solving, or conceptual development. Effect sizes of these studies are lower than those in prior syntheses and meta-analyses. On the basis of the high quality studies identified in this category by the Task Group, it is reasonable to conclude that there is no significant negative impact of calculators on students' calculation competence (only one of the studies allowed students to use calculators on the assessment). However, there are several important caveats. These findings are limited to the effect of calculators as used in the 11 studies, including studies up to a year in duration. Also, tests of computational skills did not measure the more basic processes, such as retrieval or decomposition, that students use to solve arithmetic problems, nor did they measure automaticity or procedural execution as might be assessed with timed paper-and-pencil tests (see the Learning Processes Task Group report). This is especially important when arithmetic skills are being formed, because inappropriate calculator use may interfere with the development of these skills. On the other hand, it is possible that appropriate calculator use could provide useful feedback and build a stronger association between addends and their sum, strengthening these associations. Especially given these conflicting possibilities, and the importance of this early development, the lack of rigorous studies with students earlier than third grade is especially unfortunate. Also, research on calculator use over several years—especially comparing inappropriate and appropriate use—is direly needed.

Further, given that the basic computational skills of many Americans are poor, as described in the Learning Processes report, a finding of no effect is not a promising one; more powerful instructional approaches are needed. The synthesis of previous reviews suggests that more recent calculator interventions, especially those putting calculators to "pedagogical use" as an essential element in the teaching and learning of mathematics, have a greater positive effect (the studies in the Task Group's meta-analysis did not report such comparisons). "Pedagogical use" usually implies extending mathematics learning in certain situations (and perhaps using calculators to check the accuracy of mental or other calculations), rather than using calculators when other methods would be appropriate. The overuse and inappropriate use of calculators, decried by many, may be more harmful than these (relatively short-term) studies indicate. On the other hand, an emphasis on mental arithmetic may ameliorate such problems. There is much researchers still need to study.

This report has not addressed several important educational issues. There is a dearth of research not only on broad categories of calculator use such as "functional vs. pedagogical" use, but on specific uses of calculators that may lead to negative effects (e.g., overdependence), null effects, or specific positive effects. Such research should also fill the gap in the literature if studies included *observation* of how, how much, and how well, calculators are used (this includes, for planned interventions, “fidelity of implementation” measures).

In a similar vein, the older studies in the Task Group’s meta-analysis, and more recent calculator studies, classify measures and findings by broad categories only, such as “calculation,” “problem solving,” and “concepts.” Greater specificity in terms of grade level and topic, instructional goals, and pedagogical strategy would yield more useful research results and implications. Would specifically targeted use, in which the calculator’s unique characteristics are used intentionally, result in greater benefits? For example, one might only introduce calculators in work with arithmetic with numbers of six or more digits, square roots, or scientific notion. Another project might introduce calculators in earlier grades, not to replace computational practice (mental and paper-and-pencil arithmetic) but rather to extend computational and problem-solving proficiency.

Even more fundamental, although some may argue against calculator use because it circumvents the mathematics they wish students to perform, others believe that in an age of calculators and computers, it is inappropriate to continue to focus the elementary school mathematics curriculum on pencil-and-paper arithmetic (Ralston, 1999). Research cannot address such curriculum issues of goals and values, although, it should be explicit about its assumptions. Research can clarify the ramifications of various approaches, but the work of discussing these approaches, and evaluating them through empirical research, largely remains to be done.

In summary, most of the effects in the Task Group’s meta-analysis have a similar pattern of results to those in the prior syntheses and meta-analyses, but with smaller, and usually near-zero, statistically insignificant effect sizes. Given the design flaws noted in some studies included in previous meta-analyses, this may indicate that the smaller effect sizes represent more accurate estimates of calculators’ effects. However, there are different, but still substantial, limitations to the pool of studies that met the criteria for inclusion in the present meta-analyses. First, only 11 studies of the hundreds in the literature are included in the Task Group’s meta-analysis. Only one was published after 1987, and that had but one comparison, at Grades 9 and 10. This could be important, as a previous meta-analysis indicated that effects of calculators may be becoming more positive with time (Ellington, 2003), which may suggest that technology, and especially support materials and professional development related to technology use, are improving since the introduction of calculators. Recent calculator interventions that use research-based approaches (e.g., embedding technology within a curriculum and targeting calculator use to particular pedagogical ends) to incorporate newer technologies provide suggestive results (Stroup, Pham, & Alexander, 2007). Second, most of the studies included in the Task Group’s meta-analysis measured computational skills, but only half assessed the learning of problem solving or concepts. Thus, many of the comparisons for a specific effect are from a small number of studies (with effects pooled from multiple comparisons frequently originating from a single study). Given the different limitations of each report, conclusions that they share appear trustworthy—that is, calculators as used in these studies have little or no effect on most measured outcomes in calculation or concepts, given the manner in which calculators were used and the duration of the studies—but, especially when outcomes differ, a larger body of more recent, rigorous studies that documents *how* calculators are used, including research that examines multiyear use of calculators, is needed before firm conclusions can be reached.

4. Computer Programming

Computer programming by students can be employed in a wide variety of situations using distinct pedagogies. Prior syntheses and meta-analyses indicate that programming improves students' performance compared to conventional instruction, with the greatest effects on concepts and applications, especially geometric concepts, and weaker effects on computation. Previous reviews also indicate that programming positively affects problem solving, as well as attitudes toward mathematics and instruction in mathematics, more so than other software categories. Prior reviews also provide some evidence that use of computer programming is more effective for students of average, rather than low or high SES. Earlier reviews have claimed that programming is somewhat more effective when it supplements, rather than replaces, conventional instruction, consistent with suggestions for mediated instruction of programming. Certain computer languages, especially the Logo computer language, were reported to have stronger positive effects than other computer languages.

The Task Group's meta-analysis of rigorous studies supports the conclusions of previous reviews about the impact of computer programming on mathematics performance. Further, the meta-analysis suggested that results for delayed posttests might be greater than those for immediate posttests. Additional research is needed to ascertain whether this finding is generalizable.

In summary, computer programming can be considered an effective tool, especially for elementary school students, for developing specific mathematics concepts and applications and mathematical problem-solving abilities. Effects may be larger the more computer programming is integrated into the curriculum. Although there was insufficient research on such issues, the Task Group notes that instructional use of programming has fewer "bells and whistles" than other categories of software and demands thoughtful curricula and knowledgeable teachers, all of which may have contributed to the lack of frequency in U.S. classrooms (it is more widely used in other countries, Clements & Sarama, 1997). Dissemination of research, including research-based curricula and professional development, could lead to a reversal of this trend.

5. Tools: Computer Tools

Software tools and exploratory environments (excluding calculators) have inconsistent effects on student performance. Prior syntheses and meta-analyses suggest that problem-solving software may have potential, but effect sizes have been small. Recent rigorous studies suggest that new approaches may have promise, but there are an inadequate number of such studies for the Task Group to conduct a meta-analysis of this software category.

6. Simulations and Games

Prior syntheses and meta-analyses suggest that simulation and game software packages may have positive, but relatively small, effects on student performance when compared to conventional classroom instruction. Previous studies also have shown them to have a positive effect on attitudes. Junior high, more than elementary, students may benefit from working with simulations and games. Supplemental use is indicated, consistent with the intrinsically unguided nature of simulations and games.

In summary, there is only slight evidence—based on studies of unknown rigor—indicating that simulations may be useful, especially at the middle or junior high level, to develop skills, concepts, and applications of knowledge in problem-solving settings. More needs to be known about developing and using this category of software, but it is likely that careful integration into a well-structured curriculum is critical to facilitate learning.

7. Instructional Software: Features and Pedagogical Strategies

Many questions essential to designing and selecting educational technology applications cannot be answered, because studies and reviews do not distinguish such applications on their use of specific features. Similar situations exist for practice, the role of the teacher (especially specific pedagogical strategies).

a. Software Features

The Task Group's reviews found that the previous meta-analyses and rigorous studies did not permit generalizations about critical features of software, such as those identified in Table 17. That is, prior syntheses and meta-analyses do not sufficiently distinguish such applications on their use of specific features that theoretically should contribute to learning. Such findings would be invaluable to the field, both because decisions could be guided by any software program's inclusion of critical features and because the development of new software programs could be similarly guided.

Only for the sake of illustration, a few studies that did not meet the Task Group's criteria are described here that compare CAI conditions, most of which had varied conclusions. One study reported that enhancing drill by placing it within a game context does not yield significantly different outcomes overall, but the game may distract students with learning disabilities (Christensen & Gerber, 1990). Enhancement with multimedia significantly improved learning in one study (Macaulay, 2003) but CAI with animated vs. static pictures or with or without the presentation of a cognitive strategy were equally effective (Shiah, Mastropieru, Scruggs, & Mushinski Fulk, 1994). Verbal guidance (in students' first language) may support learning from multimedia educational games (Moreno & Duran, 2004). These are single studies with little conceptual overlap; the field needs more complete and reliable guidance.

Few software programs are designed based on explicit (i.e., published) theoretical and empirical research foundations (but see Clements, 2007; Clements & Sarama, 2007a; Ritter, Anderson, Koedinger, & Corbett, 2007). More continuous, committed, iterative research and development projects are needed in this area. Research-based iterative cycles of evaluation and development, fine tuning software's mathematics and pedagogy within each cycle, can make substantial differences in learning (e.g., see Aleven & Koedinger, 2002; Clements & Battista, 2000; Clements et al., 2001; Laurillard & Taylor, 1994; Steffe & Olive, 2002).

Such research could identify how and why software designs could be improved. As one example, the pooled effect sizes in the Task Group's meta-analysis actually might be an underestimate of what can be achieved if drill and practice software were more carefully designed. Few studies use empirically validated strategies such as adaptive feedback and increasing ratio review (Siegel & Misselt, 1984).

8. Final Words

In most cases, specific uses of technology will not facilitate learning optimally unless they are implemented with fidelity. Unfortunately, information is lacking on this critical issue because reviewers and researchers generally have not measured fidelity. A similar situation exists for many specific pedagogical issues.

In addition, from the subtleties of designing features of software, to the complexities of scaling up approaches to work with entire educational systems, substantive challenges face researchers and other educators. These challenges must be met, and findings integrated across levels, before conclusions about the effectiveness of educational technology can be offered with confidence. Many difficulties stand in the way of conducting high-quality work in the field of technology in mathematics education. Applications that go beyond using the simplest features of technology to deliver a traditional curriculum face both (a) challenges of redesigning scope and sequences, pedagogies, software, and assessments (Kulik & Kulik, 1991), along with financial and logistical hurdles, and (b) barriers of a priori negative evaluation of the goals and assessment instruments they may wish to employ. Such barriers may have dampened innovative research and development in educational technology. This is unacceptable; concerted efforts are needed to meet these challenges and provide educators with clear guidelines from research. This is especially important given the poor implementation of educational technology in the field (Clements & Sarama, 1997; Cuban, 2001; Hoffer, Venkataraman, Hedberg, & Shagle, 2007).

Finally, technological advances continue to challenge practitioners and researchers. There is no research on questions that have arisen only recently. What technologies are most appropriate for students for whom multiple hand-held devices are a ubiquitous presence? How has the presence of Internet sites affected students (e.g., mathematics as presented on Wikipedia)? Both new questions and old must be better addressed with high-quality studies of high-quality implementations of computer-based tools if educational technology is to fulfill its potential.

BIBLIOGRAPHY

Aleven, V.A.W.M.M., & Koedinger, K.R. (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer-based cognitive tutor. *Cognitive Science, 26(2),* 147–179.

Ball, S. (1988). Computers, concrete materials and teaching fractions. *School Science and Mathematics, 88(6),* 470–475

Battista, M.T., & Clements, D.H. (1986). The effects of Logo and CAI problem-solving environments on problem-solving abilities and mathematics achievement. *Computers in Human Behavior, 2*, 183–193.

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

Blume, G.W., & Schoen, H.L. (1988). Mathematical problem-solving performance of eighth-grade programmers and nonprogrammers. *Journal for Research in Mathematics Education, 19(2),* 142–156.

Burns, P.K., & Bozeman, W.C. (1981). Computer-assisted instruction and mathematics achievement: Is there a relationship? *Educational Technology, 21(10),* 32–39.

Campbell, D.L., Peck, D.L., Horn, C.J., & Leigh, R.K. (1987). Comparison of computer-assisted instruction and print performance: a research note. *Educational Communication and Technology Journal, 35(2),* 95–103.

Campbell, P., & Virgin, A.E. (1976). *An evaluation of elementary school mathematics programs utilizing the mini-calculator.* Toronto, Canada: Ontario Department of Education.

Carmichael, H.W., Burnett, J.D., Higginson, W.C., Moore, B.G., & Pollard, P.J. (1985). *Computers, children and classrooms: A multisite evaluation of the creative use of microcomputers by elementary school children.* Toronto, Ontario, Canada: Ministry of Education.

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

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

Christensen, C.A., & Gerber, M.M. (1990). Effectiveness of computerized drill and practice games in teaching basic mathematics facts. *Exceptionality, 1(3),* 149–165.

Christmann, E., Badgett, J., & Lucking, R. (1997). Progressive comparison of the effects of computer-assisted instruction on the academic achievement of secondary students. *Journal of Research on Computing in Education, 29(4),* 325–337.

Clements, D.H. (1986). Effects of Logo and CAI environments on cognition and creativity. *Journal of Educational Psychology, 78(4),* 309–318.

Clements, D.H. (1990). Metacomponential development in a Logo programming environment. *Journal of Educational Psychology, 82(1),* 141–149.

Clements, D.H. (2007). Curriculum research: Toward a framework for 'research-based curricula'. *Journal for Research in Mathematics Education, 38(1),* 35–70.

Clements, D.H., & Battista, M.T. (1989). Learning of geometric concepts in a Logo environment. *Journal for Research in Mathematics Education, 20(5),* 450–467.

Clements, D.H., & Battista, M.T. (2000). Designing effective software. In A.E. Kelly & R.A. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp. 761–776). Mahwah, NJ: Lawrence Erlbaum Associates.

Clements, D.H., Battista, M.T., & Sarama, J. (2001). Logo and geometry. In *Journal for Research in Mathematics Education Monograph Series, 10.* Reston, VA: National Council of Teachers of Mathematics.

Clements, D.H., & Sarama, J. (1997). Research on Logo: A decade of progress. *Computers in the Schools, 14(1–2),* 9–46.

Clements, D.H., & Sarama, J. (2003). Strip mining for gold: Research and policy in educational technology—A response to "Fool's Gold". *Educational Technology Review, 11(1),* 7–69.

Clements, D.H., & Sarama, J. (2007). Effects of a preschool mathematics curriculum: Summative research on the *Building Blocks* project. *Journal for Research in Mathematics Education, 38(2),* 136–163.

Clements, D.H., & Sarama, J. (2008). Experimental evaluation of the effects of a research-based preschool mathematics curriculum. *American Educational Research Journal* (forthcoming 2008).

Cuban, L. (2001). *Oversold and underused*. Cambridge, MA: Harvard University Press.

Dalton, D., & Hannafin, M.J. (1988), The effects of computer-assisted and traditional mastery methods on computation accuracy and attitudes. *Journal of Educational Research, 82(1),* 27–33.

Degelman, D., Free, J.U., Scarlato, M., & Blackburn, J.M. (1986). Concept learning in preschool children: Effects of short-term logo experience. *Journal of Educational Computing Research, 2(2),* 199–205.

Duffy, J.A., & Thompson, G.E. (1980). *Improving mathematical achievement of elementary students through the use of electronic hand calculators*. Columbus, OH: Columbus Public Schools, Ohio Department of Management, Planning and Information.

Dynarski, M., Agodini, R., Heaviside, S., Novak, T., Carey, N., Campuzano, L., et al. (2007). *Effectiveness of reading and mathematics software products: Findings from the first student cohort*. Washington, DC: U.S. Department of Education, Institute of Education Sciences.

Edwards, J., Norton, S., Taylor, S., Weiss, M., & Dusseldorp, R. (1975). How effective is CAI? A review of the research. *Educational Leadership, 33(2),* 147–153.

Ellington, A.J. (2003). A meta-analysis of the effects of calculators on students' achievement and attitude levels in precollege mathematics classes. *Journal for Research in Mathematics Education, 34(5),* 433–463.

Ellington, A.J. (2006). The effects of non-CAS graphing calculators on student achievement and attitude levels in mathematics: A meta-analysis. *School Science and Mathematics, 106(7)*, 16–26.

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

Fadel, C., & Lemke, C. (2006). *Technology in schools: What the research says*. San Jose, CA: Metiri Group, commissioned by Cisco Systems.

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

Fletcher-Flinn, C.M., & Gravatt, B. (1995). The efficacy of computer-assisted instruction (CAI): A meta-analysis. *Journal of Educational Computing Research, 12(3),* 219–242.

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 & Practice, 17(2),* 90–106.

Fuchs, L.S., Fuchs, D., Hamlet, C.L., Powell, S.R., Capizzi, A.M., & Seethaler, P.M. (2006). The effects of computer-assisted instruction on number combination skill in at-risk first graders. *Journal of Learning Disabilities, 39(5)*, 467–475.

Gordon, B.W. (1992). *A quantitative analysis of the relationship between computer graphics and mathematics achievement and problem solving.* Unpublished doctoral dissertation, University of Cincinnati, Ohio.

Graham, A.T., & Thomas, M.O. (2000). Building a versatile understanding of algebraic variables with a graphic calculator. *Educational Studies in Mathematics, 41(3),* 265–282.

Hamilton, W.A. (1995). A meta-analysis of the comparative research on computer-assisted instruction and its effects on elementary and secondary mathematics achievement. *Dissertation Abstracts International, 56(05)*, 1599.

Hartley, S.S. (1978). Meta-analysis of the effects of individually paced instruction in mathematics. *Dissertation Abstracts International, 38*, 4003A. UMI No. 77-29, 926.

Hembree, R. (1984). *Model for meta-analysis of research in education with a demonstration of mathematics on education: Effects of handheld calculators.* Unpublished doctoral dissertation, University of Tennessee, Knoxville.

Hembree, R. (1992). Experiments and relational studies in problem solving: A meta-analysis. *Journal for Research in Mathematics Education, 23(3),* 242–273.

Hembree, R., & Dessart, D.J. (1986). Effects of hand-held calculators in precollege mathematics education: A meta-analysis. *Journal for Research in Mathematics Education, 17(2),* 83–99.

Henderson, R.W., Landesman, E.M., & Kachuck, I. (1985). Computer-video instruction in mathematics: Field test of an interactive approach. *Journal for Research in Mathematics Education, 16(3)*, 207–224.

Hoffer, T.B., Venkataraman, L, Hedberg, E.C., & Shagle, S. (2007). *Final report of the national survey of algebra teachers for the National Math Panel*. Chicago: University of Chicago.

Johnson-Gentile, K., Clements, D.H., & Battista, M.T. (1994). The effects of computer and noncomputer environments on students' conceptualizations of geometric motions. *Journal of Educational Computing Research, 11(2)*, 121–140.

Kapa, E. (1999). Problem solving, planning ability and sharing processes with LOGO. *Journal of Computer Assisted Learning, 15(1)*, 73–84.

Khalili, A., & Shashaani, L. (1994). The effectiveness of computer applications: A meta-analysis. *Journal of Research on Computing in Education, 27(1),* 48–61.

Khoju, M., Jaciw, A., & Miller, G.I. (2005). *Effectiveness of graphing calculators in K–12 mathematics achievement: A systematic review*. Palo Alto, CA: Empirical Education Research Report.

Kraus, W.H. (1981). Using a computer game to reinforce skills in addition basic facts in second grade. *Journal for Research in Mathematics Education, 12(2),* 152–155.

Kuchler, J.M. (1999). *The effectiveness of using computers to teach secondary school mathematics.* Unpublished doctoral dissertation. University of Massachusetts, Lowell.

Kulik, C.-L.C., & Kulik, J.A. (1991). Effectiveness of computer-based instruction: An updated analysis. *Journal of Computer-Based Instruction, 7(1–2)*, 75–94.

Kulik, J.A. (1994). Meta-analytic studies of findings on computer-based instruction. In E.L. Baker, & H.F. O'Neil, Jr. (Eds.), *Technology assessment in education and training* (pp. 9–33). Hillsdale, NJ: Lawrence Erlbaum Associates.

Kulik, J.A. (2003). *Effects of using instructional technology in elementary and secondary schools: What controlled evaluation studies say.* Arlington, VA: SRI International.

Laurillard, D., & Taylor, J. (1994). Designing the stepping stones: An evaluation of interactive media in the classroom. *Journal of Educational Television, 20(3),* 169–184.

Lee, W.C. (1990). The effectiveness of computer-assisted instruction and computer programming in elementary and secondary mathematics: A meta-analysis. (Doctoral dissertation, University of Amherst, 1990). *Dissertation Abstracts International, 51(03)*, 775.

Lehrer, R., & Randle, L. (1987). Problem solving, meta-cognition and composition: The effects of interactive software for first-grade children. *Journal of Educational Computing Research, 3(4)*, 409–427.

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

Lou, Y., Abrami, P.C., & d'Apollonia, S. (2001). Small group and individual learning with technology: A meta-analysis. *Review of Educational Research, 71(3)*, 449–521.

Macaulay, M. (2003). The effects of multimedia on learning in third world children. *Journal of Educational Multimedia and Hypermedia, 12 (2),* 185–198.

McCollister, T.S., Burts, D.C., Wright, V.L., & Hildreth, G.J. (1986). Effects of computer-assisted instruction and teaching instruction on arithmetic task achievement scores of kindergarten children. *Journal of Educational Research, 80(2),* 121–125.

Moore, B.M. (1988). Achievement in basic mathematics skills for low performing students—A study of teachers' affect and CAI. *Journal of Experimental Education, 57(1),* 38–44.

Moreno, R., & Duran, R. (2004). Do multiple representations need explanations? The role of verbal guidance and individual differences in multimedia mathematics learning. *Journal of Educational Psychology, 96*(3), 492–503.

National Council of Teachers of Mathematics (NCTM). (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.

Niemiec, R.P., & Walberg, H.J. (1984). Computers and achievement in the elementary schools. *Journal of Educational Computing Research, 1(4),* 435–440.

Oprea, J.M. (1988). Computer programming and mathematical thinking. *Journal of Mathematical Behavior*, *7(2),* 175–190.

Ortiz, E., & MacGregor, S.K. (1991). Effects of LOGO programming on understanding variables. *Journal of Educational Computing Research, 7,* 37–50.

Papert, S., Watt, D., di Sessa, A., & Weir, S. (1979). *Final report of the Brookline Logo Project: Parts 1 and 11* (Logo Memos Nos. 53 and 54). Cambridge, MA: Massachusetts Institute of Technology Artificial Intelligence Laboratory.

Podell, D., Tournakirein, N., & Lin, A. (1992). Automatization of mathematics skills via computer-assisted-instruction among students with mild mental-handicaps. *Education & Training in Mental Retardation & Developmental Disabilities, 27(3),* 200–206.

Ralston, A. (1999). Let's abolish pencil-and-paper arithmetic. *Journal of Computers in Mathematics and Science Teaching, 18*, 173–194.

Ritter, S., Anderson, J.R., Koedinger, K.R., & Corbett, A. (2007). Cognitive Tutor: Applied research in mathematics education. *Psychonomics Bulletin & Review, 14(2),* 249–255.

Roschelle, J.M., Tatar, D., Shechtman, N., Hegedus, S., Hopkins, B., Knudsen, J., et al. (2007). *Can a technology enhanced curriculum improve student learning of important mathematics?* (Vol. Technical Report 01). Menlo Park, CA: SRI International.

Ryan, A.W. (1991). Meta-analysis of achievement effects of microcomputer applications in elementary schools. *Educational Administration Quarterly, 27(2),* 161–184.

Saracho, O.N. (1982). The effects of a computer-assisted instruction program on basic skills achievement and attitudes toward instruction of Spanish-speaking migrant children. *American Educational Research Journal, 19(2),* 201–219.

Saunders, J., & Bell, F.H. (1980). Computer-enhanced algebra resources: Their effects on achievement and attitudes. *International Journal of Mathematical Education in Science and Technology, 11(4),* 465–473.

Schnur, J.O., & Lang, J.W. (1976). Just pushing buttons or learning?—A case for mini-calculators. *The Arithmetic Teacher, 23(7),* 559–562.

Shiah, R.L., Mastropieri, M.A., Scruggs, T.E., & Mushinski-Fulk, B.J. (1994). The effects of computer-assisted instruction on the mathematical problem solving of students with learning disabilities. *Exceptionality,* 5(3), 131–161.

Siegel, M.A., & Misselt, A.L. (1984). Adaptive feedback and review paradigm for computer-based drills. *Journal of Educational Psychology, 76(2),* 310–317.

Slavin, R.E., & Lake, C. (2007). *Effective programs in elementary mathematics: A best-evidence synthesis*. Baltimore, MD: Johns Hopkins University.

Smith, B.A. (1997). *A meta-analysis of outcomes from the use of calculators in mathematics education.* Unpublished doctoral dissertation, Texas A&M University.

Standifer, C.E., & Maples, E.G. (1981). Achievement and attitude of third-grade students using two types of calculators. *School Science and Mathematics, 81(1),* 17–24.

Standifer, C.E., & Maples, E.G. (1982). Mathematical achievement of third- and fourth-grade students in compensatory education when using two types of calculators. *Psychological Reports*, 82(51), 11–18.

Steffe, L.P., & Olive, J. (2002). Design and use of computer tools for interactive mathematical activity (TIMA). *Journal of Educational Computing Research, 27(1&2),* 55–76.

Stroup, W., Pham, V., & Alexander, C. (2007). *Richardson MathForward Project: Second Year Final Report: Mathematics TAKS Results.* Austin, TX: University of Texas at Austin.

Szetela, W. (1980). Calculators and the teaching of ratios in Grade 7. *Journal for Research in Mathematics Education, 11(1),* 67–70.

Szetela, W. (1982). Story problem solving in elementary school mathematics: What differences do calculators make? *Journal for Research in Mathematics Education, 13(5),* 381–389.

Szetela, W., & Super, D. (1987). Calculators and instruction in problem solving in Grade 7. *Journal for Research in Mathematics Education, 18(3),* 215–229.

Thomas, P.G., & Rickhuss, M.G. (1992). An experiment in the use of computer algebra in the classroom. *Education and Computing, 8(3),* 225–263.

Thompson, A.D., & Chen-Wang, H. (1988). Effects of a Logo microworld on student ability to transfer a concept. *Journal of Educational Computing Research, 4(3),* 335–347.

Trifiletti, J.J., Frith, G.H., & Armstrong, S.W. (1984). Microcomputers versus resource rooms for LD students: A preliminary investigation of the effects on mathematics skills. *Learning Disability Quarterly, 7(1),* 69–76.

Turner, S.V., & Land, M.L. (1988). Cognitive effects of a Logo-enriched mathematics program for middle school students. *Journal of Educational Computing Research, 4(4),* 443–452.

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

Wheatley, C.L. (1980). Calculator use and problem-solving performance. *Journal for Research in Mathematics Education, 11(5),* 323–334.

Wheatley, G.H., & Shumway, R.J. (1979). *Impact of calculators in elementary school mathematics: final report*. Washington, DC: National Science Foundation.

Wheeler, J.L., & Regian, JW (1999). The use of a cognitive tutoring system in the improvement of the abstract reasoning component of word problem solving. *Computers in Human Behavior*, *15(2),* 243–254.

Wilson, W.S., & Naiman, D.Q. (2004). K–12 calculator usage and college grades. *Educational Studies in Mathematics, 56(1),* 119–122.

VI. Instructional Practices and Mathematics Achievement: The Case of the Gifted Student

Students arrive at school with different skills and knowledge levels as well as capacities for benefiting from the opportunities provided by schools; these differences remain throughout schooling (Benbow & Stanley, 1996). This conclusion has been documented widely in the literature, going back as far as Learned and Wood (1938). Learned and Wood were among the first to show the wide range in knowledge among students in the same grade. For example, approximately 10% of high school seniors had more scientific knowledge than the average college senior. Such individual differences in knowledge and skills are evident even before entry into kindergarten, are reflected by the variance of test scores, and persist in every grade thereafter (Paterson, 1957; Pressley, 1949; Seashore, 1922; Terman, 1954; Tyler, 1965; Willerman, 1979; also see Learning Processes Task Group report). Moreover, there are differences in rate of learning. Those 13-year-olds who are in the top 1% of ability, for example, can assimilate, in three intensive weeks of schooling, a full year of high school biology, chemistry, Latin, physics, or mathematics (e.g., Lynch, 1992; Stanley & Stanley, 1986; VanTassel-Baska, 1983). Those who are in the top 1 in 10,000 in ability can accomplish even more in this time frame. Moreover, highly mathematically able students, with their exceptionally strong short-term working memory (Dark & Benbow, 1990, 1991, 1994), enjoy abstract, unstructured problems and thrive with complexity, which is different from the learning environment that is typical in the "regular" classroom. At the other end are students who need intensive work and much structured support and scaffolding over a long period of time to master basic skills in reading and mathematics. A challenge in teaching, then, is to be responsive to these individual differences so that all students make progress and are allowed to achieve their potential (National Research Council, 2000, 2002; Stanley, 2000). Particularly challenging for teachers are those students who are advanced or so challenged that the typical age-grade curriculum becomes inappropriate. In the case of the advanced student, serious adjustment is required if to teach them only what they already do not know (Stanley, 2000). In this report, the Task Group begins by briefly describing the strategies that are typically used to meet the learning needs of the advanced learner, often labeled the gifted student, and then move on to assess their effectiveness.

In American schools there are a plethora of programs that have been developed to meet the needs of gifted students. They represent the varied results obtained when the four, theoretically derived principles for adjusting the educational experiences or, more precisely, differentiating the curriculum are employed. The curriculum can be differentiated by level (e.g., grade level), complexity (e.g., abstract, unstructured), breadth or depth, and pacing to meet the learning needs of gifted students and ensure developmental appropriateness, according to an extensive literature in gifted education (Kaplan, 1986; Renzulli, 1986; VanTassel-Baska, 1998; Olszewski-Kubilus, 2007). Depending upon the relative emphasis of each one of these principles and the social context, the resulting programs fall into four broad categories: enrichment, acceleration, homogeneous grouping, and individualization. Enrichment often is seen in the regular classroom or in pullout programs or supplemental classes. It represents attempts to make the curriculum more appropriate for gifted students by adding to it or providing more depth and complexity while keeping students with their same-age peers. Acceleration and homogeneous grouping are attempts at forming groups for instruction that are at the same approximate achievement level, either by moving the advanced student to a higher grade in a (or many) subject(s) or by forming groups of same-age students on the basis of their demonstrated achievement. Indirectly, complexity is enhanced. Of course, all of these options can be used in some combination and that is what textbooks and articles in gifted education suggest (VanTassel-Baska, 1998). As well, the amount of adjustment required depends upon the level of giftedness and the difference between the individual gifted student and the average of the class. Acceleration that involves grade-skipping or putting individual students in a higher grade for a specific subject, for example, is typically reserved for the highly gifted (e.g., top 1% or even more extreme ability levels) as students much below that level often do not require such extreme adjustments.

A big debate in gifted education has been between the use of enrichment and acceleration. Most view this as a false dichotomy. For the highly gifted especially, it is recommended that both be utilized (VanTassel-Baska, 1998; Olszewski-Kubilus, 2007). There is, however, great resistance in K–12 schools toward using acceleration, even with the highly gifted (Benbow, 1991; Colangelo et al., 2004). This is not the case at the collegiate level when course placement is dependent upon having met prerequisites or scores on placement exams. The resistance by K–12 educators and fears of parents in terms of social and emotional development, however, have served to stimulate much research, admittedly of varying quality, to assess acceleration's effectiveness and whether it actually produces harm. Thus, there is an imbalance in the existing research literature in gifted education, with most of the research focused on accelerative strategies. Moreover, acceleration itself is a profound thing to do as it puts usual intellectual and social trajectories out of synchrony and often involves just one or possibly just a few students in a given school. This means that special considerations and individualization are required to make it possible and ensure its success. For example, special efforts are made to place the to-be-accelerated child with a teacher supportive of the acceleration if at all possible (VanTassel-Baska, 1998), given the frequent hostility toward such students and any interventions provided (Benbow & Stanley, 1996; Coleman, 1960; Cramond & Martin, 1987; Hofstadter, 1963; Tannenbaum, 1962). This, coupled with other issues (e.g., the child wanting to accelerate—motivation as a criterion)—makes it challenging to conduct carefully controlled research.

Previous meta-analyses have tried to make sense of this literature with all of its limitations and the varying quality of studies. They identified acceleration as the most promising strategy, followed by homogeneous grouping involving differentiation of the curriculum and adjustment of methods of teaching (Kulik & Kulik, 1982, 1984, 1992; Rogers, 2007; Olszewski-Kubilus, 2007). This represents what the field of gifted education thought the state of knowledge was before the Instructional Practices Task Group began its work.

As discussed in the introduction of this report and in Appendix A, the Instructional Practices Task Group developed criteria for which studies it would consult as part of its deliberations. The Task Group's charge was to assess the effects of instructional practices on mathematics achievement and establish a warranted claim of causality. It posed the following question: Can the Task Group conclude, without much doubt, that an intervention or mode of teaching is more effective than conventional practice or another approach? To draw such conclusions requires that studies meet a high standard of methodological rigor. Experimental and high-quality quasi-experimental studies would be consulted and confounds had to be carefully assessed to determine if valid inferences could be drawn even from this select group of studies. The Task Group also decided that non-Category 1 studies could be used only to support the conclusions of well-designed experimental or quasi-experimental studies. Groups of compromised studies (e.g., analyses, where weaknesses in one, for example, are off-set by findings in another) could provide context for the analysis conducted by the Task Group and the strength of its recommendation. The Panel also chose to limit itself for the most part to published, peer-reviewed journal articles. The approach is perhaps most similar to that used by the What Works Clearinghouse. This process eliminated all but a few relevant studies in several topical areas. This was true here as well—for the report on strategies used to serve gifted students.

Using the criteria established, the Task Group conducted a literature search for studies that assessed effectiveness of various options for serving gifted students. Key terms used included enrichment, differentiated curriculum, and acceleration. Only studies that compared gifted students participating in an intervention with a comparison group composed of nonparticipating gifted students were included. Studies that employed other comparison groups (e.g., students several grades above the treatment group, norms, or non-gifted) are not included. Finally, the Task Group generally used the term gifted to refer to students at the 90th percentile or above on standardized mathematics achievement tests, although most of the studies included here used much more selective criteria. The literature search, operating within these constraints, initially produced 11 studies, one of which was immediately eliminated due to methodological design weaknesses. The remaining ten were then reviewed by an independent methodologist, who also assessed them in relation to the Panel criteria. The Task Group followed his guidance, which resulted in two more studies being eliminated. Additional suggestions for studies to be consulted that emerged from the review process or in discussion were followed up and subjected to the same review criteria.

Of the eight studies that were included in this report on serving the needs of gifted children, all were either Category 1 or 2 studies (as described in the Methodology document in Appendix A). One was a randomized control trial (RCT) and seven were quasi-experimental. The methodological limitations of each study are clearly presented below. The Task Group organized the studies based on the type of approach toward instruction into two main categories: (i) Acceleration practices, including individualized, self-paced learning and (ii) Enrichment with or without acceleration.

In the following sections, the Task Group describes the practices used, provides study characteristics for each of the studies, and calculates effect sizes for the outcomes when possible.

A. The Role of Acceleration in Gifted Students' Math Achievement and Math-Related Outcomes

Acceleration of the curriculum, as noted above, is one form of adapting the instructional experiences received by gifted students. The curriculum is adjusted to meet the needs of the individual learner or, rather, the individual is placed in the curriculum at the approximate level of his or her functioning. Some call this placement according to competence or developmental placement (Benbow & Stanley, 1996). Acceleration may include presenting subject matter content earlier (e.g., algebra in 7th grade) or at a faster pace, or both, self-paced learning or compacting of the curriculum, participating in Advanced Placement programs (i.e., college-level classes in high school), taking college classes while in high school, skipping grades, and graduating early from high school and subsequently entering college early. It provides a differentiated curriculum for gifted students by using curricula designed for older students. The opinion of most educators in the field of gifted education, however, is that good acceleration does not stop there (VanTassel-Baska, 1998). It also should explore topics more deeply, probe interconnectedness of concepts, and adjust the content to make it more complex and abstract. This can occur in special accelerated classes for gifted students or in the regular classroom with a truly excellent teacher.

Several points need to be considered when evaluating the value of acceleration for gifted students. Acceleration, beyond self-paced learning or offering algebra to eighth-graders, is reserved for the highly gifted. Second, because of social and academic disruptions it causes, acceleration is used only with students who want to accelerate (VanTassel-Baska, 1998). No matter how positive the effects of acceleration could be, it is a widely held professional opinion that it is inadvisable to accelerate a child if there is significant resistance (Benbow, 1998). Thus, this educational intervention is different from others (e.g., choosing a specific text-book or teaching method) because student choice is a factor in its use. It may be that those students who choose to accelerate are more academically motivated or desire academic challenges more than those who choose not to. Alternatively, those who choose to not accelerate may need more of other, nonacademic, factors in structuring a satisfying life. That is, accelerates and non-accelerates may have different priorities and this is confounding when the aim is to assess effects of acceleration specifically. Thus, any recommendations would pertain only to academically motivated students, the very ones for whom acceleration is to be used according to practice guidelines developed on the basis of professional judgment.

Third, when gifted students are accelerated by putting them together for special classes, this creates a different academic and social environment that appears to be highly valued by and motivating for gifted students (Benbow, Lubinski, & Suchy, 1996). In this descriptive study, students report feeling affirmed and challenged in ways that the regular classroom does not provide. Also, the nature of the discourse changes, becoming much more high-level and intellectually challenging (Fuchs, Fuchs, Hamlett, & Karns, 1998). So, accelerated classes are more than just content taught at a fast pace. This makes it hard, if not impossible, to separate out the effects attributable only to the acceleration in these types of programs.

The analyses presented in Table 30 include six quasi-experimental studies that looked at acceleration and include both short- and long-term outcomes; that is, students were assessed shortly after having been accelerated (e.g., completion of self-paced learning program) or several years later. In the latter studies, the short-term impact on learning is not assessed (e.g., covering two years of mathematics in one) but rather subsequent participation in mathematics in the years following the advancement (e.g., college course-taking four years later) is assessed. In the analyses, the Task Group presents individual but not pooled effects sizes of the acceleration practices. The "interventions" were seen as sufficiently different to preclude pooling. Findings associated with both the short-term and long-term outcomes are presented in Table 30.

Table 30: Studies That Examine the Impact of Acceleration on Gifted Students' Math Achievement and Math Related Outcomes

Study	Design	Sample	Duration/ Content	Contrast	Measure	Hedge's g	Standard Error	
Short-Term Effects								
Brody & Benbow, 1990: Study 1[c]	Quasi	80 seventh-grade participants in the Talent Search sponsored by the Center for Talented Youth (CTY) at Johns Hopkins University. All subjects were screened with SAT-M to meet eligibility requirements for participation in gifted educational programs	Three weeks/ Algebra 1 or above	Fast-paced accelerated summer math class vs. No summer program	SAT-Math	0.241	(ns)	0.227
Ma, 2005[b d e]	Quasi	276 gifted seventh-grade students randomly selected from the Longitudinal Study of American Youth (LSAY)	One school year/ Algebra 1	Took Algebra 1 in grades 7 or 8 vs. Did not take Algebra 1 in grades 7 or 8	LSAY Math Achievement (combination of sub-tests: math basic skills, algebra, geometry)[f]	0.167	(ns)	0.166
Parke, 1983	Quasi	44 gifted students in Grades K–2 from two elementary schools	10 weeks/ Addition, subtraction, place value, sets, and measurement	Used self-instructional math materials vs. Control	Skill Mastery[g]	Insufficient data to calculate effect sizes		
Ysseldyke et al., 2004[a]	Quasi	100 gifted students in Grades 3–5 that were part of a larger study in which instructional software was implemented in 15 different states in the U.S.	Four months/ Individually based mathematics curriculum tailored to grade level/skill level	Personalized Computer Instruction vs. Control	STAR math computer adaptive test of mathematics skills in numeric concepts, computation, and math applications	0.449	~	0.271

Continued on p. 6-160

Table 30, continued

Study	Design	Sample	Duration/ Content	Contrast	Measure	Hedge's g	Standard Error	
Long-Term Effects								
Swiatek & Benbow, 1991[a]	Quasi	37 qualifying participants and 58 nonparticipants of a fast-paced/self-paced accelerated math classes that were followed up 10 year after participation. The subjects were initially identified through the Study of Mathematically Precocious Youth (SMPY) referrals and screened through additional testing	Saturday mornings for approximately one year/ Algebra 1 and 2, plane geometry, trigonometry, and analytic geometry	Participation in extracurricular fast-paced and self-paced accelerated math classes vs. No participation	Percent taking elective undergraduate math courses (ES = 0.066), Percent undergraduate majors in math (ES = 0.271), Percent graduate majors in applied math (ES = 0.514~)	0.284	(ns)	0.317
Swiatek & Benbow, 1991[b]	Quasi	107 pairs of gifted students that were followed up 10 years after participation in SMPY	N/A	Students who chose to undergo acceleration and enter college at least one year early vs. Students who chose traditional educational route	Number of non-required math courses (ES = 0.381**), Number of undergraduate math courses (ES = 0.091)	0.236	~	0.137

~ p < .10, * p < .05, ** p < .01, *** p < .001
[a] Data were adjusted for clustering that occurred within classrooms.
[b] Data were adjusted for clustering that occurred within schools.
[c] The standard deviations used for the calculation of effect size were estimates provided by the author.
[d] The number of students used for the gifted analytic sample was adjusted by the 12% attrition that the overall study had in the first two years.
[e] The seventh and ninth grade achievement tests were used as pre- and posttests to estimate effect of taking Algebra 1 in seventh or eighth grade.
[f] The three sub-measures of LSAY Math Achievement Test—math basic skills, algebra, geometry—were combined to create a score comparable to the SAT-M.
[g] The assessment included 170 items testing mastery of 82 skills in five areas: addition, subtraction, place value, sets, and measurement.

Brody and Benbow (1990; Study 1), in a quasi-experimental study, investigated whether short-term, accelerative academic training had an effect on SAT scores of middle school students who were in the top 1% in ability. Program participants were enrolled in a fast-paced, three-week summer academic program that was focused on increasing content knowledge in pre-algebra. Their performance on the SAT was compared to a nonrandomized control group not participating in any accelerative learning experience during the summer but in the top 1% in ability (with lower SAT-M scores initially but not SAT-V) or to students enrolled in other, nonmathematics accelerative classes in the academic summer program. The results from the ANCOVA, adjusting for relatively large initial differences in ability across groups, revealed that in-depth instruction over a short period of time in specific mathematical or verbal areas had little or no impact on SAT scores at the conclusion of the program. The data reported in Table 30 reflects no significant effect between the participants in the fast-paced mathematics class and those not enrolling in any special mathematics class (ES = 0.241).

Ysseldyke et al. (2004), in a quasi-experimental study, compared third through sixth grade gifted students whose mathematics curriculum was differentiated and adjusted to the needs of the students through a instructionally-based curriculum management system, called Accelerated Math (Renaissance Learning, 1998). It was a self-directed, four-month long mathematics program with assessment of skill level, tailoring of the instruction to match skill level, individual pacing and goal setting, ample practice, and immediate feedback to student and teacher on performance. The effect size of 0.449 bordered on statistical significance, and favored the personalized instruction group on outcome measures of mathematics skills, numeric concepts, computation, and mathematics applications.

Parke (1983), in another quasi-experimental study, used a 10-week self-instruction program in mathematics with gifted Kindergarten through second-graders to differentiate the curriculum and compared their performance to another equally able, high-ability sample and a comparison group, both of whom were enrolled in a regular class with no differentiation. In terms of design limitations, there was no random assignment; the sample size was small; and there were large pretest ability differences among the groups. Data needed to compute an effect size also were not reported. The ANCOVA results reported by the author, adjusted for initial differences in ability, were statistically significant. The intervention group mastered significantly more concepts and skills than the comparison groups. The adjusted means were 52 learned concepts for the participants, 38 for the high-ability comparison group, and 41 for the random comparison groups. This finding lends support to the value of differentiation through individualization via self-paced, accelerative learning.

Ma (2005) compared the mathematics achievement at the end of high school for students in the top 10% in ability who took formal algebra either in seventh or eighth grade, an increasing trend in the U.S. for capable students, and equally able students who took such algebra in ninth grade or beyond. Ma used a subsample of the Longitudinal Study of American Youth that was divided into gifted, honors, and regular students. For each, differences between students who took Algebra I early (accelerated) versus those who did not, mainly reflecting practices in different schools, were examined. There were relatively balanced numbers of accelerated versus not accelerated for the gifted (49%) and honors (21%) students, but few "regular" students were accelerated in this way (4%) as would be expected. The mathematics achievement outcome variable, which captured performance on a combination of basic mathematical skills, algebra, geometry, and quantitative literacy items, was a growth curve of achievement measures from Grade 7 to 12. The effect size for the score differences favoring accelerates was 0.167 and not statistically significant. All accelerated students seemed to perform well on this test, however, even though reservation has been expressed about learning algebra early (e.g., Prevost, 1985).

Swiatek and Benbow (1991a), in a quasi-experimental study, assessed participants 10 years after the completion of two homogeneously grouped and fast-paced mathematics classes. The individuals in these classes had learned algebra and possibly all the content up through precalculus at a rapid rate. These classes were the model for the fast-paced programs that have sprung up across the country in the past 35 years and now serve over 100,000 gifted students annually. The initial class was taught by an experienced math teacher at a rate dictated by the capacity for learning of the most able students in the class. Most students in the class completed four years of mathematics in 14 months and their standardized achievement test scores were well above the 90th percentile on relevant tests of mastery. A subgroup completed just two years of math in that time frame. They were less able initially and, thus, experienced difficulty in keeping up the pace of the faster moving group. The participants in the two fast-paced mathematics classes were compared to students who had been matched on ability but did not attend the class and students who dropped out of the class. All were at least in the top 1% in ability, but there may have been motivational and other differences between participants and nonparticipants. Another limitation was that the same teacher taught both classes (reassuringly, similar results in mathematics has been found with other teachers, e.g., Lunny, 1983; Mezynski, Stanley, & McCoart, 1983). At the end of high school, the participants scored higher on standardized mathematics achievement tests, such as the College Board Math Achievement test, than the nonparticipants or dropouts, despite their younger age, and did not regret their acceleration (Benbow, Perkins, & Stanley, 1983). Ten years after the initiation of the class few statistically significant differences on academic achievement variables emerged between the participants and the comparison group (Table 30). The one effect that borders on significance (ES = 0.514) favored the participants (who also tended to be several years ahead in their educational progress and so were younger at time of comparison on specific variables than nonparticipants). This comparison was the percent of students at age 23 who were attending graduate school in applied mathematics, engineering, and computer science (50% of participants vs. 28% of comparison group).

Swiatek and Benbow (1991b), in a quasi-experimental study, compared, via a 10-year follow-up, mathematically talented students (at least top 1% in ability) who had managed to accelerate their education so that they entered college at least one year early with equally able students who had not entered college early. This was a nonrandomized comparison, but the groups had been matched on gender and pretest SAT scores (within 10 points for mathematics, 20 for verbal). The mathematics achievement outcomes were indirect—number of undergraduate mathematics courses taken, number of non-required mathematics courses taken, mathematics major as an undergraduate or graduate student, and interest, confidence, and perceived ease of mathematics. Only one statistically significant effect size was found on the various outcome variables (see Table 30)—on the number of non-required mathematics courses taken (ES = 0.381). The difference favored accelerates who, of course, also had the advantage of being advanced in their education.

In a correlational study, Sadler and Tai (2007) have demonstrated that learning mathematics at an earlier age than typical or at a faster pace is related to allowing students to become more advanced in their mathematics education and to be better prepared for college science classes. No long-term negative consequences have been found and the evidence suggests that there are possibly some small additional advantages.

B. The Role of Enrichment on Gifted Students' Mathematics Achievement

The following section presents findings from the remaining two studies that utilized primarily enrichment to differentiate the curriculum for gifted students (Robinson et al., 1990; Robinson, 1997). Because the interventions differed in that one also explicitly adjusted the pace of instruction, the effects are presented individually (see Table 31).

.

Table 31: Studies That Examine the Role of Computer Instruction, Enrichment, and Cooperative Learning on Gifted Students' Math Achievement

Study	Design	Sample	Duration/ Content	Contrast	Measure	Hedge's g	Standard Error	
Robinson et al., 1990	Quasi	78 elementary age students who participated in a special program for mathematically talented children and 185 program alternates	One school year/ Variety of math content	Curriculum replacement program with a focus on enrichment, self-paced computer instruction, and acceleration vs. Regular curriculum	Math Applied to Novel Situations (MANS)[c]	0.648	***	0.114
Robinson et al., 1997	RCT	310 gifted Kindergarten and 1st-grade students in 158 schools who scored at or above the 98th percentile on a screening test	14 2 1/2 hour sessions per year for two years/ Kindergarten and first-grade curriculum	Extracurricular constructivist enrichment activities (Saturday Club) vs. No enrichment	Pooled measures (Stanford-Binet IV quantitative subtest, Number knowledge test; Woodcock-Johnson calculation subtest)	0.401	**	0.127

~ p < .10, * p < .05, ** p < .01, *** p < .001
[a] Data were adjusted for clustering that occurred within classrooms.
[b] Data were adjusted for clustering that occurred within schools.
[c] Standardized measure MANS measures skills in computation, estimation, mental arithmetic, number representations, relations, number patterns, elucidation, word problems.

Robinson et al. (1997), the only experimental study to emerge out of the literature search, randomly assigned equally able gifted kindergarten and first-grade students to supplemental enrichment mathematics classes conducted on Saturdays over two years or to no treatment. The enrichment classes, with 28 sessions in all, were described by the authors as constructivist in philosophy, "developmentally appropriate," and adhering to NCTM (1989) guidelines. Teachers created social communities that engaged in open-ended problem-solving. At the end of two years, the participants significantly outperformed nonparticipants on a combined mathematics achievement measure. However, there was differential attrition—5% in the control condition, 20% in treatment condition—and it is possible that the least able students left the program at higher rates than the most able. A statistically significant effect size of 0.401 was found favoring the students who participated in the enrichment program. Here, the regular curriculum in the school was not differentiated in any way. Rather, gifted children were challenged through the provision of extra activities, a pull-out model of sorts, and were not explicitly accelerated.

The Robinson et al. (1990) study is similar to the Ysseldyke et al. (2004) study in that it utilized CAI to adjust pace and is quasi-experimental. However, this after-school mathematics program for gifted elementary students also provided specific enrichment activities for the class that allowed students to add breadth and depth to their learning. Hence, the curriculum was differentiated even further than what was possible in Ysseldyke et al. and Parke (1983). This was, however, one single class and hence involved just one teacher. Performance of participants was compared to non-randomized control groups comprised of students who were selected but did not attend or were selected as alternates. Although there was no reporting on pretest mean differences in ability among the groups, a regression discontinuity analysis was used with pretest proxy measures as covariates. With a statistically significant effect size of 0.648, the results lend support to the value of differentiating and enhancing the pace of the curriculum for gifted students.

C. Conclusions

It is generally agreed that good teaching is responsive to individual differences, tailoring instruction to meet the needs of individual learners (Robinson, 1983). In the case of gifted students who are advanced in their skill and concept attainment and can learn new material at a much more rapid rate than their same-age peers (e.g., Lynch, 1992; Stanley & Stanley, 1986; VanTassel-Baska, 1983), it is the professional opinion of those in gifted education that these students need a curriculum that is differentiated (by level, complexity, breadth and depth), developmentally appropriate, and conducted at a more rapid rate (Van Tassel-Baska, 1998). This is typically accomplished to some degree through some combination of acceleration, homogeneous grouping, enrichment, or individualization.

As the Instructional Practices Task Group began its work, it was aware that there were hundreds of studies over decades evaluating the effectiveness of acceleration, in which results have been interpreted as indicating positive academic benefits and no negative effects social-emotionally (see Colangelo, Assouline, & Gross, 2004; and Rogers, 2007 for the latest syntheses). The Task Group did not know the overall quality of these studies or their usefulness for drawing causal attributions. So, it was impossible to decipher the strengths of the signal emitted from these studies and into which category the support for this instructional practice fell (see the Panel Standards of Evidence Document). From a descriptive study, the Task Group learned, however, that gifted students report satisfaction with acceleration (even wishing as adults that they had accelerated more) and that they feel they would not have achieved as much without it (Benbow et al., 2000; Benbow, Lubinski, & Suchy, 1996). But, such data, although valuable, are from the world of perceptions and beliefs and cannot speak to effectiveness.

Enrichment, which attempts to add breadth and depth to the regular curriculum, as well as complexity, also has been studied and has exhibited some positive effects under the same circumstances, limitations, or conditions affecting the interpretability of findings from the literature on acceleration. Yet, many seemingly excellent enrichment programs have not been rigorously evaluated, perhaps because this option for meeting the needs of gifted students has faced less negativity and resistance than is the case for acceleration.

Homogeneous grouping is an educational approach that meets with much controversy as well. Enrichment tends to dominate in homogeneously grouped classes, but it often includes some increased pace of learning. So, there can be settings wherein both acceleration and enrichment is utilized, which most professionals in gifted education would prefer. Before the Task Group began our review of this literature, the state of knowledge, as captured by the results of meta-analyses, revealed positive effects of homogeneously grouped classes, with the value-added gain in one year being about four to five months (Kulik & Kulik, 1982, 1987, 1992). These meta-analyses, however, lumped together studies of various methodological quality, making them less rigorous tests of effectiveness and hence compromised generalizability (also see Delcourt, Cornell, & Goldberg, 2007; Robers, 2007). In terms of perceptions, nonetheless, gifted students when reflecting back as an adult 20 years later seem to favor homogeneous grouping (Benbow et al., 2000). Finally, although often utilized to stimulate gifted children, the effects of mathematics contests on the mathematics achievement of gifted students have not been well studied. This, then, was the state of knowledge before Instructional Practices undertook its analysis.

The Task Group's review of the literature assessing the effectiveness of the various means for tailoring instruction to meet the needs of gifted students yielded surprisingly few studies that met the methodologically rigorous criteria for inclusion adopted by the Task Group. The Task Group actually had to use somewhat less stringent criteria than in other instructional practices reports in order to fulfill the charge of evaluating the "best available scientific evidence." The Task Group could formulate recommendations only on the basis of one randomized control trial study and seven quasi-experimental studies that met the Category 1 and 2 criteria. This was disappointing, especially because even the few studies included in the analyses contained some methodological limitations. For example, almost all studies on acceleration, although essentially positive in their reported outcomes (Colangelo, Assouline, & Gross, 2004 and Rogers, 2007 provide a comprehensive review), are limited to students who are highly gifted and motivated to accelerate. Thus, motivation is a confound just as it is a selection criterion for being considered a candidate for acceleration.

Nonetheless, the studies reviewed above that met our criteria provided some support for the value of differentiating the mathematics curriculum, especially when acceleration is a component (i.e., pace and level of instruction is adjusted). Individualized instruction in which pace of learning is increased, often managed via computer instruction, also showed positive benefits.

The challenge of implementing random assignment or well-matched comparison groups in programs for gifted students is substantial. Parents are unlikely to agree to let their child participate in anything but the treatment that is designed for acceleration or enrichment. Thus, to gain insights about the impact and nature of different approaches to the mathematical education of the gifted, it would be useful to look at some research that does not meet the inclusion criteria because it has been designed to be more descriptive or correlational. That research when coupled with the analyses reported above suggests several positive directions. For instance, there is evidence that gifted students who are accelerated by other means gained time and reached educational milestones earlier (e.g., college entrance) than their equally able same-age peers. They also demonstrate comparable or stronger performance than their same-age peers (although with small effect sizes) on a variety of indicators, at younger ages. Together these studies help to illuminate the conclusions drawn from the scientific literature as summarized above. The Task Group has no evidence that acceleration harms the mathematical achievement of the gifted student.

Gifted students who are accelerated also appear to become more strongly engaged in science, technology, engineering, or mathematics areas. This finding fits well with the results of a recent correlational study showing that the more mathematics courses taken in high school, which is facilitated through acceleration, the more likely students are to perform well in science (Sadler & Tai, 2007). Although some have seen acceleration as a cause for concern, there is no evidence in the studies that met our criteria that gaps and holes in knowledge have occurred as a result of acceleration.

Support also was found for supplemental enrichment programs. Of the programs analyzed here, one explicitly utilized acceleration as a program component and the other did not. Of the two studies that met our criteria for inclusion as Category 1 or 2, both studies had significant effect sizes favoring the enrichment treatment. So, although the evidence is somewhat mixed, it suggests a positive effect of enrichment approaches. Other research (e.g., Benbow, 1998; Delcourt, Cornell, & Goldberg, 2007; Rogers, 2007; VanTassel-Baska, 1998; VanTassel-Baska & Brown, 2007) has examined varied enrichment approaches. Clearly understanding the nature of the enrichment activity is crucial to efforts to improve opportunities for gifted students. Self-paced instruction supplemented with enrichment seemed to have a positive impact on student achievement. This supports the widely held view in the field of gifted education that acceleration and enrichment combined should be the intervention of choice.

Underscored by the analysis undertaken by the Task Group is the need for more high-quality experimental and quasi-experimental research to study effectiveness of interventions designed to meet the learning needs of gifted students. Especially missing are evaluations of academically rigorous enrichment programs, the mathematical content explored in such programs, and their goals. The Task Group concludes, however, that it is important for school policies to support appropriately challenging work in mathematics for gifted and talented students. Acceleration, combined with enrichment, is certainly a promising, possibly moderately supported (if the entire literature is considered), practice.

BIBLIOGRAPHY

Benbow, C.P. (1991). Meeting the needs of gifted students through use of acceleration: A neglected resource. In M.C. Wang, M.C. Reynolds, & H.J. Walberg (Eds.), *Handbook of special education* (Vol. 4, pp. 23–36). Elmsford, NY: Pergamon Press.

Benbow, C.P. (1998). Grouping intellectually advanced students for instruction. In J. VanTassel-Baska (Ed.), *Excellence in educating the gifted* (pp. 261–278). Denver, CO: Love.

Benbow, C.P., Lubinski, D., Shea, D.L., & Eftekhari-Sanjani, H. (2000). Sex differences in mathematical reasoning ability: Their status 20 years later. *Psychological Science, 11(6),* 474–480.

Benbow, C.P., Lubinski, D., & Suchy, B. (1996). Impact of the SMPY model and programs from the perspective of the participant. In C. P. Benbow & D. Lubinski (Eds.), *Intellectual talent: Psychometric and social issues* (pp. 266–300). Baltimore, MD: Johns Hopkins University Press.

Benbow, C.P., Perkins, S., & Stanley, J.C. (1983). Mathematics taught at a fast pace: A longitudinal evaluation of SMPY's first class. In C. P. Benbow & J.C. Stanley (Eds.), *Academic precocity: Aspects of its development* (pp. 51–78). Baltimore, MD: Johns Hopkins University Press.

Benbow, C.P., & Stanley, J.C. (1996). Inequity in equity: How current educational equity policies place able students at risk. *Psychology, Public Policy, and Law, 2,* 249–293.

Brody, L.E., & Benbow, C.P. (1990). Effects of high school course-work and time on SAT scores. *Journal of Educational Psychology, 82(4),* 866–875.

Colangelo, N., Assouline, S.G., & Gross, M.U.M. (Eds.) (2004). *A nation deceived: How schools hold back America's brightest students*. Iowa City, IA: University of Iowa.

Coleman, J. S. (1960). The adolescent sub-culture and academic achievement. *America Journal of Sociology*, *65(4)*, 337–347.

Cramond, B., & Martin, C.E. (1987). In-service and preservice teachers' attitudes towards the academically brilliant. *Gifted Child Quarterly, 31(1)*, 15–19.

Dark, V.J., & Benbow, C. (1990). Mathematically talented students show enhanced problem translation and enhanced short-term memory for digit and spatial information. *Journal of Educational Psychology, 82,* 420–429.

Dark, V.J., & Benbow, C.P. (1991). Differential enhancement of working memory with mathematical and vertical precocity. *Journal of Educational Psychology, 83(1),* 48–60.

Dark, V.J., & Benbow, C.P. (1994). Type of stimulus mediates the relationship between performance and type of precocity. *Intelligence, 19(3),* 337–357.

Delcourt, M.A.B., Cornell, D.G., & Goldberg, M.D. (2007). Cognitive and affective learning outcomes of gifted elementary school students. *Gifted Child Quarterly, 51(4),* 359–381.

Fuchs, L.S., Fuchs, D., Hamlett, C.L., & Karns, K. (1998). High-achieving students' interactions and performance on complex mathematical tasks as a function of homogeneous and heterogeneous pairings. *American Educational Research Journal, 35(2),* 227–268.

Hofstadter, R. (1963). *Anti-intellectualism in American life*. New York: Knopf.

Kaplan, S. (1986). The grid: A model to construct differentiated curriculum for the gifted. In J.S. Renzulli (Ed.), *Systems and models for developing programs for the gifted and talented* (pp. 180–193). Mansfield Center, CT: Creative Learning Press.

Kulik, C.C., & Kulik, J.A. (1982). Effects of ability grouping on secondary school students: A meta-analysis of evaluation findings. *American Educational Research Journal, 19(3)*, 415–428.

Kulik, J.A., & Kulik, C.C. (1984). Effects of accelerated instruction on students. *Review of Educational Research, 54(3)*, 409–425.

Kulik, J.A., & Kulik, C.C. (1987). Effects of ability grouping on school achievement. *Equity and Excellence, 23,* 22–30.

Kulik, J.A., & Kulik, C.C. (1992). Meta-analytic findings on grouping programs. *Gifted Child Quarterly, 36(2),* 73–77.

Learned, W.S., & Wood, B.D. (1938*). The student and his knowledge. A report to the Carnegie foundation on the results of the high school and college examinations of 1928, 1930, and 1932*. New York: Carnegie Foundation for the Advancement of Teaching.

Lunny, J.F. (1983). Fast-paced mathematics classes for a rural county. In C.P. Benbow & J.C. Stanley (Eds.), *Academic precocity: Aspects of its development* (pp. 79–85). Baltimore, MD: Johns Hopkins University Press.

Lynch, S.J. (1992). Fast-paced high school science for the academically talented. A six-year perspective. *Gifted Child Quarterly, 36(3)*, 147–154.

Ma, X. (2005). A longitudinal assessment of early acceleration of students in mathematics on growth in mathematics achievement. *Developmental Review, 25(1),* 104–131.

Mezynski, K., Stanley, J.C., & McCoart, R.F. (1983). Helping youth score well on AP examinations. In C.P. Benbow & J.C. Stanley (Eds.), *Academic precocity: Aspects of its development* (pp. 86–112). Baltimore, MD: Johns Hopkins University Press.

National Council of Teachers of Mathematics (NCTM) (1989). *Curriculum and evaluation standards for school mathematics.* Reston, VA: National Council of Teachers of Mathematics.

National Research Council (2000). *How people learn: Brain, mind, experience, and school.* Washington, DC: National Academy Press.

National Research Council (2002). *Learning and Understanding*. Washington, DC: National Academy Press.

Olszewski-Kubilus, P. (2007). Best Practices in Gifted Education (Special Issue). *Gifted Child Quarterly, 51*(4).

Parke, B.N. (1983). Use of self-instructional materials with gifted primary-aged students. *Gifted Child Quarterly, 27(1),* 29–34.

Paterson. D.G. (1957). Conservation of human talent. *American Psychologist*, *12*, 134–144.

Pressley, S.L. (1949). *Educational acceleration: Appraisals and basic problems.* Columbus, OH: Ohio State University.

Prevost, F.J. (1985). Soundoff: Let's not teach algebra to eighth graders! *Mathematics Teacher, 78(8),* 586–587.

Renaissance Learning (1998). *Accelerated math [computer software]*. Wisconsin Rapids, WI: Author.

Renzulli, J.S. (1986). The three-ring conception of giftedness: A developmental model for creative productivity. In R.J. Sternberg & J.E. Davidson (Eds.), *Conceptions of giftedness* (pp. 53–92). New York: Cambridge University Press.

Robinson, A., Bradley, R.H., & Stanley, T.D. (1990). Opportunity to Achieve: Identifying Mathematically Gifted Students. *Contemporary Educational Psychology, 15(1),* 1–12.

Robins, H. (1983). A case for radical acceleration. In J.C. Stanley & C.P. Benbow (Eds.), *Academic Precocity: Aspects of its Development.* Baltimore, MD: Johns Hopkins University Press.

Robinson, N.M., Abbott, R.D., Berninger, V.W., Busse, J., & Mukhopadhyay, S. (1997). Developmental changes in mathematically precocious young children: Longitudinal and gender effects. *Gifted Child Quarterly, 41(4),* 145–158.

Rogers, K.B. (2007). Lessons learned about educating the gifted and talented: A synthesis of research on educational practice. *Gifted Child Quarterly, 51(4),* 382–396.

Sadler, P.M., & Tai, R.H. (2007). The two high-school pillars supporting college science. *Science, 317,* 457–458.

Seashore, C.E. (1922). The gifted student and research. *Science, 56(1458),* 641–648.

Stanley, J.C. (2000). Helping students learn only what they don't already know. *Psychology, Public Policy, and Law, 6(1),* 216–222.

Stanley, J.C., & Stanley, B.S.K. (1986). High school biology, chemistry, and physics learned well in three weeks. *Journal of Research in Science Teaching, 23(3),* 237–250.

Swiatek, M.A., & Benbow, C.P. (1991a). A 10-Year longitudinal follow-up of participants in a fast-paced mathematics course. *Journal for Research in Mathematics Education, 22(2)*, 138–150.

Swiatek, M.A., & Benbow, C.P. (1991b). Ten-year longitudinal follow-up of ability matched accelerated and unaccelerated gifted students. *Journal of Educational Psychology, 83(4),* 528–538.

Tannenbaum, A.J. (1962). *Adolescent attitudes towards academic brilliance*. New York: Bureau of Publications, Teachers College, Columbia University.

Terman, L.M. (1954). The discovery and encouragement of exceptional talent. *American Psychologist, 9,* 221–230.

Tyler, L.E. (1965). *The psychology of individual differences* (3rd Edition). New York: Meredith.

Van Tassel-Baska, J. (1983). Illinois' state-wide replication of the Johns Hopkins Study of Mathematically Precocious Youth. In C.P. Benbow & J.C. Stanley (Eds.), *Academic Precocity: Aspects of its Development* (pp. 179–191). Baltimore: Johns Hopkins University Press.

Van Tassel-Baska, J. (1998). *Excellence in Educating Gifted & Talented Learners.* Denver, CO: Love.

Van Tassel-Baska, J., & Brown, E.F. (2007). Toward best practice: An analysis of the efficacy of curriculum models in gifted education. *Gifted Child Quarterly, 51(4)*, 342–358.

Willerman, L. (1979). *The psychology of individual and group differences*. San Francisco: Freedman.

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

VII. Teachers' Use of Formative Assessments to Improve Learning of Mathematics: Results from a Meta-Analysis of Rigorous Experimental and Quasi-Experimental Research

Formative assessment—ongoing monitoring of student learning to inform instruction—is generally considered a hallmark of effective instruction in any discipline. In the past 20 years, the term has been used in two complementary but distinct traditions of scholarship and research. One tradition—which has played an influential role in the field of mathematics education—is represented, for example, in two recent publications by the National Research Council—one on the teaching and learning of mathematics (National Research Council, 2001) and one on human learning and cognition (National Research Council, 2005). Donovan and Bransford (National Research Council, 2005) established three goals for formative assessments: 1) "to make students' thinking visible to both teachers and students;" 2) "to monitor student progress (in mastering concepts as well as factual information);" and, 3) to "design instruction that is responsive to student progress" (p. 16).

The second tradition, developed primarily within the fields of school psychology, educational psychology, and special education shares only one of the three aforementioned goals. This approach is typified, for example, by the definition provided in the recent *Encyclopedia of School Psychology* (Lee, 2005). In this tradition, formative assessments are tools "used to monitor progress and to provide feedback about the progress being made. ... In the classroom setting, formative evaluation is used to inform students and teachers about progress" (p. 209). The goal of this type of formative assessment is to determine whether specific students—or in some cases, an entire class—require additional instruction devoted to learning a particular concept or acquiring proficiency in a particular mathematical procedure or strategy for problem solving. Typically, measures are administered weekly or biweekly, often using computer-assisted assessment. They are brief and efficient, taking approximately 5 minutes to administer.

Note that both research traditions stress monitoring progress toward an instructional goal and adjusting instruction for students based on the formative measures. However, the school psychology tradition devotes a good deal of attention to empirical establishment of the validity and reliability of the assessment procedures (e.g. Fuchs, 2004; Foegen, Jiban, & Deno, 2007).

The system described by Donovan and Bransford (2005) rarely addresses psychometric issues. It presents an ambitious agenda that includes not only progress monitoring but also interpretation of students' errors and misconceptions to guide the types of questions teachers ask to probe for student understanding. This tradition is embodied for example, in *Adding It Up (*National Research Council, 2001):

> *Information about students is crucial to a teacher's ability to calibrate tasks and lessons to students' current understanding.… In addition to tasks that reveal what students know and can do, the quality of instruction depends on how teachers interpret and use that information. Teachers' understanding of their students' work and the progress they are making relies on … their ability to use that understanding to make sense of what the students are doing. (pp. 349–350).*

Ruiz-Primo, Shavelson, Hamilton, and Klein (2002) discuss a continuum of assessment distance that traverses both traditions. The continuum includes the following distance classifications, with the classifications ranging from formative to summative assessments, by proximity:

- **Immediate—**informal observation or artifacts from a lesson.
- **Close—**embedded assessments and semi-formal quizzes following several activities.
- **Proximal—**formal classroom exams following a particular curriculum.
- **Distal—**criterion-referenced achievement tests such as those required by NCLB.
- **Remote—**broad outcomes measured over time—norm-referenced tests, such as the Scholastic Aptitude Test.

Formative assessment would be identified as immediate, close and perhaps proximal in the continuum above and is used to regularly monitor instruction. Freudenthal (1973) noted, “It is more informative to observe a student during a mathematical activity than to grade his papers” (p. 84). Informal assessments which include observations and informal probes of students to assess their level of understanding, according to Freudenthal, need to inform day-to-day teaching.

Sueltz, Boynton, and Sauble (1946) noted that observation, discussion, and interviews serve better than paper-pencil tests in evaluating a pupil’s ability to understand the principles he or she uses. Spitzer (1951) and others have long advocated the interview as a formative assessment strategy that is closely associated with the use of observations. Decades later the National Council of Teachers of Mathematics noted that information is best collected through informal observation as students participate in class discussions, attempt to solve problems, and work on various assignments individually or in groups (National Council of Teachers of Mathematics, 1989, p. 233). However, Glaser and Silver (1994) note that aside from teacher-made classroom tests, the integration of assessment and learning as an interacting system has been too little explored. The meta-analysis completed by the Instructional Practices Task Group revealed no methodologically acceptable studies that examined the impact of using this type of assessment on student performance.

The goal of this section is to review the experimental and quasi-experimental research on the extent to which teachers’ use of formative assessments in mathematics enhances students’ acquisition of mathematics content. Although the Task Group reviewed the literature for studies using any type of formative assessment from any tradition, the only studies located that met the criteria for adequate experimental design emanated from the school psychology or educational psychology traditions.

This report describes studies that indicate the extent to which use of formative assessments improves students’ mathematics proficiency. The Task Group also describes the impacts of various enhancements, i.e., procedures and strategies for helping teachers use this information to provide differentiated instruction, and, thus enhance mathematics achievement. The Task Group centered the meta-analysis on two research questions. The first was whether teachers’ use of formative assessments enhanced student achievement in mathematics. The second question explored the effectiveness of various tools or enhancements that can assist teachers in their use of formative assessments. Before discussing the findings of the meta-analysis, the Task Group begins by providing a brief historical overview.

A. Historical Overview

In the 1970s and 1980s, a good deal of research and effort went into the field of formative assessment. Researchers examining high-performing schools invariably found that the school used some system to monitor all students' academic progress on a regular basis. Mastery learning (e.g., Bloom, 1980; Guskey, 1984), a form of differentiated instruction was widely implemented in large school districts such as Chicago and San Diego. Mastery learning calls for frequent assessment of student progress using brief tests at the end of each unit, which is typically once a week. Students who did not reach a mastery level (typically defined as 80% correct) are retaught the material. Those with scores above the mastery level are provided with extension or enrichment activities. During this time in history, publishers of the major mathematics curricula began to include unit tests along with their programs. This practice continues to this day.

B. Validity and Reliability Concerns for Formative Assessments in Mathematics

In developing formative assessments in mathematics, the goal has invariably been to develop measures that are valid and reliable in the psychometric sense (AERA, APA, NCME, 1999) and that are relatively easy to administer and score.

Contemporary conceptions of test validity include indices that a measure is correlated with other measures of mathematics achievement which can include teacher appraisals (*criterion-related validity*), that the mathematical content is valid and important (*content validity*) and that there is evidence concerning the impact of use of the measure, including both intended and unintended consequences (*consequential validity*) (e.g., Messick, 1988). A valid formative assessment system should actually help teachers "make specific instructional decisions" (National Research Council, 2001, p. 35) and according to Gersten, Keating, and Irvin (1995), it should also provide data that indicates that use of the system is beneficial to students.

A group of researchers found the unit mastery tests problematic for several reasons. When they examined the psychometric characteristics of these measures (Fuchs, Tindal, & Fuchs, 1986; Tindal, Fuchs, Fuchs, Shinn, Deno, & Germann, 1985) they found them to be weak. In addition, difficulty levels varied from week to week, and the cut score of 80 or 85% seemed increasingly arbitrary. Unit mastery tests, sometimes called criterion-referenced tests in this era, did nothing to assess retention of previously taught material.

Almost 25 years ago, two seminal articles called for a radically different, seemingly counterintuitive approach to formative assessment (Fuchs, Deno, & Mirkin, 1984; Deno, 1986). This approach entailed a sampling of items representing major instructional objectives for the year and periodic use of assessments with items that randomly sampled across the year's objectives. This approach seems counterintuitive in that, during the early parts of the year, students are asked to solve problems involving material not yet covered. Toward the end of the year, they are asked about material they may have covered 6 to 8 months ago.

Yet, therein lies the power of a formative assessment that contains items from across the year's objectives. It is a far more accurate means to measure progress because the difficulty remains more or less the same across the year, and teachers and students can actually see the progress they have made toward acquiring the material. In contrast, typical mastery learning tests' difficulty level varied from unit to unit, depending on both the difficulty of the topic and the difficulty of the items selected. In addition, with this type of assessment system, students could actually see their progress; from say a score of 20% correct to 90% correct, as the year progressed. Even in the best of unit mastery tests, students will typically score at about the same level from unit to unit. Another advantage of this type of system is that it automatically assesses both retention of material taught months ago and, to some extent, a student's ability to generalize what she learned to unfamiliar material. Because each of the brief measures samples broadly across the years' objectives, criterion-related validity is far superior to assessments that only cover one week's unit. For all these reasons, these measures have consistently shown far superior reliabilities and criterion-related validity than traditional unit mastery tests (see Fuchs (2004) and Foegen et al. (2007) for extensive reviews). They also have consistently demonstrated *construct validity,* in particular, in terms of sensitivity to instruction, i.e. use as a means to reliably monitor student progress.

Especially in the field of reading, a second type of formative assessment was used, which also possessed strong psychometric qualities in terms of criterion-related validity (i.e., correlation with state- or nationally-normed achievement test) and reliability. These measures are typically called *robust indicators*. Foegen et al. (2007) define them as:

> *Measures that represent broadly defined proficiency in mathematics … Effective measures are not necessarily representative of a particular curriculum but are instead characterized by the relative strength of their correlations to various overall mathematics proficiency criteria (p. 4).*

These measures are "not necessarily drawn from the student's … (actual) ... curriculum, yet offer strong correlations to a host of criterion measures of overall subject area proficiency" (p. 4).

A potential advantage of robust indicators is that they can "create a seamless and flexible system of progress monitoring measures in mathematics … across multiple grade levels. The search for robust indicators represents an effort to identify aspects of core competence in mathematics … that are predictive of important outcomes in mathematics, regardless of the vagaries of specific curriculum programs or high stakes state tests."

Several robust indicators have been developed in the field of mathematics, especially for students in the primary grades. These include measures of number naming (e.g., VanDerHeyden, Witt, Naquin, & Noell, 2001; Chard, Clarke, Baker, Otterstedt, Braun, & Katz, 2005), magnitude comparison (e.g., Clarke & Shinn, 2004) and counting proficiency. Validity coefficients tend to be higher for first grade assessments than kindergarten assessments and highest for magnitude comparison measures. A drawback of these measures is that they are only useful for one grade level so that they cannot be used to assess progress over multiple years.

At the middle school level, Foegen (2000) developed two robust indicator measures: one of fluency with basic arithmetic combinations (i.e., facts) and the second, an estimation task. The estimation task was a timed measure and attempted to measure students' number sense (as opposed to computational skill).

Helwig and Tindal (2002) developed a measure that focused on conceptual understanding using an item bank developed for eighth-graders. In general, these measures demonstrated adequate criterion related validity, although the Helwig et al. conceptual measure demonstrated the strongest correlations with high-stakes assessments.

The authors note that, with one or two exceptions, neither the robust indicators approach nor the sampling from annual state curricular objectives approach have generated the same high levels of criterion related validity that oral reading fluency has in the field of reading. A major benefit of sampling from annual objectives is that teachers can use these data to obtain a sense of topics that require additional attention for groups of students. There are, however, several drawbacks.

The first is that, in order to be efficient, the sample of items should be limited. However, a limited sample of items may cause potential reliability issues. The second is that, at the current point in time, state standards in mathematics are quite variable in terms of quality, and different states provide differing emphases to topics and sequence topics differently (Reys, Dingman, Sutter, & Teuscher, 2005). Current efforts to use a common framework such as the *NCTM Focal Points* (National Council of Teachers of Mathematics, 2006) may help alleviate this problem in the future.

Both these types of measures were, in our view, unfortunately given the term, *curriculum based measurement.* That term seems to imply that they are valid, for example, only for a given curricula. Yet, in reality they are aligned to various district or states' mathematics standards. In 2007, Deno reported that the term used to describe this formative assessment approach was an unfortunate choice (Deno, 2007). Unlike unit mastery tests, the curriculum objective can basically be used for any curricula in use in a district because they gauge progress toward state standards.

Virtually all the applied experimental research on formative assessments has involved the use of these types of measures and an understanding of (a) the extent to which providing teachers and students with this information enhances mathematics achievement and, increasingly, (b) the efficacy of various tools and procedures for helping teachers use this information to provide differentiated instruction. This is the focus of the remainder of this section.

Most of the formative assessments used in mathematics demonstrate criterion-related validities in the 0.5 to 0.7 range (Foegen et al., 2007). Although these are weaker than those found in reading, they appear to be reasonable. It is in the area of *consequential validity*, that formative assessment measures have shown their greatest utility (Gersten et al., 1995). In the next section, the Task Group reviews the research on this topic that (a) examines impacts/effects of use of formative assessments and (b) addresses the standards for experimental and quasi-experimental design utilized by the Task Group.

C. Results

Table 32 presents the contrast between use of formative assessment on a regular (typically biweekly) basis versus a control condition. Six of the studies analyzed data at the individual student level, and the remaining three used the classroom or teacher as the unit of analysis. (One study (Calhoon & Fuchs, 2003) did not compare formative assessment to a control condition; it measured the impact of an enhanced version of formative assessment to a control condition. This study is thus excluded from this analysis.)

Table 32: Studies that Investigate the Impact of Formative Assessment (FA) Versus a Control Condition

Study	Design	Sample	Duration/ Content	Contrast	Measure	Subgroup	Hedge's g		Standard Error
Student-level analyses									
Allinder et al., 2000[a]	RCT	38 learning disabled elementary students and 22 teachers in a large midwestern school district	School year/ Curriculum based measurement in math computation	FA only vs. Control	Math Computation Test-Revised	Overall	-0.012	(ns)	0.349
Fuchs et al., 1990[a]	RCT	56 learning disabled students (Grades 3–9) and 20 elementary special education teachers in a southeastern metropolitan school district	15 weeks/ Individualized math programs	FA with performance indicator only vs. Control	Math Computation Test-Revised (Combined problems and digits)	Overall	0.239	(ns)	0.323
Fuchs et al., 1994	RCT	30 students and teachers in Grades 2–5 in a southeastern district	25 weeks/ Classwide program - general math operations	FA only vs. Control	Math Operations Test-Revised	Average achieving	0.310	(ns)	0.379
		30 students and teachers in Grades 2–5 in a southeastern district				Low achieving	0.015	(ns)	0.377
		30 students and teachers in Grades 2–5 in a southeastern district				Learning disabilities	0.189	(ns)	0.378
Fuchs et al., 1996[a]	RCT	22 learning disabled students (Grades 3–7) and 12 special education teachers in Tennessee metro school district	School year/ Aim was to reintegrate students into mainstream math	FA vs. Control	Math Operations Test-Revised (Digits)	Overall	0.468	(ns)	0.467
		25 learning disabled students (Grades 3–7) and 15 special education teachers in Tennessee metro school district		FA plus TP (trans-environmental programming) vs. TP only			0.220	(ns)	0.429
Fuchs et al., 1999[a]	RCT	272 students (Grades 2–4) and 16 teachers in four schools in one southeastern school district	23 weeks/ Problem solving	Performance Assessments (PA) vs. no PA	Novel problem-solving (based on ITBS problem)	Overall	0.355	(ns)	0.249
Spicuzza et al., 2001[a]	Quasi	495 students in Grades 4 and 5 in multiple schools in a large, midwestern school district	4 months/ Accelerated Math - individualized assignments	Accelerated Math vs. regular math program	NALT (Northwest Evaluation Association) – annual district test	Overall	0.139	(ns)	0.213
Heterogeneity									
Q-value	*Df (Q)*	*P-value*	*I-squared*	*Pooled ES: student level (six studies, nine effect sizes)*			*Hedge's g*		*Standard Error*
1.509	*8*	*0.993*	*0.000*				*0.206*	*~*	*0.107*

Continued on p. 6-177

Table 32, continued

Study	**Design**	**Sample**	**Duration/ Content**	**Contrast**	**Measure**	**Subgroup**	**Hedge's g**	**Standard Error**	
Classroom-level analyses									
Allinder, 1996	RCT	58 students (Grades 3–6) of 29 special education teachers in multiple schools in a large midwest school district	16 weeks/ Math computation	FA vs. Control	Math Computation Test-Revised	Overall	0.558	(ns)	0.387
Fuchs et al., 1989	RCT	40 students (Grades 2–9) of 20 special education teachers in elementary and middle schools in a southeastern metropolitan area	15 weeks/ Individualized math programs	Dynamic goal FA vs. Control	Math Computation Test	Overall	0.600	(ns)	0.439
Fuchs et al., 1991	RCT	42 learning disabled students (Grades 2–8) and 22 teachers in multiple elementary and middle schools in a southeastern metropolitan area	20 weeks/ Math operations	FA only vs. Control	Math Operations Test (combined problems and digits)	Overall	0.045	(ns)	0.426
Heterogeneity									
Q-value	*Df (Q)*	*P-value*	*I-squared*	*Pooled ES: student level (three studies, three effect sizes)*			*Hedge's g*	*Standard Error*	
1.069	*2*	*0.586*	*0.000*				*0.408*	*~*	*0.240*

~ p < .10, * p < .05, ** p < .01, *** p < .001
[a] Data were adjusted for clustering that occurred within teachers or classrooms.

Effect sizes were calculated for studies in which analysis was conducted at the individual student level (N = 6) and those where analysis was conducted at the classroom level (N = 3).

For the student level set, nine effect sizes from the six relevant studies were calculated. There are two reasons for this. The first is that Fuchs, Fuchs, et al. (1994) intentionally sampled students from three strata: those with learning disabilities, a below-average low-performing group, and a group of students performing at or near the class average. For Fuchs, Roberts, Fuchs, and Bowers (1996), there were four conditions, including two involving formative assessment (FA). Thus, effect sizes for two orthogonal contrasts could be calculated.

For the six studies where analyses were conducted at the student level, the mean effect size is 0.206, bordering on significance. For the three studies that use the class as the unit of analysis, the effect size is 0.408, also bordering on significance.

A reasonable inference is that merely providing teachers and students with feedback on how they are progressing is consistently helpful to students. This is a consistent replicable phenomenon across a large number of studies that involve well over a hundred classrooms. The reader should keep in mind that two-thirds of the research has been conducted at the elementary school level and only two include a middle school sample, and a one high school sample. In addition, all have used formative assessment systems that have empirical data to indicate validity and reliability. In addition, all but one (Ysseldyke et al., 2003), have used formative assessments that are based on sample problems selected to represent a randomly selected set of state standards for the year. Thus, there is insufficient evidence to determine whether or not the use of formative assessments is effective in the secondary grades. The next set of studies investigates what additional information is required to assist teachers in how to use these data.

D. Enhancements to Assist Teachers in Use of Formative Assessment

Early on, researchers realized that teachers might not know how to use formative assessment to enhance instruction unless some type of additional guidance was provided. Thus, a set of *enhancements* was developed and field-tested in a series of research studies. These appear in Tables 33 and 34. Table 33 compares the use of formative assessments with enhancement to a control condition (i.e., no formative assessment). Table 34 attempts to estimate the value added to formative assessment by each of these enhancements. Thus, the contrasts in Table 34 compare use of formative assessment with an enhancement to use of formative assessment.

Table 33: Studies That Investigate the Impact of Formative Assessment (FA) Plus Enhancements Versus a Control Condition

Study	Design	Sample	Duration/ Content	Contrast	Measure	Subgroup	Hedge's g	Standard Error	
Student-level analyses									
Allinder et al., 2000[a]	RCT	37 learning disabled elementary students and 20 teachers in a large midwestern school district	School year/ Curriculum based measurement in math computation	FA + teacher self-monitoring of instructional changes vs. Control	Math Computation Test-Revised	Overall	0.588	(ns)	0.369
Calhoon & Fuchs, 2003[a]	RCT	92 high school students with disabilities (Grades 9–12) and three teachers in 10 classrooms in three schools in a southeastern urban district	15 weeks/ Computation, concepts and applications	FA with PALS (Peer-assisted learning strategies) vs. Control	Math Operations Test-Revised (computation)	Overall	0.355	(ns)	0.340
Fuchs et al., 1990[a]	RCT	54 learning disabled students (Grades 3–9) and 20 elementary special education teachers in a southeastern metropolitan school district	15 weeks/ Individualized math programs	FA with performance indicator and skills analysis vs. Control	Math Computation Test-Revised (Combined problems and digits)	Overall	0.398	(ns)	0.325
Fuchs et al., 1994	RCT	30 students and teachers in Grades 2–5 in a southeastern district	25 weeks/ Classwide program - general math operations	FA + instructional recommendations vs. Control	Math Operations Test-Revised	Average achieving	0.292	(ns)	0.379
		30 students and teachers in Grades 2–5 in a southeastern district				Low achieving	0.546	(ns)	0.383
		30 students and teachers in Grades 2–5 in a southeastern district				Learning disabilities	0.172	(ns)	0.377
Fuchs et al., 1996[a]	RCT	24 learning disabled students (Grades 3–7) and 13 special education teachers in Tennessee metro school district	School year/ Aim was to reintegrate students into mainstream math	FA + transenvironmental programming (TP) vs. Control	Math Operations Test-Revised (Digits)	Overall	0.304	(ns)	0.445
Heterogeneity									
Q-value	*df (Q)*	*P-value*	*I-squared*	*Pooled ES: student level (five studies, seven effect sizes)*			*Hedge's g*	*Standard Error*	
0.901	*6*	*0.989*	*0.000*				*0.383*	**	*0.140*
Classroom-level analyses									
Fuchs et al., 1991	RCT	43 learning disabled students (Grades 2–8) and 22 teachers in multiple elementary and middle schools in a southeastern metropolitan area	20 weeks/ Math operations	FA with expert system instructional consultation vs. Control	Math Operations Test (combined problems and digits)	Overall	0.657	(ns)	0.439

~ p < .10, * p < .05, ** p < .01, *** p < .001
[a] Data were adjusted for clustering that occurred within teachers or classrooms.

Table 34: Studies That Investigate the Impact of Formative Assessment (FA) Plus Enhancements Versus Formative Assessment Only

Study	Design	Sample	Duration/ Content	Contrast	Measure	Subgroup	Hedge's g	Standard Error	
Student-level analyses									
Allinder et al., 2000[a]	RCT	33 learning disabled elementary students and 18 teachers in a large midwestern school district	School year/ Curriculum based measurement in math computation	FA + teacher self-monitoring of instructional changes vs. FA only	Math Computation Test-Revised	Overall	0.603	(ns)	0.386
Fuchs et al., 1990[a]	RCT	72 learning disabled students (Grades 3–9) and 20 elementary special education teachers in a southeastern metropolitan school district	15 weeks/ Individualized math programs	FA with performance indicator and skills analysis vs. FA with performance indicator only	Math Computation Test-Revised (Combined problems and digits)	Overall	0.234	(ns)	0.291
Fuchs et al., 1994	RCT	20 students and teachers in Grades 2–5 in a southeastern district	25 weeks/ Classwide program—general math operations	FA + instructional recommendations vs. FA only	Math Operations Test-Revised	Average achieving	-0.021	(ns)	0.428
		20 students and teachers in Grades 2–5 in a southeastern district				Low achieving	0.453	(ns)	0.434
		20 students and teachers in Grades 2–5 in a southeastern district				Learning disabilities	-0.037	(ns)	0.428
Fuchs et al., 1996[a]	RCT	24 learning disabled students (Grades 3–7) and 13 special education teachers in Tennessee metro school district	School year/ Aim was to reintegrate students into mainstream math	FA + transenvironmental programming (TP) vs. FA only	Math Operations Test-Revised (Digits)	Overall	-0.229	(ns)	0.444
Heterogeneity									
Q-value	*df (Q)*	*P-value*	*I-squared*	*Pooled ES: student level (4 studies, 6 effect sizes)*			*Hedge's g*	*Standard Error*	
2.949	*5*	*0.708*	*0.000*				*0.194*	*(ns)*	*0.159*
Classroom-level analyses									
Fuchs et al., 1991	RCT	41 learning disabled students (Grades 2–8) and 22 teachers in multiple elementary and middle schools in a southeastern metropolitan area	20 weeks/ Math operations	FA with expert system instructional consultation vs. FA without expert system	Math Operations Test (combined problems and digits)	Overall	0.750	~	0.443

~ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$
[a] Data were adjusted for clustering that occurred within teachers or classrooms.

Specific enhancements included:

1) *Providing teachers with detailed analysis indicating strengths and weaknesses based on formative assessment data*. The analysis describes student proficiency in specific mathematics subskills, and visually presents student proficiency in each subskill across the school year. The detailed information provided by the analysis allows teachers to evaluate how students maintain skills over time and helps teachers decide what to teach. Further, specific areas for instructional change can be targeted based on the information provided (Fuchs et al., 1990) (effect size of 0.398; value-added effect size of 0.234; these effect sizes did not reach statistical significance).

2) *Using data from formative assessments and sophisticated software to provide specific instructional suggestions to teachers for individual students.* Instructional consultation on teacher planning and student achievement was provided by a computerized expert system. Using the formative assessment data, the expert consultation system recommended instructional adjustments and provided detailed instructions on how to implement that adjustment. The consultation not only helped teachers isolate what material to re-teach but also how to restructure their instruction (Fuchs et al., 1991) (ES = 0.657, ns; value added ES = 0.750, bordering on significance.).

3) *Using the data from the formative assessments as a basis for the content of peer-assisted learning sessions.* Results of the formative assessments were entered into a computer program that produced a graph of students' progress overtime and a skills profile of each "student's performance on each skill in the annual curriculum" (p. 240). Teachers used this report to group students into pairs for Peer-Assisted Learning (PALS). The teachers also used the reports to determine the content of the PALS lesson (Calhoon & Fuchs, 2003) [ES = 0.355, nonsignificant (ns)]. This is the only study that addressed high school mathematics instruction.

4) *Using the data from formative assessments as the basis for consultation between the classroom teacher and the special educator to determine what content to emphasize.* The data from the formative assessments was used by the special education teacher to provide classroom teachers with specific information on which curricular areas require additional attention. Feedback from the formative assessments was also used to provide teachers with data on the effectiveness of various instructional strategies used to promote math achievement (Fuchs et al., 1996) (ES = 0.304; value added ES = -0.229, neither ES reached statistical significance).

5) *Self monitoring.* The self-monitoring process was completed each time the formative assessment data suggested the need for an instructional change. Teachers answered a set of questions regarding their students' progress and their future instructional plans. Using the information provided by the twice-weekly formative assessment, teachers responded to questions such as, "On what skill(s) has the student done well in the preceding 2 weeks? On what skill(s) has the student improved compared to the previous 2-week period? What skill(s) should be

targeted for the coming 2-week period? How will the teacher attempt to improve student performance on the targeted skill(s)?" (Allinder et al., 2000, p. 5). (ES = 0.588, value added ES = 0.603, neither ES reached statistical significance.)

6) *Using formative assessment data as a means for teachers to provide specific instructional suggestions for small group instruction and computer-assisted instruction.* Each teachers' weekly report, based on the formative assessment data included the following: (a) content that needed to be taught or retaught during whole class instruction, (b) specific suggestions on how to break the class into small groups for small group instruction and which content to cover, (c) individualized computer-assisted problems for each student, and (d) suggested material to cover during peer tutoring sessions (Fuchs et al., 1994). (Effect sizes were 0.546 for low-achieving students, 0.292 for average achieving students and 0.172 for students with learning disabilities; none of these effect sizes were statistically significant.) Note that for value added, effect size was 0.453 for low achievers, -0.021 for average achievers, and 0.37 for students with learning disabilities. However, effect sizes were negligible as none of them reached statistical significance.

The overall picture provided by the data in Table 33 indicates that the set of enhancements are effective in enhancing students' mathematics achievement. The average effect size, in this case, is significant for studies conducted at the student level [average ES = 0.383, statistically significant, ES = 0.657, (ns) for the class level study]. As was the case for the first set of analyses, the same pattern emerges whether or not the full set of studies is included, or only those where an effect size could be computed at the student level.

Note that the effect of formative assessment with enhancements increases dramatically from formative assessment alone. It appears that the approaches that provided specific suggestions directly to the teacher about what to teach during small group instruction or partner work, or provided specific instructional suggestions worked better than this more indirect method.

Table 34 presents effect sizes for the value added by the enhancement. In other words, the comparison condition involves use of formative assessment only. As one would expect, with the more stringent criterion, the mean effect size is much lower, 0.194, which is not statistically significant. In two cases, the enhancement did not provide any additional gain in terms of student achievement to the mere use of formative assessment. On average the effect size doubles when an enhancement is added.

It is important to note that the majority of these studies focus on students with diagnosed learning disabilities. Only two samples (both from the same study) involve students from the general population. It is also important to note that these studies involve only one of two dependent measures, the Mathematics Operations Test, or the Mathematics Concepts and Applications Tests. These tests were developed by the researchers. However, they do possess solid psychometric properties.

The final caveat is that many of the studies involve special education teachers. Thus, one should be cautious in interpreting implications for classroom teachers because few of the enhancements involved only the classroom teacher.

E. Summary and Conclusions

The set of ten well-designed and well-executed studies on formative assessment demonstrates that regular use of formative assessments in mathematics can enhance students' mathematics achievement in the elementary grades, in both the areas of computation and concepts and applications. These studies were conducted with moderately large numbers of teachers in "real-world" settings; thus the external validity is high. The average effect size boost provided by use of formative assessments for studies conducted at the individual level is 0.206 and approaches significance. This corresponds to a boost of 9 percentile points.

In addition, the set of studies describes a set of tools and procedures (what the Task Group calls "enhancements") that can accompany formative assessment. These tools include specific activities that are linked to a student's current needs. Activities range from a list of ideas for alternate means of teaching the material, to specific materials for use in peer tutoring, to a listing of skills and concepts that require additional explanation and discussion.

Although many of the effect sizes doubled in value with these enhancements, the net contribution of 0.194 was not significant. Thus, the Task Group would more cautiously call these practices promising as opposed to evidence-based.

Two other issues need to be considered in framing specific recommendations for improving practice. The first is that the preponderance of studies were conducted at the elementary school. Second, to date, only one type of formative assessment has been studied with rigorous experimentation. These are assessments that include random sampling of items that address state standards. These assessments tend to take between 2 and 8 minutes to administer and thus are feasible for regular use. However, as discussed in the Introduction, many other types of formative assessments have been developed. The Task Group simply cannot comment on how useful these other types are in terms of enhancing students' performance at this point in time since the Task Group was unable to uncover any rigorous experiments involving their use. Hopefully, such research will be conducted in the near future.

F. Proposed Recommendations

Schools should seriously consider regular use of formative assessments in mathematics.

This might entail weekly assessments of students experiencing difficulties and less frequent (perhaps three times a year) assessments for others.

The Task Group would recommend use of formative assessments with known validity and reliability. However, the Task Group is aware that at the current point in time, there is a paucity of such measures. The Task Group advocates serious research and development in this area. It appears such work has already begun and federal support of this effort seems critical. In particular, validity studies of methods other than those that sample from annual state or district goals could and should be explored.

Research findings suggest that several enhancements can help teachers use formative assessment information more effectively.

Here, the research base is smaller, and less consistent. Several major tools appear to be promising. The first is linking formative assessment information (via technology) with specific recommendations for a teacher in areas such as a) content and concepts that require additional work with the majority of the class and b) specific activities that could and should be used by a given student for either tutoring or computer-assisted intervention or intervention work provided by an adult (teacher, mathematics specialist, or trained paraprofessional).

Use of formative assessments in mathematics can lead to increased precision in how time is used in class and assist teachers in providing appropriate instruction to students who need help on topics for which they need help. This should seriously be considered as districts consider the development and implementation of response-to-intervention models in mathematics.

G. Suggestions for Future Research

Several future research areas seem worth pursuing. The first is extending this line of research to the middle school and high school area. The second entails the same type of rigorous research studies of other, more clinical types of formative assessments such as those described in recent publications by NCTM.

The Task Group also needs to know more about how formative assessment measures relate to mathematics tests that include items that are more mathematically sophisticated than those on current standardized achievement tests. It also is important to update the studies of the reliability and validity of publisher-developed tests. That research is now over 20-years-old and the nature of mathematics instruction has changed dramatically.

BIBLIOGRAPHY

Allinder, R.M. (1996). When some is not better than none: Effects of differential implementation of curriculum-based measurement. *Exceptional Children, 62(6)*, 525–535.

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 & Special Education, 21(4)*, 219–226.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: AERA Publications.

Bloom, B.S. (1976). *Human characteristics and school learning.* New York: McGraw-Hill.

Bloom, B.B. (1980). *All our children learning.* New York: McGraw-Hill.

Calhoon, M.B., & Fuchs, L. (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(4)*, 235–245.

Chard, D., Clarke, B., Baker, B., Otterstedt, J., Braun, D., & Katz, R. (2005). Using measures of number sense to screen for difficulties in mathematics: Preliminary findings. *Assessment Issues in Special Education*, *30*(2), 3–14.

Clarke, B., & Shinn, M.R. (2004). A preliminary investigation into the identification and development of early mathematics curriculum-based measurement. *School Psychology Review, 33(2)*, 234–248.

Deno, S.L. (1986). Formative evaluation of individual student programs: A new role for school psychologists. *School Psychology Review*, *15(3)*, 358–374.

Deno, S. (2007, April). Distinguished Research Award. Presentation at the American Educational Research Association Annual Meeting. Chicago, IL.

Donovan, M.S., & Bransford, J.D. (2005). Introduction. In the National Research Council (Ed.), *How students learn: Mathematics in the classroom* (pp. 1–30). Washington, DC: National Academy Press.

Foegen, A. (2000). Technical adequacy of general outcome measure for middle school mathematics. *Diagnostique*, *25*, 175–203.

Foegen, A., Jiban, C., & Deno, S. (2007). Progress monitoring measures in mathematics: A review of the literature. *Journal of Special Education*, *41*, 121–139.

Freudenthal, H. (1973). *Mathematics as an educational task.* Dordrecht, Holland: Reidel.

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

Fuchs, L.S. (2004). The past, present, and future of curriculum-based measurement research. *School Psychology Review*, *33*, 188–192.

Fuchs, L.S., Deno, S.L., & Mirkin, P.K. (1984). The effects of frequent curriculum-based measurement and evaluation on pedagogy, student achievement, and student awareness of learning. *American Educational Research Journal, 21(2),* 449–460.

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

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(1),* 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(3),* 617–641.

Fuchs, L.S., Fuchs, D., Hamlett, C.L., Thompson, A., Roberts, P.H., Kupek, P., et al.(1994). Technical features of a mathematics concepts and applications curriculum-based measurement system. *Diagnostique, 19(4),* 23–49.

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

Fuchs, L.S., Tindal, G., & Fuchs, D. (1986). Effects of mastery learning procedures on student achievement. *Journal of Educational Research, 79(5),* 286–291.

Gersten, R., Keating, T.J., & Irvin, L.K. (1995). The burden of proof: Validity as improvement of instructional practice. *Exceptional Children, 61(6),* 510–519.

Glaser, R., & Silver, E. (1994). Assessment, testing and instruction: retrospect and prospect. *Review of Research in Education, 20,* 393–419.

Guskey, T.R. (1984). The influence of change in instructional effectiveness upon the affective characteristics of teachers. *American Educational Research Journal, 21(2),* 245–259.

Helwig, R., & Tindal, G. (2002). Using general outcome measures in mathematics to measure adequate yearly progress as mandated by Title I. *Assessment for Effective Intervention, 28(1)*, 9–18.

Jiban, C., Deno, S.L., & Foegen, A. (in press). Progress monitoring in mathematics: A review of the literature. *Exceptional Children*.

Lee, S.W. (2005). *Encyclopedia of school psychology*. Thousand Oaks, CA: Sage Publications, Inc.

Messick, S. (1988). The once and future of validity: Assessing the meaning and consequences of measurement. In H. Wainer & H. I. Braun (Eds.), *Test Validity* (pp. 33–46). Hillsdale, NJ: Erlbaum.

National Council of Teachers of Mathematics (NCTM) (1989). *Curriculum and evaluation standards for school mathematics.* Reston, VA: Author.

National Council of Teachers of Mathematics (NCTM). (2006). *Curriculum focal points for prekindergarten through Grade 8 mathematics: A quest for coherence.* Reston, VA: Author.

National Research Council. (2001). *Adding it up: Helping children learn mathematics*. In J. Kilpatrick, J. Swafford, and B. Findell (Eds.), Mathematics learning study committee, center for education, division of behavioral and social sciences, and education. Washington, DC: National Academy Press.

National Research Council (2005). *How students learn: Mathematics in the classroom.* Washington, DC: National Academy Press.

Renaissance Learning (1998). Accelerated Math [computer software]. Wisconsin Rapids, WI: Author.

Reys, B.J., Dingman, S., Sutter, A., & Teuscher, D. (2005) *Development of state-level mathematics curriculum documents: Report of a survey.* Columbia, MO: University of Missouri, Center for the Study of Mathematics Curriculum, 2005. Retrieved on February 14, 2006, from http://www.mathcurriculumcenter.org/resources/ASSMReport.pdf.

Ruiz-Primo, M.A., Shavelson, R.J., Hamilton, L., & Klein, S. (2002). On the evaluation of systemic science education reform: Searching for instructional sensitivity. *Journal of Research in Science Teaching, 39(5)*, 369–393.

Spicuzza, R., Ysseldyke, J., Lemkuil, A., Kosciolek, S., Boys, C., & Telluchsingh, E. (2001). Effects of curriculum-based monitoring on classroom instruction and math achievement. *Journal of School Psychology, 39(6)*, 521–542.

Spitzer, H.F. (1951). Learning and Teaching Arithmetic. In N.B. Henry, & G.T. Buswell, *The teaching of arithmetic—Fiftieth yearbook, part 11, National Society for the Study of Education*. Chicago: University of Chicago Press.

Sueltz, B.A., Boynton, H., & Sauble, I. (1946). The measurement of understanding in elementary-school mathematics. In *Measurement of understanding, 45th yearbook of the national society for the study of education, part I*. Chicago: University of Chicago Press.

Tindal, G., Fuchs, L.S., Fuchs, D., Shinn, M.R., Deno, S.L., & Germann, G. (1985). Empirical validation of criterion-referenced tests. *Journal of Educational Research, 78(4)*, 203–209.

VanDerHeyden, A.M., Witt, J.C., Naquin, G., & Noell, G. (2001). The reliability and validity of curriculum-based measurement readiness probes for kindergarten students. *School Psychology Review, 30(3),* 363–382.

Ysseldyke, J., Spicuzza, R., Kosciolek, S., Teelucksingh, E., Boys, C., & Lemkuil, A. (2003). Using a curriculum-based instructional management system to enhance math achievement in urban schools. *Journal of Education for Students Placed at Risk*, 8(2), 247–265.

Conclusion

Mathematics instruction is a complex professional practice. Researchers in the educational research community have made important forays into several of the most controversial and pressing questions about the effectiveness and impact of various types of instructional practice and, in particular, have conducted some studies that examine the effects of various interpretations and implementations of practices that have been advocated in the "reform" documents in mathematics education during the past two decades.

The question we asked is: *What can be learned from a review of the best available evidence in six important aspects of practice?* These practices included: the use of cooperative groups and peer instruction, the use of direct instruction with learning disabled students, the use of "real-world" problems in mathematics teaching, the use of technology, the enrichment and acceleration of instruction for mathematically precocious students, and the use of formative assessment.

For none of the areas examined did the Task Group find sufficiently strong and comprehensive bodies of research to support all-inclusive policy recommendations of any of the practices addressed. Nor did the Task Group find sufficient evidence to support policy recommendations favoring the status quo in mathematics teaching.

Across all of the areas, the Task Group found that **several instructional practices in mathematics teaching show some promise, in comparison to typical practice, for affecting student learning.** In each case the "promising" practice is clearly specified, somewhat prescriptive, and involves a mix, or combination, of particular, distinct practices. Thus, for example, it cannot be said that cooperative learning is a practice whose effectiveness is supported by research—but the Team Assisted Individualization (TAI) approach, with particular students in a particular area of mathematics, is effective. Although formative assessment to inform instruction is useful, it is enhanced when teachers use assessment tools with known validity and reliability. For students performing in the lower third of their grade level expectations, explicit instruction involving clear models of proficient performance, many opportunities to verbalize their problem solving strategies, and adequate practice and review should be a part of the mathematics program. It is not surprising that what the Task Group found about effective instructional practice is far more subtle and nuanced than direct answers to the starkly stated questions investigated.

The Task Group found some rather robust findings but they must be accompanied by a caveat. When a practice is demonstrated by high-quality experimental research to have some promise, it is critical to be clear about the promise "for what aspects of mathematics proficiency?" Different practices and approaches impact different kinds of outcomes, ranging from computational performance, to "real-world" problem solving, to identifying extraneous problem information, to long-term participation and interest in studying mathematics.

Because researchers and practitioners use different definitions to describe their interventions, it is conceptually problematic to place too much stock in either generalizing that a broad category of practice (e.g., using technology, or using "real-world" problems) has impact because a set of studies working on the same particular component of this category have impact, which was the case in some of the Task Group's reviews.

The Task Group's process included asking mathematicians and mathematics education reviewers to examine the mathematical content of the research studies—to look at the assessments and interventions, to the extent possible, based on the published reports. They expressed important concerns, including the possibility that an outcome measure item purported to measure computation might not do so because it really measured ability to use the context, for instance. They expressed concern that some topics were underdeveloped (i.e., failed to help students access the underlying mathematics in the topic covered), or that items were mislabeled (e.g., as "problem solving") when a mathematics expert might classify them otherwise. However, they also did note that several of the studies we reviewed seemed to help students increase their knowledge of mathematics and their ability to apply that knowledge to novel situations is a fashion that is valid from a mathematical perspective.

The reader may feel disappointed at this juncture, seeing how few robust findings emanated from a review of the rigorous research on the topics addressed. Yet even the inconclusive and limited findings can provide a real service to the profession. If an administrator, a curriculum developer or a parent comments, "Research says that lessons must start with 'real-world' problems," or "Students will really learn mathematics only if they are taught using direct instruction," consumers and professionals now know that research is inconclusive on these topics. This is a necessary step in the evolution of educational research into a more mature science. The paucity of findings and the paucity of high-quality experimental research in the field led the Task Group to realize, early on in the process, that few definitive answers to the research questions posed would be found.

However, the Task Group did see this work as the starting point for creating a base of knowledge to answer the questions posed at the onset of this work. We also see the application of the rigorous standards (developed in large part through earlier work of the Institute of Education Sciences of U.S. Department of Education) as serving as guidelines for the next generation of researchers.

The questions and topics studied and findings are briefly summarized below.

A. How Effective Is Teacher-Directed Instruction in Mathematics in Comparison to Student-Centered Approaches, Including Cooperative and Collaborative Groups, in Promoting Student Learning? Is One Approach Preferable to Another? If So, in Which Areas of Mathematics?

A controversial issue in the field of mathematics teaching and learning is whether classroom instruction should be more teacher-directed or student-centered. These terms have come to incorporate a wide array of meanings, with teacher-directed ranging from highly scripted direct instruction approaches to interactive lecture styles, and with student-centered ranging from students having primary responsibility for their own mathematics learning to highly structured cooperative groups. Schools and districts must make choices about curricular materials or instructional approaches that often seem more aligned with one instructional orientation than another. This leaves teachers wondering about when to organize their instruction one way or the other, whether certain topics are taught more effectively with one approach or another, and whether certain students benefit more from one approach than the other.

In the review, the Task Group limited the search to studies that directly compared these two extreme positions. Teacher-directed instruction was defined as instruction in which it is the teacher who is primarily communicating the mathematics to the students directly, and student-centered instruction as instruction in which primarily students are doing the teaching.

Only eight studies were found that met the Task Group's standards for quality that were consistent with this definition. The studies presented a mixed and inconclusive picture of the relative impact of these two forms of instruction. High-quality research does not support the contention that instruction should be either entirely "child-centered" or "teacher-directed." Research indicates that some forms of particular instructional practices can have a positive impact under specified conditions. All-encompassing recommendations that instruction should be entirely "child-centered" or "teacher-directed" are not supported by research. The limited research base of rigorous research does not support the exclusive use of either approach.

1. Cooperative and Collaborative Groups

One of the major shifts in education over the past 25–30 years has been advocacy for the increased use of cooperative learning groups and peer-to-peer learning (e.g., structured activities for students working in pairs) in the teaching and learning of mathematics.

Research has been conducted on a variety of cooperative learning approaches. One such approach, Team Assisted Individualization (TAI) has been shown to significantly improve students' computation skills. This instructional approach involves heterogeneous groups of students helping each other, individualized problems based on student performance on a diagnostic test, and rewards based on both group and individual performance. Effects on conceptual understanding and problem solving were not significant. There is evidence suggesting that working in dyads with a clear structure also improves computation skills in the elementary grades. However, additional research is needed.

B. What Is the Impact of Use of Formative Assessment in Mathematics Teaching?

Formative assessment—the ongoing monitoring of student learning to inform instruction—is generally considered a hallmark of effective instruction in any discipline. The Task Group's review of the high-quality studies of this topic produced several conclusions.

Teachers' regular use of formative assessment is marginally significant in improving their students' learning. This is especially true if teachers have additional guidance on using the assessment to design and individualize instruction.

Although the research base is smaller, and less consistent than that on the general effectiveness of formative assessment, the research does suggest that several specific tools and strategies can help teachers use formative assessment information more effectively. The first promising strategy is providing formative assessment information to teachers (via technology) on content and concepts that require additional work with the whole class. The second promising strategy involves using technology to specify activities needed by individual students. Both of these aids can be implemented via tutoring, computer-assisted instruction, or help provided by a professional (teacher, mathematics specialist, trained paraprofessional).

We caution that only one type of formative assessment has been studied with rigorous experimentation. These are assessments that include random sampling of items that address state standards. These assessments tend to take between 2 and 8 minutes to administer and thus are feasible for regular use.

The regular use of formative assessment particularly for students in the elementary grades is recommended. These assessments need to provide information not only on their content validity but also on their reliability and their criterion-related validity (i.e., correlation of these measures with other measures of mathematics proficiency). For struggling students, frequent (e.g., weekly or biweekly) use of these assessments appears optimal, so that instruction can be adapted based on student progress.

Research is needed regarding the content and criterion-related validity and reliability of other types of formative assessments (such as unit mastery tests included with many published mathematics programs, performance assessments, and dynamic assessments involving "think alouds"). This research should include studies of consequential validity (i.e., the impact they have on helping teachers improve their effectiveness).

Use of formative assessments in mathematics can lead to increased precision in how instructional time is used in class and can assist teachers in identifying specific instructional needs. Formative measures provide guidance as to the specific topics needed for assistance. Formative assessment should be an integral component of instructional practice in mathematics.

C. What Instructional Strategies for Teaching Mathematics to Students With Learning Disabilities and to Low-Achieving Students Show the Most Promise?

A review of 26 high-quality studies, mostly using randomized control designs, was conducted. These studies provide a great deal of guidance concerning some defining features of effective instructional approaches for students with learning disabilities (LD) as well as low-achieving (LA) students.

Explicit systematic instruction typically entails teachers explaining and demonstrating specific strategies, and allowing students many opportunities to ask and answer questions and to think aloud about the decisions they make while solving problems. It also entails careful sequencing of problems by the teacher or through instructional materials to highlight critical features. More recent forms of explicit systematic instruction have been developed with applications for these students. These developments reflect the infusion of research findings from cognitive psychology, with particular emphasis on automaticity and enhanced problem representation.

Our analysis of the body of research indicated that explicit methods of instruction are consistently and significantly effective with students with learning disabilities in computation, solving word problems, and solving problems that require the application of mathematics to novel situations.

Only a small number of studies were located that investigated the use of visual representations or student “think alouds.” Therefore no inferences about their effectiveness can be drawn. The research suggests that they are most useful when they are integrated with explicit instruction.

Based on this admittedly small body of research, we conclude that students with learning disabilities and other students with learning problems should receive some time on a regular basis with explicit systematic instruction. There is no reason to believe that this type of instruction should comprise all the mathematics instruction these students receive. However, it does seem essential for building proficiency in both computation and the translation of word problems into appropriate mathematical equations and solutions. Some of this time should be dedicated to ensuring that students possess the foundational skills and conceptual knowledge necessary for understanding the mathematics they are learning at their grade level.

D. Do "Real-World" Problem Approaches to Mathematics Teaching and Efforts to Ensure that Students Can Solve 'Real-World' Problems, Lead to Better Mathematics Performance Than Other Approaches?

The meaning of the term "real-world" problem varies by mathematician, researcher, developer, and teacher. Conducting research in this area is complex; fidelity of the teachers' implementation of the instructional materials or instructional strategy is difficult to assess. Although not addressed in the studies we examined, teachers' knowledge and capacity to use such problems effectively varies greatly. Given these caveats, the Task Group addressed the question of whether using "real-world" contexts to introduce and teach mathematical topics and procedures is preferable to more typical instructional approaches.

The body of high-quality studies for this topic is small. Five studies addressed the question of whether the use of "real-world" problems as the instructional approach led to improved performance on outcome measures of ability to solve "real-world" problems, as well as on more traditional assessments. Four of these studies were similar enough to combine in a meta-analysis. The meta-analysis revealed that if mathematical ideas are taught using "real-world" contexts, then students' performance on assessments involving similar problems is improved. However, performance on assessments of other aspects of mathematics learning, such as computation, simple word problems, and equation solving, is not improved.

For certain populations (upper elementary and middle grade students and remedial ninth-graders) and for specific domains of mathematics (fraction computation, basic equation solving, and function representation), instruction that features the use of "real-world" contexts can have a positive impact on certain types of problem solving. Additional research is needed to explore the use of "real-world" problems in other mathematical domains, at other grade levels, and with varied definitions of "real-world" problems.

E. What Is the Relative Impact on Mathematics Learning When Students Use Technology Compared to Instruction That Does Not Use Technology?

1. Calculators

A review of 11 studies that met the Task Group's rigorous criteria (only one study was less than 20 years old) found limited to no impact of calculators on calculation skills, problem-solving, or conceptual development over periods of up to one year. Unfortunately, these studies cannot be used to judge the advantages or disadvantages of multiyear calculator use beginning in the early years, because such long-term use has not been adequately investigated. The Panel cautions that to the degree that calculators impede the development of automaticity, fluency in computation will be adversely affected.

2. Computer-Assisted Instruction and Computer Programming

We found that CAI drill and practice, if of high quality, can improve students' performance compared to conventional instruction, with the greatest effect on computation, and less effect on concepts and applications. Drill and practice programs **can** be considered as a useful tool in developing students' automaticity, or fast, accurate, and effortless performance on computation, freeing working memory so that attention can be directed to the more complicated aspects of complex tasks.

Research has demonstrated that tutorials (CAI programs, often combined with drill and practice) that are well designed and implemented can have a positive impact on mathematics performance, particularly at the middle and high school levels. CAI tutorials have been used effectively to introduce and teach new subject-matter content. However, these studies also suggest several important caveats. Care must be taken to ensure that there is evidence that the software to be used has been shown to increase learning in the specific domain and with students who are similar to those who are under consideration. Educators should critically inspect individual software packages and studies that evaluate them critically. Furthermore, the requisite support conditions to use the software effectively (sufficient hardware and software; technical support; adequate professional development, planning, and curriculum integration) should be in place, especially in large-scale implementations, to achieve optimal results.

Research indicates that computer programming improves students' performance compared to conventional instruction, with the greatest effects on understanding of concepts and applications, especially geometric concepts, and weaker effects on computation. However, computer programming by students can be employed in a wide variety of situations using distinct pedagogies, not all of which may be effective. Therefore, the findings are limited to the careful, targeted application of computer programming for learning used in the studies reviewed.

F. What Instructional Arrangements for Engaging with Mathematics Are Most Promising for Mathematically Precocious Students?

The Task Group's review of the literature about what kind of mathematics instruction would be most effective for gifted students focused on the impact of programs involving acceleration, enrichment, and the use of homogeneous grouping. The extensive literature searches we conducted yielded few studies that met the Task Group's methodologically rigorous criteria for inclusion. Thus for this topic—and this topic only—we relaxed these criteria in order to fulfill our charge of evaluating the "best available scientific evidence." One randomized control trial study and seven quasi-experimental studies were located. All but one of these studies have limitations.

Despite the flaws in any one study, the set of studies suggests there is value to differentiating the mathematics curriculum for students who are gifted in mathematics and possess sufficient motivation, especially when acceleration is a component (i.e., pace and level of instruction are adjusted). A small number of studies suggest that individualized instruction, in which pace of learning is increased and often managed via computer instruction, produces gains in learning.

Gifted students who are accelerated by other means not only gained time and reached educational milestones earlier (e.g., college entrance) but appear to achieve at levels at least comparable to those of their equally able same-age peers on a variety of indicators even though they were younger when demonstrating their performance on the various achievement benchmarks. One study suggests that gifted students also appear to become more strongly engaged in science, technology, engineering, or mathematical areas of study.

Some support also was found for supplemental enrichment programs. Of the two programs analyzed, one explicitly utilized acceleration as a program component and the other did not. Self-paced instruction supplemented with enrichment seemed to have a positive impact on student achievement. This supports the view in the field of gifted education that acceleration and enrichment combined should be the intervention of choice. We believe it is important for school policies to support appropriately challenging work in mathematics for gifted and talented students.

G. What Would the Instructional Practices Task Group Say to the Practitioner?

There is no one ideal approach to teaching mathematics; the students, the mathematical goals, the teacher's background and strengths, and the instructional context, all matter. The findings here do suggest that it is especially important:

- to monitor what students understand and are able to do mathematically;
- to design instruction that responds to students' strengths and weaknesses, based on research when it is available; and
- to employ instructional approaches and tools that are best suited to the mathematical goals, recognizing that a deliberate and conscious mix of strategies will be needed.

Also, it is important for teachers, school administrators, and the public to understand the importance of helping to formulate research questions and being willing to participate in the types of experimental and quasi-experimental studies that are described here.

H. What Would the Instructional Practices Task Group Say to the Researcher?

More research that can identify causal claims is needed to guide both policy and practice. Building the mathematics education research portfolio to include this work will involve:

- Formulation of research questions that are of interest to practitioners and policymakers;
- Collaborations among mathematicians, mathematics education researchers, methodologists, and psychometricians; and
- Motivation to design and undertake rigorous studies.

The work of this Task Group has substantiated our understanding of the complexity and challenge of effective mathematics instruction. It is now up to practitioners, policymakers, mathematicians, and mathematics education researchers to take up the challenges of clarifying the definitions of mathematics instructional practices, debunking myths about mathematics instruction, and formulating the types of research studies that can answer the pressing questions that need to be addressed.

APPENDIX A: Methodological Procedures

Methodology for the Instructional Practices Task Group Research Reviews

From the onset, the Instructional Practices Task Group was committed to assembling the most rigorous scientific research addressing questions of effectiveness about the types of interactions that occur in mathematics classrooms relative to student performance. The Task Group was aware that there might be a paucity of such studies. This issue of understanding the quality of evidence and design needed to lead to causal inference is discussed in the *Standards of Evidence* document approved by the National Mathematics Advisory Panel. However, it is particularly germane to this topic, in that before requiring widespread implementation of a particular instructional practice or intervention, or committing significant resources toward such implementation, it seems critical to know that it will in all likelihood lead to higher levels of mathematics proficiency than alternatives. Of the six topics investigated as part of the Panel report, the topic of instructional practices was the topic for which the most experimental research was available. The Task Group thus chose to review and synthesize only the highest quality experimental and quasi-experimental research, research that can lead to causal inference, as the primary goal. In some cases, the Task Group also relied on the best available evidence suitable to a particular issue.

The recent report by the National Research Council (NRC), *Scientific Research in Education* (2002), was influential in the decision. The authors note, "[They] believe that attention to the development and systematic testing of theories and conjectures across multiple studies and using multiple methods—a key scientific principle … is currently undervalued in education relative to other scientific fields" (p. 124). They go on to note, "While large-scale education policies and programs are constantly undertaken… they are typically launched without an adequate evidentiary base to inform their development, implementation or refinement over time…" (p. 124). The report also states: "Randomized experiments are not perfect.… For instance, they typically test complex causal hypotheses, they may lack generalizability to other studies, and they can be expensive. *However, we believe that these and other issues do not generate a compelling rationale against their use in education research and that issues related to ethical concerns, political obstacles and other potential barriers often can be resolved"* (p. 125, emphasis added). Whereas the field of reading instruction has made great strides through a combination of randomized controlled trials (RCTs), longitudinal research, descriptive research and qualitative research, there is less of a history in mathematics education research of using RCTs until recently.

Selection of Topics

The original members of the Task Group[16] devoted the first two meetings to decisions about the array of topics to study. The Task Group followed a process similar to that used by the National Reading Panel. After extended brainstorming and discussion, a list of approximately 20 topics was developed and then each member selected the top six.

No particular theoretical framework was used to generate this list. Panelists selected topics that were perceived as: a) high interest to the teachers and policymakers, b) areas requiring additional attention in terms of implementation of recent federal policies such as No Child Left Behind (NCLB) and Individuals with Disabilities Act (IDEA) of 2004, or c) topics deemed critical by organizations such as National Council of Teachers of Mathematics (NCTM).

In addition, based on cumulative knowledge of the research literature, the Task Group wanted to include at least one or two topics for which adequate research was available to provide empirically based recommendations. This seemed particularly important because three recent reports by the National Research Council (2001, 2002, 2004)[17] noted an extremely limited amount of rigorous research in this field; too few studies were available to draw causal inferences.

This resulted in a list of 12 topics. Due to time constraints, the Task Group was unable to address all of the 12 topics. The following eight received the most support:

1) “Real-world” problem solving
2) Relative effectiveness of explicit or teacher-centered instruction vs. child-centered or inquiry based instruction
3) Formative assessment
4) Cooperative, collaborative learning and peer-assisted instruction
5) Instructional strategies for students with learning disabilities
6) Instructional strategies for low-performing students
7) Instructional strategies for mathematically precocious students
8) Technology with a particular focus on use of graphing calculators and single function calculators

Among the topics that generated a good deal of interest, but were excluded due to time constraints, were: a) importance of time spent engaged in mathematics, b) guidelines for developing homework assignments, c) best practices in terms of review of previously taught material, and d) types of practice problems and the sequencing of practices.

[16] One of the original members of the Task Group (Diane Jones) left when she was assigned to another position in the federal government; she was replaced by Irma Arispe in June, 2007. In April, 2007, Bert Fristedt and Douglas Clements joined the group and Joan Ferrini-Mundy replaced Kathie Olsen in January 2007. By that point, most of the topics had been finalized, although three topics were subsequently eliminated.

[17] *Adding It Up (2001), Scientific Research in Education (2002), On Evaluating Curricular Effectiveness (2004).*

The topic of curriculum and instructional materials was assigned to a small task force to preclude even the appearance of conflict of interest, upon advice from the ethics attorney of the U.S. Department of Education.

Instructional Practices Task Group Methodology Statement

The Instructional Practices Task Group organized the available scientific evidence into several categories of evidence for consideration as they reviewed studies related to each topic of mathematics instruction investigated. A discussion of the categories for studies with quantitative designs that were considered for inclusion as rigorous evidence is provided below followed by a discussion of the procedures for identifying relevant research and synthesizing the research. Any deviations from the general practices outlined below are specified in the individual sections of the report.

Category 1: Experimental and Quasi-Experimental Studies that Meet or Meet with Reservations What Works Clearinghouse (WWC) Standards

Studies in this category provide evidence of causal claims and include randomized control trials (RCT; use random assignment to create experimental groups) and strong quasi-experimental studies (QED; experimental groups created by a method other than random assignment) that meet WWC Criteria. These criteria can be found at http://ies.ed.gov/ncee/wwc/overview/review.asp?ag=pi and http://ies.ed.gov/ncee/wwc/twp.asp.

The only cases where exceptions to WWC criteria were allowed are:

- Differential attrition rates of up to 30% are permitted for RCTs and for QEDs if there is evidence that attrition does not affect the nature of the sample on a salient pretest variable.
- Studies that assign only one school per condition are acceptable provided that there are several teachers per condition.

In all other areas, the Task Group followed the WWC policies expressed on www.whatworks.org. QEDs were excluded if they fail to either provide evidence of pretest comparability or control for pretest differences. Thus, the Task Group downgraded RCTs and exclude QEDs if a) there is evidence of contamination; b) there is only one teacher per experimental condition. However, if both the treatment and control conditions were taught by the same teacher, these were reviewed on a case-by-case basis, and the study may have been included is there was reason to believe that there was no bias in delivery.

Category 1 studies are the core of the results section for the Instructional Practices Task Group as they represent clear evidence to support causal claims. Category 1 studies correspond to high and moderate quality studies, as defined by the National Mathematics Advisory Panel Guidelines for Standards of Evidence.

Category 2: Weak Group Comparison Studies and Other Quantitative Designs that Attempt to Infer Causality

Category 2 consisted of weak group comparison studies (failed RCTs and weak nonequivalent comparison designs). Category 2 studies are always open to multiple interpretations with regard to causal inferences, however, they are not necessarily weak studies for other purposes (e.g., descriptive). Category 2 studies correspond to moderate and low-quality studies, as defined by the National Mathematics Advisory Panel Guidelines for Standards of Evidence.

Category 2: Weak Group Comparison Studies. These are attempts at experiments or quasi-experiments that are seriously flawed (e.g., one teacher per condition, widely differential attrition across the experimental groups, quasi-experiments with no evidence of pretest equivalence). This category would also include studies in which all the dependent measures are closely aligned to the instructional content of the intervention and not at all covered in the control condition. These studies are considered biased. Studies in which the experimental sample consists only of volunteers and the control group only of those who declined to participate would also be considered a weak comparison study. Because of the serious nature of the flaws, the Task Group would *not* consider these as providing valid causal evidence.

Category 2 studies are used only when there is an insufficient body of information from the evidence provided by Category 1-level studies. Flawed studies can never compensate for high-quality experimental or quasi-experimental studies. However, if there are no acceptable experimental studies, the report may include brief discussion of Category 2 studies. If there is a pattern of findings across the studies—and if the design flaws that compromise the studies are dissimilar (e.g., one study has differential attrition, another compares volunteers to non-volunteers)—the report may indicate that a pattern emerges that might be considered worthy of mentioning. Studies in this category, however, are highly variable in the nature of their flaws and will be assessed case by case by two Panelists and a researcher at Abt Associates before being used for this purpose.

These two categories are studies that attempt to determine causal inference. However, panelists were free to use any type of research (descriptive, correlational, qualitative) to set the context for their meta-analysis. The reader will note that all of these types of research have been used to help explain the concepts examined in the chapter, and to help interpret findings from the experiments. However, these studies were not used to make claims of causality or effectiveness.

Procedures

Literature search and study inclusion

Literature searches were conducted to locate studies on evidence-based practices and learning in mathematics. Electronic searches were made in PsycInfo and the Social Sciences Citation Index (SSCI) using search terms identified by the Instructional Practices Task Group. A full list of the search terms used follows on page 206. A total of 1,733 studies[18] were identified based on these search terms. The identification of studies using formative assessments was based on work conducted by the Urban Institute and is described in Appendix B. Any other deviations from this general literature search procedure are specified in the report for each topic. Additional studies were identified through manual searches of relevant journals, reference lists, and recommendations from experts. Abstracts from these searches were screened for relevance to research questions and appropriate study design. For each of the 381 studies that met the screening criteria, the full study report was examined to determine whether it met the inclusion criteria specified below. Additionally, citations from relevant articles and research syntheses in each of the areas were reviewed to identify additional candidate studies.

Criteria for Inclusion:

- Study was published between 1976 and 2007.
- Study involved K–12 students studying mathematics through algebra.
- Study was available in English.
- Study was published in peer-reviewed journal or government report.
- Study design was (a) a randomized experiment or (b) a quasi-experiment with techniques to control for bias (matching, statistical control) or demonstration of initial equivalence on a salient pretest variable.
- The study included at least two classrooms per condition if the intervention was performed at the classroom level. In cases where a single teacher or investigator administered both the treatment and control, one classroom may have been sufficient if there was evidence that no bias existed.
- The intervention was not confounded with teacher, instructional time, or any other variable. Studies with potential confounds were reviewed on a case-by-case basis.
- There was no evidence of contamination (i.e., that control group teachers were using experimental curriculum or ideas from the experimental curriculum).
- Attrition was less than 30% or evidence showed that the remaining sample was equivalent to the original sample on a salient variable.

Effect size calculations

For all studies that met the criteria for inclusion, the Panel applied the WWC guidelines to calculate standardized mean differences in mathematics achievement (see http://www.whatworks.ed.gov/reviewprocess/conducted_computations.pdf). Using *Comprehensive Meta-Analysis, Version 2,* software, Hedges g standardized mean

[18] This number does not include studies that were identified from searches using combinations of terms that led to hundreds of largely irrelevant citations, studies that were identified from manual reviews, or studies that were recommended from experts.

differences were calculated for each of the studies. The standardized mean difference is defined as the difference between the mean score for the treatment group minus the mean score for the comparison group, divided by the pooled standard deviation of that outcome for both the treatment and comparison groups.

For all quasi-experiments and for randomized controlled trials that showed differences in pretest scores at baseline, the effect size measure was calculated as an adjusted mean difference as per WWC guidelines. Specifically, whenever possible, the numerator in the effect size was calculated as the difference between the posttest means of the treatment and control groups minus the difference in the pretest means for those groups, divided by the pooled unadjusted between-student standard deviation on the posttest.

In cases in which schools, teachers, or classrooms were assigned (either randomly or nonrandomly) into intervention and comparison groups and the unit of assignment was not the same as the unit of analysis, the effect size and accompanying standard error were adjusted for clustering within schools, teachers, or classrooms. This analysis used WWC guidelines to adjust for clustering,[19] applying an intraclass correlation (ICC) adjustment of 0.20 when actual ICC values were unavailable, which is the default ICC for achievement outcomes recommended by the WWC.

Pooling effect sizes across study samples

When judged appropriate, the Task Group pooled effect sizes across studies meta-analytically using random effects models. Specifically, weighted mean effect sizes were computed using inverse variance weights to reflect the statistical precision of the respective studies stemming from both the subject-level and study-level sampling error.

Multiple contrasts: For each study that included at least three conditions, effect sizes were calculated for all relevant contrasts, provided that they were orthogonal. When pooling the effects using meta-analytic techniques, only independent effect sizes per study were included, i.e., those not based on the same participant samples.

Multiple outcomes: For studies that reported effects on more than one mathematics achievement outcome, Panel reviewers decided either to choose one outcome or to average the results from multiple outcomes on a case-by-case basis. Assessments that were overly aligned with an intervention were either not used or noted when used.

Multiple independent samples within a study: In cases where impacts on independent samples within a study were reported, all independent effect sizes were included separately in the pooled analysis.

[19] See http://ies.ed.gov/ncee/wwc/pdf/rating-scheme.pdf for more information on this issue.

Study Identification Procedure for Formative Assessment

Studies that were included in the formative assessment analysis were based primarily on the literature search conducted by the Urban Institute as part of the U.S. Department of Education's Promising Practices Initiative (Olsen, 2006). The Urban Institute study inclusion criteria focused on issues of relevance, appropriate research methods, and adequate reporting of program effects (See Olsen, 2006, for further detail). The Urban Institute research staff identified relevant studies by gathering studies recommended by content experts at the U.S. Department of Education's Center on Instruction, examining reference lists, conducting database searches in Google Scholar, and searching through a dissertation database. Their efforts yielded 92 potentially relevant studies for consideration. Of these, nine studies met the criteria for inclusion in the Urban Institute's meta-analysis. Reasons for exclusion were: qualitative studies, quantitative studies with no measure of program impacts, no relevance to formative assessment, formative assessment unrelated to mathematics. In addition, studies were excluded if they provided insufficient data to calculate effect sizes. The criteria for the search were virtually identical to WWC except that standards for differential attrition, and confounding of intervention with school were not as rigorous (See Methodology report for further discussion). In the search, the keyword string included was ("mathematics" OR "math") ("formative assessment" OR "curriculum-based measurement") ("estimate" OR "coefficient" OR "correlation") ("student achievement" OR "teacher use") ("random" OR "random assignment" OR "randomly assigned" OR "matched"). In order to identify dissertations in the area of formative assessment in mathematics using Proquest's Digital Dissertation database, the keyword string included ("formative assessment" OR "curriculum based measurement" OR "ongoing assessment").

The Urban Institute put "on hold" studies that measured the effect of various enhancements to the formative assessments. The Task Group viewed the "enhancements" as important for understanding best ways for teachers and school districts to use formative assessments.

The National Mathematics Panel retrieved and reviewed all studies that had been excluded by the Urban Institute but were coded as quantitative with a comparison group. As a result, two additional studies were added to the Panel review: one that studied the effect of an enhanced formative assessment program against a control group (Calhoon & Fuchs, 2003), and another where it was possible to estimate a student-level sample size and thus calculate effect sizes and standard errors (Allinder, Bolling, Oats, & Gagnon, 2000).

In addition, after reviewing the nine original studies included by the Urban Institute, the Task Group determined that two of the papers (Spicuzza, Ysseldyke, Lemkuil, Kosciolek, Boys, & Telluchsingh, 2001; Ysseldyke, Spicuzza, Kosciolek, Teelucksingh, Boys, & Lemkuil, 2003) reported usable data based on the same study and sample. As a result, the Panel analysis includes a total of ten studies.

Search Terms Used for Instructional Practices Task Group By Research Question

All of the terms in the lists below were searched with the term *math**

Teacher-Directed and Student-Centered Instruction

active instruction
active teaching
CGI
cognitively guided instruction
constructivist
cumulative review
direct instruction
discovery learning
drill
explicit instruction
guided inquiry
guided learning
learner centered
student directed strategies
student explanations
student feedback
student reasoning
teacher centered instruction
teacher demonstration
teacher-directed instruction
teacher-directed strategies
teacher explanations
teacher feedback
teacher led instruction
teacher modeling

Additional Searches Specifically for Cooperative Learning

classwide peer tutoring
collaboration
cooperative learning
cooperative mastery learning
peer assisted learning
peer tutoring

Real World

aligning everyday and mathematical reasoning
anchored instruction
applications project
applied curricul*
applied problems
Arise
authentic
case-based
complex mathematical tasks
Connected Math
contextual curricul*
contextual problems
Core Plus
effectiveness of real world problem solving
engagement potential
everyday reasoning
Freudenthal Institute
integrated mathematics curricul*
interactive mathematics program
interactive mathematics project
Jasper
math in context
mathematical complexity
mathematical modeling
mathematical reasoning
mathematical word problems
mathematization
Middle School Math*
modeling curricul*
modeling our world
multiple solution paths
PISA
problem-based curricul*
problem-based learning
realistic math*
real-life mathematical problem solving
real world problems
SimCalc
simulations
situated cognition
solution paths
solving word problems
video

Students With Learning Disabilities, Low-Achieving Students, and English Language Learners

academically disadvantaged
anchored instruction
at-risk
classroom practices
cognitive strategy instruction
cooperative learning
curricula adaptation
differentiated instruction
direct instruction
dyscalculia
elementary education
English as a second language
English language learners
heterogeneous group
instructional design
instructional practice
intervention
learning disabil*
limited English proficient
low achiev*
math disability
math dyslexia
peer assisted learning
problem solving strategy
reform curricula
school-based intervention
secondary education
slow learners
teaching methods

Gifted Students

acceleration
developmental placement
differentiated curriculum
differentiated instruction
differentiation
enrichment
exceptional
gifted
grouping
high achiev*
talent*

Technology

artificial intelligence
CAI
calculator*
calculator-based ranger
CampOS
CBL
cellular
computer manipulatives
computer*
computer-assisted instruction
development
education
electronic blackboard
enhanced anchored instruction
graphing calculator
handheld
hypermedia
instruction
instructional tools
interactive whiteboard
interactive*
internet
learning
Logo
PDA
pedagogy
portable
programming
screen projection
screen-based technology
smart board
software
spreadsheet
teaching
technology
turtle graphics
tutor*
virtual manipulatives
visual representation
web-based

BIBLIOGRAPHY

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 & Special Education*, *21(4)*, 219–226.

Calhoon, M.B., & Fuchs, L. (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(4)*, 235–245.

National Reading Panel (2000). *Teaching children to read.* Washington, DC: U.S. Department of Education.

National Research Council (2004). *On Evaluating Curricular Effectiveness.* Washington, DC: National Academy Press.

National Research Council (2002). *Scientific research in education*. Washington, DC: National Academy Press.

National Research Council. (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, DC: National Academy Press.

Spicuzza, R., Ysseldyke, J., Lemkuil, A., Kosciolek, S., Boys, C., & Telluchsingh, E. (2001). Effects of curriculum-based monitoring on classroom instruction and math achievement. *Journal of School Psychology, 39(6)*, 521–542.

Ysseldyke, J., Spicuzza, R., Kosciolek, S., Teelucksingh, E., Boys, C., & Lemkuil, A. (2003). Using a curriculum-based instructional management system to enhance math achievement in urban schools. *Journal of Education for Students Placed at Risk*, 8(2), 247–265.

APPENDIX B: Research Questions

This appendix lists only topics that fall within the areas that we addressed in our literature review. We recognize that there is much needed research on other topics about instructional practice, such as the teaching of specific mathematical topics and content (e.g., fractions).

For future studies on teacher-directed versus student-centered instruction, our major suggestions include:

- Studies that further unpack the underlying variables behind terms such as student-centered, guided inquiry, teacher-centered, direct instruction, and explicit instruction. These studies should entail a strong classroom observational component.
- Studies that describe and evaluate the impact of: 1) various models of teaching, 2) explicit instruction for specific topics, and, 3) the use of visual representations with manipulatives to link abstract concepts (i.e. equations, algorithms).
- Studies that use curricula that are deemed to be mathematically accurate and rigorous and detail teaching practices that enhance understanding.
- Both qualitative and correlational studies of classes with exceptionally high student growth in mathematics to provide deeper insights into the nature of effective practice.

For future studies on the role of technology, our major suggestions include:

- Improved measures and analyses of fidelity of Computer Based Instruction (CBI) to best ascertain the effectiveness of interventions and to reveal "true" effect sizes that are the result of high-quality interventions. Research must also reveal what actions support high-quality large-scale implementation of these interventions.
- Studies that illuminate the particular cognitive and learning processes that different categories of software do or do not support.
- The linking of CBI features to student outcomes so that software engineers and curriculum designers can improve the use of technology in the school setting.
- The role of additional contextual variables (e.g., settings, such as urban, suburban, or rural and student or family characteristics), and implementation variables (e.g., duration, support and availability of resources, funds, and time) should be conscientiously addressed in future research.
- The initiation of longitudinal studies that will assess whether the consistent use of computer-based tools, including computer programming, can benefit learning and improve student skills.
- The implications of technology for the content of mathematics education must be adequately addressed philosophically, theoretically, and empirically.

For future studies on the role of calculators, our major suggestions include:

- The contextualization of advances in technology, curricula, and pedagogical strategies within research that examines the benefits of using calculators. This research should standardize the use of graphing calculators to address education research questions.

- An examination of whether appropriate pedagogical uses reinforce, or at least maintain, students' learning of basic arithmetical facts and properties *while simultaneously* garnering educational advantages. The fidelity of these approaches should be evaluated alongside student outcomes.

- An exploration of the cognitive processes that students use (e.g., in assessment situations) when calculators are available and the ramifications that these findings have for instruction and assessment with calculators.

- An investigation as to why about two-thirds of algebra teachers use graphing calculators infrequently. Are there practical barriers to their use, does curriculum and professional development discourage their use, and do student experiences convince teachers they are not useful? Would the provision of resources and professional development change this situation?

For future studies about instructional practices with low-achieving students and students with learning disabilities:

- Studies of the issues discussed above that focus particularly on impact with students who experience difficulty in mathematics.

- Studies to determine the amount of additional practice with feedback that these students require, the amount of highly systematic instruction needed, and the areas in which this instruction is required needs to be determined.

- Studies that examine how various approaches that are linked to specific mathematical topics are needed.

For future studies about instructional practices with gifted and mathematically precocious students:

- Evaluations of academically rigorous enrichment programs.

- Explorations of the extent to which effective enrichment programs are, in fact, acceleration programs. As students explore the mathematics that underlie their current work, the enrichment activities can develop skills in more advanced areas of mathematics, areas that the student may not cover in a formal sense for several more years.

- Longitudinal studies examining career choices and persistence in mathematics for mathematically gifted students who have participated in various intervention programs.

For future studies about the use of "real-world" problems in mathematics instruction:

- Studies to examine, describe, and clarify the multiple definitions of "real-world" problems and "real-world" problem-based instruction, and to relate those definitions to the interventions and to student learning.
- Development of valid and reliable outcome measures that clearly distinguish what is being assessed (mathematical concepts, mathematical procedures, problem solving, etc.).
- Studies that explore the possibly differential impact of "real-world" approaches to instruction for specific mathematical topics and concepts.
- Studies to examine the nature of the impact of "real-world" problem instructional approaches on student motivation and interest in mathematics, for different student groups.

APPENDIX C: Additional Technology Tables

Table C-1: Results from Prior Meta-Analyses on Drill and Practice

Study	**Pooled effect size**			
	Achievement	*Attitudes*	*Study information*	
Hamilton, 1995	.19			
Burns & Bozeman, 1981	.34			
Hartley, 1978	.34			
Lee, 1990	.35	.23		
Slavin et al., 2007 (one ES only, not "pooled")	.36			
Kuchler, 1999	.51		Junior/Senior students	
Specific education goals	*Computation*	*Concept development*	*Applications*	*Combination of goals*
Burns, 1981, cited in Kuchler, 1999	.38	.18	.14	
Lee, 1990	.45	.19 (concepts & applications)		.28
Contextual variables				
Age/grade level	*Preschool*	*Elementary*	*Middle/Junior*	*Junior/Senior*
Burns & Bozeman, 1981		.35		.24
Lee, 1990		.34	.41	
Hamilton, 1995		.17 (Grades 0–6)		.25 (Grades 9–12)
Ability level	*Low*	*Average*	*Average*	*High*
Burns & Bozeman, 1981	.31	.14	.47 (high/avg)	.32
Lee, 1990	.36	.16		.16
Hamilton, 1995	.12 (very low)	.57 (low/avg)	-.04 (average)	.27
Gender	*Males*	*Females*		
Burns & Bozeman, 1981	.42	.17		
Lee, 1990	.31	-.06		
Hamilton, 1995	.26	.14		
Implementation variables				
Duration	*1–18 weeks*	*19–36 weeks*	*37+ weeks*	
Lee, 1990	.44	.25	.46	
Substitute vs. supplement	*Substitute*	*Supplement*		
Lee, 1990	.57	.33		
Developer	*Experimenter /teacher*		*Commercial*	
Lee, 1990	.42		.34	

Table C-2: Results from Prior Meta-Analyses on Tutorials

Study	**Pooled effect size**			
Overall	*Achievement*	*Attitudes*	*Study information*	
Hamilton, 1995	.20			
Lou et al., 2001	.20			
Kuchler, 1999	.25		Secondary school	
Hartley, 1978	.34			
Kulik, 1994	.38			
Kulik, 2003	.38[b]		Study focused on ILS	
Becker, 1992	.40		Study focused on ILS	
Burns & Bozeman, 1981	.45			
Lee, 1990	.55	.02[a]		
Specific education goals	*Computation*	*Concept & Applications*		*Combination*
Lee, 1990	.42	.63		.62
Topic	*Arithmetic*	*Geometry*	*Algebra*	*General*
Lee, 1990	.47	.47	1.19	.41
Contextual variables				
Age/grade level	*Preschool*	*Elementary*	*Middle/Junior*	*Secondary*
Hartley, 1978		.66 (Grades K–8)		.58 (Grades 9–12)
Lee, 1990		.49	.85	
Hamilton, 1995		.14 (Grades 0–6)		.29 (Grades 9–12)
Burns & Bozeman, 1981		.43		.52
Ability level	*Low*	*Average*	*Average*	*High*
Burns & Bozeman, 1981	.57	.58		.28
Lee, 1990	.62	.20		.18
Hamilton, 1995	.12	.08 (low/avg)	.02	.57
Gender	*Males*	*Females*		
Lee, 1990	.82 [a]	1.58 [a]		
Hamilton, 1995	.58 [c]	.14 [c]		
Implementation variables				
Duration				
	1–18 weeks	*19–36 weeks*	*37+ weeks*	
Lee, 1990	.55	.53	.57	
Substitute vs. supplement	*Substitute*	*Supplement*		
Lee, 1990 (achievement)	.30	.58		
Lee, 1990 (problem solving)	.09	.25		
Developer	*Experimenter / teacher*	*Commercial*	*Both*	
Lee, 1990	.62	.39	.58	
Audience	*Specific*	*General*		
Lee, 1990	.58	.29		

[a] Only two effect sizes were included.
[b] Pooled effect size was .40 when the ILS instruction was in mathematics only, and .17 when it was in both mathematics and reading.
[c] Only three effect sizes were included.

Table C-3: Results from Prior Meta-Analyses on Calculators

Study	**Pooled effect size**			
	Achievement	*Attitudes*	*Study information*	
Ellington, 2003a[a]		0.20[b]		
Hembree, 1984; Hembree & Dessart, 1986		0.190		
Hembree, 1992	0.29			
Smith, 1997		0.3655		
	Calculator type			
	Basic/scientific	*Graphing*	*All*	
Ellington, 2003a (operational skills, testing with calculators)	0.55	0.40	0.25	
Ellington, 2003a (conceptual skills, testing with calculators)	0.13	0.69		
Ellington, 2003a (problem solving skills, testing with calculators)	0.23	0.61[3]		
	Specific education goals			
	Operational	*Computational*	*Conceptual*	*Problem solving*
Ellington, 2003a[a] (testing without calculators)	0.14[b]	-0.02[b]	-0.05[b]	0.16
Ellington, 2003a, [a] (testing with calculators)	0.32[b]	0.41[b]	0.44[b]	0.22[b]
Hembree, 1984; Hembree & Dessart (1986) (testing with calculators)		0.636		
Hembree, 1984; Hembree & Dessart, 1986 (testing without calculators)			0.018	
Hembree, 1984 (special calculator instruction, testing without calculators)	0.798	0.564	-0.268	0.534
Smith, 1997		0.2054	0.1972	0.1468
Smith, 1997 (Graphing calculators, graphing skills)	-0.523		(concept development)	
Contextual variables				
Age/grade level	*Preschool*	*Elementary*	*Middle/Junior*	*Secondary*
Ellington, 2003)[a] (conceptual skills, testing without calculators)		-0.06	0.52[2]	-0.15[2]
Ellington, 2003a[a] (operational skills, testing with calculators)		0.48[1]	0.57	0.32
Ellington, 2003a[a] (conceptual skills, testing with calculators)		-0.14[2]	0.70	0.43

Continued on p. 6-218

Table C-3, continued

Study	**Pooled effect size**			
Ability level	*Low*	*Average*	*High*	*Mixed*
Ellington, 2003a[a] (operational skills, testing with calculators)			0.69[e]	0.35
Ellington, 2003a[a] (conceptual skills, testing with calculators)			0.84	0.29
Ellington, 2003a[a] (problem solving skills, testing with calculators)	-0.18[c]		0.15[c]	0.43
Hembree, 1984; Hembree & Dessart, 1986 (operational skills, testing without calculators)	-0.107		-0.031	
Hembree, 1984 [Hembree & Dessart, 1986 (operational skills, testing with calculators)]	0.325	0.737		
Hembree, 1984; Hembree & Dessart, 1986 (computation skills, testing without calculators)	-0.009		-0.024	
Hembree, 1984; Hembree & Dessart, 1986 (problem solving composite skills, testing without calculators)	0.005		-0.118	
Hembree, 1984; Hembree & Dessart, 1986 (problem solving composite skills, testing with calculators)	0.436	0.271	0.458	
Implementation variables				
Duration	*0–3 weeks*	*4–8 weeks*	*9+ weeks*	
Ellington, 2003a[a] (operational skills, testing without calculators)	0.31	-0.17[e]	0.24	
Ellington, 2003a[a] (computational skills, testing without calculators)	0.14[e]	-0.25[e]	0.06	
Ellington, 2003a [a] (conceptual skills, testing without calculators)	0.26[d]	-0.29[d]	0.08[e]	
Ellington, 2003a[a] (operational skills, testing with calculators)	0.47	0.34	0.49	
	.285	-.21	.16	

[a] Included both graphing and scientific calculators.
[b] Outliers removed.
[c] Only one study.
[d] One two studies.
[e] Only three studies.

Table C-4: Results from Prior Meta-Analyses on Graphing Calculators Only

Study	Pooled effect size		
	Achievement	*Attitudes*	*Study information*
Khoju, Jaciw, & Miller, 2005	.85		One study was of college students
Ellington, 2006	0.19	0.21[a]	Calculators not allowed in testing
	0.29[a]		Calculators allowed in testing
	Specific education goals		
	Procedural	*Conceptual*	*Combined skills*
Ellington, 2006 (testing without calculators)	–0.21	0.29[a]	0.19
Ellington, 2006 (testing with calculators)	0.32[a]	0.42[a]	0.29[a]

[a] Outliers removed.

Table C-5: Results from Prior Meta-Analyses on Programming

Study	**Pooled effect size**			
Overall	*Achievement*	*Problem solving*	*Attitudes*	*Study information*
Kulik, 1994	.09			All *but* Logo
Lou et al., 2001	.22			
Gordon, 1992	.26	.34		
Kuchler, 1999	.35			Secondary
Lee, 1990	.36	.23	.29	
Khalili, 1994	.45			Logo
Kulik, 1994	.58			Logo
Specific education goals	*Computation*	*Concept & Applications*		*Combination of goals*
Lee, 1990	-.04[a]	.56		.15
Topic	*Arithmetic*	*Geometry*	*Algebra*	*General*
	.40[b]	.68	-.02	.29
Contextual variables				
Age/grade level	*Pre-school*	*Elementary*	*Middle/Junior*	
Lee, 1990 (achievement)		.76	.09	
Lee, 1990 (problem solving)		.29	.27	
Ability level	*Low*	*Middle*	*High*	
Lee, 1990 (achievement)	.22	.11	.37	
Lee, 1990 (problem solving)	.20	-.02	-0.4	
SES	*Low*	*Average*	*High*	
Lee, 1990	.10	.33	.19	
Implementation variables				
Duration	*1–18 weeks*	*19–36 weeks*	*37+ weeks*	
Lee, 1990	.45	.30	.03	
Substitute vs. supplement	*Substitute*	*Supplement*		
Lee, 1990 (achievement)	.40	.34		
Lee, 1990 (problem solving)	.08	.43		
Specific languages	*Logo*	*BASIC*	*Scientific languages*	*Other*
Kuchler, 1999[c]	.78	.34	.42	.47
Khalili, 1994	.45			.33
Lee, 1990	.41	.48		-.15
Kulik, 1994	.58			.09

[a] Only three effect sizes.
[b] Only two effect sizes.
[c] Secondary students.

Table C-6: Results from Prior Meta-Analyses on Tools and Problem Solving Environments

Study	**Pooled effect size**		
Overall	*Achievement*	*Attitudes*	*Study information*
Lou et al., 2001	.04		Tool and exploratory environments
Kulik & Kulik, 1991	.10		Computer-enhanced instruction
Kuchler, 1999	.24		Problem solving software

Table C-7: Results from Prior Meta-Analyses on Simulation and Games

Study	**Pooled effect size**			
Overall	*Achievement*	*Attitudes*	*Study information*	
Kulik, 1994	.10			
Kuchler, 1999	.23		Secondary school	
Lee, 1990	.28	.24[a]		
Specific education goals	*Computation*	*Concept & Applications*		*Combination of goals*
Lee, 1990	.61[b]	.24		.63[b]
Topic	*Arithmetic*	*Geometry*	*Algebra*	*General*
Lee, 1990	.61[b]	.24	.15[a]	1.12[a]
Contextual variables				
Age/grade level	*Elementary*	*Junior*		
Lee, 1990	.24	.45		
Gender	*Males*	*Females*		
Lee, 1990	.31	.12		
Implementation variables				
Duration	*1–18 weeks*	*19–36 weeks*		
Lee, 1990	.33	.14		
Substitute vs. supplement	*Substitute*	*Supplement*		
Lee, 1990 (achievement)	.18	.39		
Lee, 1990 (problem solving)	-.83			
Developer	*Experimenter / teacher*	*Commercial*		
Lee, 1990	.26	.29		

[a] Only one effect size included.
[b] Only two effect sizes included.

Table C-8: Subgroup Analysis for Calculator Studies

	Computation				Applications				Concepts		
	N studies/ ES	**Hedges g**		**se**	**N studies/ ES**	**Hedges g**		**se**	**N studies/ ES**	**Hedges g**	**se**
Contextual variables											
Grade level											
Elementary	4 / 5	0.367		0.330	2 / 3	0.074		0.371	3 / 3	0.267	0.236
Secondary	3 / 4	0.113		0.169	3 / 4	0.437	**	0.164	1 / 1	0.328	0.489
Mixed	1 / 1	0.855	~	0.512	0 / 0	na		na	0 / 0	na	na
Implementation variables											
Duration		*									
Less than 3 months	5 / 7	0.503	*	0.218	3 / 5	0.295		0.239	2 / 2	0.084	0.307
3 months or greater	3 / 3	-0.134		0.198	2 / 2	0.328		0.249	2 / 2	0.458	0.295

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

Table C-9: Calculator Effect Sizes not Included in Meta-Analytic Tables

Study	Grade Level	Contrast	Measure	Hedge's g	Standard Error	
Assessments in which calculator group was able to use calculator						
Szetela, 1982	Grade 3	Regular instruction plus calculator-specific materials vs. Regular instructional activities	*Problem-solving*: Posttest 2: 20 researcher-designed items	0.568	~	0.297
Szetela, 1982	Grade 5	Regular instruction plus calculator-specific materials vs. Regular instructional activities	*Problem-solving*: Posttest 2: 20 researcher-designed items	-0.29		0.342
Szetela, 1982	Grade 7	Regular instruction plus calculator-specific materials vs. Regular instructional activities	*Problem-solving*: Posttest 2: 20 researcher-designed items	0.633	*	0.294
Szetela, 1982	Grade 8	Regular instruction plus calculator-specific materials vs. Regular instructional activities	*Problem-solving*: Posttest 2: 20 researcher-designed items	0.589	*	0.274
Alternate interventions/enhancements						
Standifer & Maples, 1981	Grade 3	Programmed feedback calculator vs. No calculator in regular math curriculum	*Computation*: Science Research Associates	0.296		0.395
Standifer & Maples, 1981	Grade 3	Programmed feedback calculator vs. No calculator in regular math curriculum	*Concepts*: Science Research Associates	-0.28		0.395
Standifer & Maples, 1982	Grades 3&4	Programmed feedback calculator vs. General remedial math curriculum	*Computation*: Science Research Associates	0.363		0.326
Standifer & Maples, 1982	Grades 3&4	Programmed feedback calculator vs. General remedial math curriculum	*Concepts*: Science Research Associates	0.064		0.325
Duffy & Thompson, 1980	Grade 4	Calculator plus materials vs. No calculator (classroom-level effect size)	*Computation*: CTBS	-0.01		0.428
Duffy & Thompson, 1980	Grade 4	Calculator plus materials vs. No calculator (classroom-level effect size)	*Problem-solving*: CTBS Applications	-0.15		0.429

Continued on p. 6-223

Table C-9, continued

Study	Grade Level	Contrast	Measure	Hedge's g	Standard Error
Duffy & Thompson, 1980	Grade 5	Calculator plus materials vs. No calculator (classroom-level effect size)	*Concepts*: CTBS	-0.27	0.430
Duffy & Thompson, 1980	Grade 5	Calculator plus materials vs. No calculator (classroom-level effect size)	*Computation*: CTBS	0.222	0.450
Duffy & Thompson, 1980	Grade 5	Calculator plus materials vs. No calculator (classroom-level effect size)	*Problem-solving*: CTBS Applications	-0.090	0.449
Duffy & Thompson, 1980	Grade 6	Calculator plus materials vs. No calculator (classroom-level effect size)	*Concepts*: CTBS	-0.09	0.449
Duffy & Thompson, 1980	Grade 6	Calculator plus materials vs. No calculator (classroom-level effect size)	*Computation*: CTBS	-0.22	0.440
Duffy & Thompson, 1980	Grade 6	Calculator plus materials vs. No calculator (classroom-level effect size)	*Problem-solving*: CTBS Applications	0.191	0.440
Duffy & Thompson 1980	Grade 6	Calculator plus materials vs. No calculator (classroom-level effect size)	*Concepts*: CTBS	-0.21	0.440
Total math achievement (calculator vs. control)					
Standifer & Maples, 1981	Grade 3	Hand-held, four function calculator vs. No calculator in regular math curriculum	*Total achievement*: Science Research Associates	0.309	0.396
Standifer & Maples, 1982	Grades 3&4	Hand-held, four function calculator vs. General remedial math curriculum	*Total achievement*: Science Research Associates	0.341	0.330
Duffy & Thompson, 1980	Grade 4	Calculator only group vs. No calculator (classroom-level effect size)	*Total achievement*: CTBS	0.062	0.428
Duffy & Thompson, 1980	Grade 5	Calculator only group vs. No calculator (classroom-level effect size)	*Total achievement*: CTBS	0.008	0.449
Duffy & Thompson, 1980	Grade 6	Calculator only group vs. No calculator (classroom-level effect size)	*Total achievement*: CTBS	-0.113	0.439
Total math achievement (calculator + enhancement vs. control)					
Standifer & Maples, 1981	Grade 3	Programmed feedback calculator vs. No calculator in regular math curriculum	*Total achievement*: Science Research Associates	-0.047	0.394
Standifer & Maples, 1982	Grades 3&4	Programmed feedback calculator vs. General remedial math curriculum	*Total achievement*: Science Research Associates	0.194	0.325
Duffy & Thompson, 1980	Grade 4	Calculator plus materials vs. No calculator (classroom-level effect size)	*Total achievement*: CTBS	-0.089	0.429
Duffy & Thompson, 1980	Grade 5	Calculator plus materials vs. No calculator (classroom-level effect size)	*Total achievement*: CTBS	0.207	0.450
Duffy & Thompson, 1980	Grade 6	Calculator plus materials vs. No calculator (classroom-level effect size)	*Total achievement*: CTBS	0.021	0.428

Table C-10: Computer Programming Effect Sizes (Comparing Programming to CAI) not Included in Meta-Analytic Tables

Study	Design	Sample	Duration/ Content	Contrast	Measure	Hedge's g	Standard Error
Programming							
Battista & Clements, 1986	RCT	11 fourth-graders in two midwestern middle schools	42 sessions (2 40-min per week)/LOGO	Gr 4: Logo vs. CAI	Problem-solving Combine Test 1&2, Total	0.409	0.611
		26 sixth-graders in two midwestern middle schools		Gr 6: Logo vs. CAI		0.326	0.395
Clements, 1986	RCT	24 first-grade students from a middle-class midwestern school system	44 sessions (22 weeks)/ LOGO	Gr 1: Logo vs. CAI	WRAT Math score	0.486	0.415
		24 third-grade students from a middle-class midwestern school system		Gr 3: Logo vs. CAI		0.473	0.414
Clements, 1987	RCT	16 third-grade students who had received Logo or CAI experience in first-grade	3 months/ LOGO	Gr 3: Logo vs. CAI	CAT - Total	0.452	0.511
Emihovich & Miller, 1988	RCT	36 first-grade students in five classrooms in an elementary school in the southeast	20, 30-min sessions (3 months)/ LOGO	Gr 1: Logo vs. CAI	CTBS - Math	0.214	0.410

~ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$
[a] Data were adjusted for clustering that occurred within classrooms.