Using Contrasting Examples to Support Procedural Flexibility and Conceptual Understanding in Mathematics
Bethany Rittle-Johnson and Jon Star

RESEARCH NARRATIVE

Current educational reforms in mathematics advocate that the teacher act more as a facilitator, encouraging students to explain their own thinking and problem-solving procedures to other students (e.g., Hiebert & Carpenter, 1992; National Council of Teachers of Mathematics, 1989, 2000). Despite the intuitive appeal of such educational approaches, the psychological literature is inconclusive about the benefits of this type of reform pedagogy. In this proposal, we rigorously evaluate a potentially pivotal component of this instructional approach that is supported by basic research in cognitive science: the value of students explaining contrasting examples. We evaluate this approach within domains of critical need for improvement in United States middle-school mathematics instruction: computational estimation and equation solving. This research will be conducted in intact classrooms, initially using a controlled, experimental paradigm within the classrooms and later under more typical classroom conditions.

Contribution of Project to Solving an Educational Problem

Current reform efforts in U. S. education are motivated by endemic problems with students gaining only inert knowledge—rigid knowledge that is not accessed or transferred to solve novel problems (National Research Council, 2000). Inert knowledge often results when students memorize information rather than learn for understanding. Judd’s water darts experiment is a classic example of the dangers of learning without understanding: Boys who learned to throw darts underwater through trial-and-error did not transfer this knowledge to hitting a target at a new water depth. In contrast, boys who were taught the relevant principle of light refraction were able to transfer their knowledge to a new water depth (Judd, 1908).

Mathematics is one of the most critical domains for overcoming the inert knowledge problem. Both international and national assessments indicate that U.S. students may learn to execute rote procedures, but they often fail to gain robust, flexible knowledge. In the Third International Mathematics and Science Study (TIMSS), U.S. students were above the international average in fourth grade, but by eighth grade had fallen to below average and by 12th grade were at the bottom of the rankings (Beaton et al., 1996). As the results from this and other assessments indicate, although early mathematics instruction may be at least adequate, mathematics instruction beginning in the middle-school years is in critical need of improvement (National Research Council, 2001).

Large-scale assessments, as well as more focused studies, indicate that the largest gap in students’ mathematical abilities is in their ability to flexibly solve novel problems (Hiebert & Carpenter, 1992; Schmidt, McKnight, Cogan, Jakwerth, & Houang, 1999). Flexible problem solving requires that students integrate their conceptual knowledge of principles in the domain with their procedural knowledge of specific actions for solving problems (National Research Council, 2001). This integrated knowledge allows students to identify prior knowledge that is relevant—including potentially relevant procedures—and to adapt existing procedures to solve novel problems. Thus, integrated knowledge is not inert—rather, it can be used to flexibly solve both routine and novel problems accurately and efficiently.
Such integrated, flexible knowledge is a key characteristic of expertise (Larkin, McDermott, Simon, & Simon, 1980; National Research Council, 2000).

This goal of integrated and flexible knowledge is a holy grail of cognitive science and education—everyone agrees that it exists, but no one is sure how to reliably develop it. Fortunately, research in cognitive science offers powerful suggestions for how teachers and schools might inculcate this kind of productive knowledge. As there are numerous possible cognitive science findings to pursue, we used the lens of current “best practices” in mathematics education to identify a potentially powerful teaching approach that arises from both basic cognitive science research and mathematics education best practices—namely, having students explain contrasting examples (Gentner & Loewenstein, 2002; Lampert & Cobb, 2003; Schwartz & Bransford, 1998). For example, students could be asked to share solution procedures for a particular problem and then to discuss the similarities and differences in the different procedures. When presented with contrasting examples of solution procedures and then asked to compare them, learners should come to notice critical features of the procedures and to abstract their common underlying structure. Such analogical encodings support flexible and transferable knowledge (Gentner & Loewenstein, 2002).

This kind of instructional approach lies at the core of current reform pedagogy in mathematics. At all grade levels, teachers are encouraged to create an environment where students can engage in thinking and communicating deeply about mathematics—in discussing, collaborating, justifying, conjecturing, experimenting, and responding to the ideas of their peers (National Council of Teachers of Mathematics, 2000, p. 18). The development of mathematical understanding is believed to be enhanced by classroom discussions, where students share procedures and evaluate the procedures of others (Lampert, 1986, 1992a, 1992b), including informal and non-standard algorithms (Carroll, 2000; Mack, 1990). In fact, a key distinction between novice and expert teachers is whether and how they orchestrate such a discussion of procedures—is it a simple sharing of alternative methods, or is this initial discussion followed by an in-depth comparison of the procedures and the mathematics underlying the similarities and differences between different approaches (Gravemeijer & van Galen, 2003; Lampert & Cobb, 2003)? This difference in teaching approach is not simply a difference of expertise. Rather, it is also part of a fundamental debate in mathematics education between a more laissez-faire, non-directive teaching approach advocated by radical constructivists (e.g., Steffe & Kieren, 1994; von Glasersfeld, 1995, 1996) and a more directive and guided approach to teaching advocated by more moderate constructivists (e.g., Cobb, 1994; Simon, 1995). Despite the increasingly wide adoption of reform pedagogy and its intuitive appeal, most research evaluating the effectiveness of this approach has been descriptive, with few controlled empirical evaluation studies. Yet there are at least two lines of research in cognitive science that converge on the importance of having students explain contrasting examples as a pathway to flexible knowledge: research on 1) analogical encoding with contrasting cases, and 2) learning from worked examples and self-explanations. Each of these lines of research will be discussed below.

In the proposed set of studies, we will compare learning from contrasting two solutions to the same problem with learning from sequential consideration of individual procedures. Thus, we will evaluate the benefit of contrasting alternative procedures over and above mere exposure to multiple procedures (which is the more common educational approach if alternative procedures are discussed at all). The first set of studies (Studies 1 and 2) will be conducted in intact classrooms over several days. The experimental manipulation
will occur at the level of pairs of students; the classroom teacher and the research staff will be involved in the implementation of the intervention. This will allow for random assignment of pairs to condition, and for tight control of the activities students are engaged in. In the later studies (Studies 3, 4, and 5), the manipulation will occur at the classroom level, and regular classroom teachers will implement the intervention.

Research Plan

Supporting Literature

The proposed studies will evaluate a promising instructional technique that has emerged from best practices in mathematics education as well as research in cognitive science: explaining contrasting cases. In this section, we discuss findings from cognitive science—(a) research on analogical encoding with contrasting cases, and (b) learning from worked examples and self-explanations—that support the potential of this approach to promote the development of procedural flexibility and conceptual understanding.

**Analogical encoding with contrasting cases.** A key feature of expertise is selective attention to important information in problems (Gentner, 1983; Larkin, 1989; National Research Council, 2000; Novick, 1988). People do not simply “absorb” what they see in the world; rather, their prior knowledge and the current context influences what information is encoded and used (Garner, 1974; Siegler, 1976). As a novice in a domain, it is very difficult to know what information is important and how to use that information, especially with a single example. For example, consider what one may notice when shown a single phrase—“Mathematics is fun”—as compared to when this phrase is paired with a second phrase, “Mathematics is fun.” When shown only the single phrase, one probably does not notice the font size and style, but the features become very noticeable when the contrasting phrases are presented as a pair. Perceptual learning theorists such as Garner (1974) and Gibson and Gibson (1955) have long emphasized the importance of contrasting cases for noticing what is important. Recently, two separate lines of cognitive science research have expanded this idea into the learning of cognitive tasks.

Working with contrasting cases facilitates deep and meaningful abstraction of information from problems, and prepares students for future learning (Bransford, Franks, Vye, & Sherwood, 1989; Bransford & Schwartz, 2001; Schwartz & Bransford, 1998). For example, college students who actively compared simplified data sets on memory experiments before hearing a lecture transferred their knowledge much more readily than students who read and summarized a text before hearing the lecture (Schwartz & Bransford, 1998). However, students who compared the data sets a second time rather than hearing the lecture did not learn more than the summarize + lecture group. Schwartz and Bransford (1998) emphasize that working with contrasting cases guides students’ attention to important features and thus prepares them to learn from an organizing lesson. It also highlights that there is a time for telling—that direct instruction under certain conditions can be more effective than continued discovery learning.

Gentner and her colleagues also focus on comparison of examples as a critical pathway to flexible, transferable knowledge. Analogical encoding, or the abstraction of common underlying structure from multiple examples, requires the use of contrasting cases (Gentner, Loewenstein, & Thompson, 2003; Loewenstein, Thompson, & Gentner, 1999). For example, college students studied two cases illustrating a contract-negotiating strategy. Those who were prompted to compare the two cases by reflecting on the similarities were much
more likely to transfer the negotiation strategy to a new case than were students who read and reflected on the cases independently. Evidence on the quality of students’ responses during the learning phase suggested that comparing cases led to higher understanding of key principles than studying the cases independently. Young children derive similar benefits when they are helped to compare two examples in a spatial task and a word-learning task (Gentner & Namy, 2000; Loewenstein & Gentner, 2001; Namy & Gentner, 2002).

The evidence from research in cognitive science strongly suggests that comparing and contrasting multiple examples helps avoid the development of inert knowledge. However, this research has largely focused on learning and transferring concepts and ideas, and has not examined the impact of contrasting cases on learning problem-solving procedures. It also has not been conducted within mathematics or in K-12 classrooms. Will contrasting cases promote procedural flexibility and conceptual understanding in mathematics? In typical classrooms? Descriptive evidence from studies in mathematics education suggests that they will. Expert mathematics teachers engage students in discussions of similarities across alternative solution methods (Ball, 1993; Lampert, 1990). In the current studies, we will rigorously evaluate whether comparing contrasting examples promotes greater flexibility and understanding of mathematics than studying examples one at a time.

**Worked examples and self-explanations.** Research on worked examples provides additional support for having students explain contrasting examples, as well as additional guidance in how to implement a contrasting cases approach in mathematics. Using worked examples to develop novices’ ability to solve problems has a relatively long history in psychological and educational research (see Atkinson, Derry, Renkl, & Wortham, 2000, for a review). Worked examples typically include a problem statement and a procedure for solving the problem. This worked example tradition grew out of research in cognitive psychology on schemas, domain-specific knowledge, automaticity, and expert-novice differences (Sweller, 1988). Repeated studies have shown that students from elementary school to university—both in the laboratory and in the classroom—learn more efficiently and deeply if they study worked examples paired with practice problems rather than solve the equivalent problems without guidance (Carroll, 1994; Mawer & Sweller, 1982; Sweller & Cooper, 1985; Ward & Sweller, 1990; Zhu & Simon, 1987). Further, students who used a geometry and algebra curriculum rich in worked examples learned the same content in two-thirds the time of a traditional curriculum, with equivalent success on a standard year-end mathematics assessment (Zhu & Simon, 1987).

The use of worked examples may be viewed by some as an example of traditional, direct instruction and thus contradictory to aspects of mathematics education reform that advocate for student invention of procedures (Carpenter, Franke, Jacobs, Fennema, & Empson, 1998; Hiebert & Carpenter, 1992; Kamii & Dominick, 1998; National Council of Teachers of Mathematics, 1989). In fact, the use of worked examples differs significantly from traditional teacher-led instruction of procedures. Whereas traditional classroom pedagogy typically involves the teacher demonstrating (and students memorizing) a single way to solve problems, worked examples (particularly when presented as sample student solutions, as proposed here) guarantee efficient exposure to multiple ways to solve problems while avoiding implied judgments that there is only one way to solve a problem. Indeed, think-aloud data from students using worked-out examples suggests that they do not just memorize the examples; rather, students develop knowledge of when the procedures are applicable and link to related concepts explaining the procedures (Zhu & Simon, 1987).
Overall, evidence suggests that it is students’ active cognitive processing, not necessarily the source of the procedures (e.g., teacher demonstration, worked-out example, or student invention), that seems most important for learning (Carpenter et al., 1998; Hiebert & Wearne, 1996; Kamii & Dominick, 1998; Klahr & Carver, 1988; Klahr & Nigam, 2004; Mayer, 2004; Schwartz & Bransford, 1998).

Although worked examples are generally effective for improving learning, there are individual differences in learning from them. Think-aloud protocol studies of college students studying worked examples revealed that good learners generated their own explanations (self-explanations) to justify steps in the worked example, whereas poorer learners did not (Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Renkl, 1997a). Subsequent studies using random assignment to condition indicate that prompting students to generate self-explanations when studying worked examples leads to greater learning, as compared to no prompts (Bielaczyc, Pirolli, & Brown, 1995; Chi, DeLeeuw, Chiu, & LaVancher, 1994; Renkl, Stark, Gruber, & Mandl, 1998; Siegler, 1995). However, even with prompting, individual differences in the quality of self-explanations sometimes persist, and these differences predict the amount of learning. Prompting students to explain contrasting examples is one promising method for improving the quality of self-explanations. Analogical encoding of examples prompted by contrasting cases should lead to deeper explanations that focus on critical problem features and structural invariants. In the current studies, we evaluate whether prompts to explain contrasting examples improve students’ explanation quality and subsequent learning, compared to explaining an example in isolation.

An additional method for improving the quality of students’ self-explanations may be to include a co-learner. Learners who collaborate with a partner tend to learn more than those who work alone, especially when the partner brings different prior knowledge to the collaboration (e.g., Johnson & Johnson, 1994; Teasley, 1995; Webb, 1989, 1991; Weinstein & Bearison, 1985). When talking with another person, people are more explicit, more interpretive, and use more qualifiers (Krauss, 1987; Loewenthal, 1967; Teasley, 1995), suggesting that the social context of collaboration should improve the quality of self-explanations. Contrary to these general findings, some recent research has found that having a student assume the role of a teacher and explain worked examples to another student may not improve the quality of the self-explanations or the amount of learning, perhaps because of the stress and added cognitive load induced by having to teach another person (Renkl, 1995, 1996, 1997b, 1998). However, Renkl’s work also suggests that the benefits of collaboration are more likely to be realized when partners are of equal status rather than as teacher-student dyads. Finally, working with a partner may be particularly important for eliciting explanations in a classroom setting where students are unlikely to talk aloud to themselves (and are unlikely to write as much as they would verbalize). In the proposed work, participants work together as equal-status partners in order to encourage the generation of high-quality explanations.

Extending research on self-explaining—which has primarily been conducted in lab settings with students working individually—to typical classroom conditions of talking with peers is an important step in bridging cognitive science laboratory research and classroom practice. It is also important in light of current educational reform efforts that focus on students learning by sharing problem-solving procedures and listening and commenting on other students’ explanations (e.g., National Council of Teachers of Mathematics, 1989, 2000). Despite the clear potential benefits of self-explanations to student learning, the psychological
literature is unclear as to whether this instructional approach advocated by recent reforms is reliably effective under more typical classroom conditions (Ploetzner, Dillenbourg, Praier, & Traum, 1999). Variability in the quality of explanations may be a critical factor underlying the contradictory findings on the effectiveness of generating explanations for peers. We hypothesize that having students contrast different solution procedures will lead to deeper explanations and greater learning than will explaining the solution procedures in isolation. We will evaluate this prediction under typical classroom conditions of working with a partner and participating in teacher-facilitated whole-class discussions.

**Mathematical Domains**

To investigate the effectiveness of the contrasting cases approach, as well as its generality, we will evaluate this instructional intervention in two domains within middle-school mathematics: with seventh graders learning about algebra equation solving, and with fifth graders learning about mental math and computational estimation. Both of these topics are pivotal components of mathematical competence, and both are topics students struggle to learn.

First, consider algebra equation solving. The transition from arithmetic to algebra is a notoriously difficult one, and improvements in algebra instruction are greatly needed (National Research Council, 2001). Algebra historically has represented students’ first sustained exposure to the abstraction and symbolism that makes mathematics powerful (Kieran, 1992); its symbolic procedures enable students to consider relationships, variable quantities, and situations in which change occurs (Fey, 1990). In addition to its central role in the discipline of mathematics, algebra also serves as a critical “gatekeeper” course, in that earning a passing grade has become a de facto requirement for many educational and workplace opportunities. Some have gone so far as to refer to algebra as the new civil right (Moses, 1993). Regrettably, students’ difficulties in algebra have been well documented in national and international assessments (e.g., Beaton et al., 1996; Blume & Heckman, 1997; Lindquist, 1989; Schmidt et al., 1999). For example, data from the NAEP indicates that many 12th graders can solve only the most simple and routine algebra tasks (Blume & Heckman, 1997).

Equation solving is considered a “basic skill” by many in mathematics education (Ballheim, 1999), and this domain has been widely studied in both the mathematics education (e.g., Adi, 1978; Kieran, 1989, 1990; Whitman, 1975) and cognitive science communities (Bernard & Cohen, 1988; Bundy & Silver, 1981; Lewis, 1981; Matz, 1980, 1982; Sleeman, 1982, 1984). When introduced, the procedures used to solve equations are among the longest and most complex to which students have been exposed. Current mathematics curricula typically focus on standard procedures for solving equations, rather than on flexible and meaningful solving of equations (Kieran, 1992). In contrast, prompting students to solve problems in multiple ways leads them to greater procedural flexibility (Star, 2001; Star & Rittle-Johnson, 2004). Table 1 shows examples of different procedures that students generate to solve equations (Star, 2001).
Table 1: Procedures used by students to solve the equation $3(x + 1) = 15$

<table>
<thead>
<tr>
<th>Procedure 1</th>
<th>Procedure 2</th>
<th>Procedure 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$3(x + 1) = 15$</td>
<td>$3(x + 1) = 15$</td>
<td>$3(x + 1) = 15$</td>
</tr>
<tr>
<td>$3x + 3 = 15$</td>
<td>$3x + 3 = 15$</td>
<td>$x + 1 = 5$</td>
</tr>
<tr>
<td>$3x = 12$</td>
<td>$x + 1 = 5$</td>
<td>$x = 4$</td>
</tr>
<tr>
<td>$x = 4$</td>
<td>$x = 4$</td>
<td></td>
</tr>
</tbody>
</table>

Next, consider mental math and computational estimation. A large majority of students have difficulty doing simple calculations in their heads or estimating the answers to problems (e.g., Case & Sowder, 1990; Hope & Sherrill, 1987; Reys, Bestgen, Rybolt, & Wyatt, 1980; Sowder, 1992). For example, on the third National Assessment of Educational Progress in Mathematics (NAEP), less than half of the 13-year-olds tested could mentally calculate 60 x 70, and only 38% thought that 945 x 1000 should be solved mentally (See also Blume & Heckman, 1997; Lindquist, 1989; Lindquist, Carpenter, Silver, & Matthews, 1983).

This disuse or inability to use mental math or estimation is a significant barrier to using mathematics in everyday life. We often must make quick computations or judgments of numerical magnitude without the aid of calculator or paper and pencil. In addition to being a fundamental, real-world skill, the ability to quickly and accurately perform mental computations and estimations has two additional benefits: 1) It allows students to check the reasonableness of their answers found through other means, and 2) it can help students develop a better understanding of place value, mathematical operations, and general number sense (Beishuizen, van Putten, & van Mulken, 1997; National Research Council, 2001). These benefits are encapsulated in the recent “Adding It Up” report from the National Research Council: “The curriculum should provide opportunities for students to develop and use techniques for mental arithmetic and estimation as a means of promoting deeper number sense” (2001, p. 415).

While the benefits of mental math are increasingly recognized in mathematics education, most mathematics curricula address the topic by having students focus only on relatively low-level mental math procedures such as rounding. Such approaches are unlikely to benefit people’s need for flexible procedures and greater understanding of our number system (Sowder, 1992). In order to develop this type of knowledge, students require experiences that use multiple different procedures in a variety of settings. Table 2 illustrates a variety of solution procedures that students have been found to use in estimating answers to multiplication problems (LeFevre, Greenham, & Waheed, 1993; Reys et al., 1980).

Table 2: Sample procedures students use to estimate 78 x 289

<table>
<thead>
<tr>
<th>Procedure</th>
<th>What student might do in his/her head</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rounding both numbers</td>
<td>80 x 300</td>
</tr>
<tr>
<td>Front-end multiplying</td>
<td>7 x 2 = 14, so 14,000</td>
</tr>
<tr>
<td>Prior compensation</td>
<td>75 x 300</td>
</tr>
<tr>
<td>(e.g. round one number up</td>
<td></td>
</tr>
<tr>
<td>and the other down)</td>
<td></td>
</tr>
<tr>
<td>Post compensation</td>
<td>80 x 300 – (2 x 300)</td>
</tr>
</tbody>
</table>
This brief review of equation solving and computational estimation clearly illustrates that these are core mathematical domains (advocated for in national documents and covered in most mathematics books), topics that students struggle to master, and topics where procedural flexibility and conceptual understanding are particularly important. There is a critical need for innovations that both improve learning in these areas and have implications for improving mathematics instruction more broadly. The current studies evaluate one potential innovation that addresses this need.

**Overview of Intervention**

Across five studies, we will compare learning from contrasting examples (compare group) to learning from sequentially presented examples (sequential, or control, group). In Studies 1 and 2, pairs of students will be randomly assigned to condition and the manipulation will occur while student pairs study worked examples and solve practice problems in their classrooms. In Studies 3, 4, and 5, classrooms will be randomly assigned to condition, and the manipulation will occur both in partner activities and in whole-class discussions led by the regular classroom teachers. This expansion to include teacher-led discussion is critical for at least three reasons: 1) individual pairs of students may fail to recognize or discuss critical issues, so whole-class discussion strengthens the manipulation by ensuring that students are exposed to the range of critical ideas and comparisons; 2) student-generated solutions, in addition to worked examples, can be the basis of instruction since the teacher can select students to share solutions that are different (something that cannot be ensured in partner work), and 3) teacher-led discussion more closely reflects typical classroom conditions, where the teacher usually plays a more central role. Studies 3 and 4 will be conducted in a small number of classrooms as an initial field test of the approach across two domains, and Study 5 will scale up to a greater diversity of classrooms for one of the domains.

Multi-step equation solving (e.g., more than one operation must be performed to solve an equation) is typically introduced in the seventh grade (both in general and in our participating schools). Thus, Studies 1, 3, and 5 will be on equation solving with seventh-grade students. Computational estimation of multidigit multiplication problems is typically discussed in fifth grade, and Studies 2 and 4 will address this topic with fifth-grade students.

In all studies, students will study worked-out examples of mathematics problems and answer questions about the examples. Students in the compare group will be shown a pair of worked examples illustrating different solution procedures to the same problem and be asked to compare and contrast the steps used and to reflect on why two different procedures lead to the same answer. Students in the sequential group will study the same two solution procedures, but will be shown each worked example separately on two isomorphic problems (i.e., problems with the same structure but with different numbers, such as $3(x + 1) = 9$ and $5(x + 2) = 20$) and be asked to think about the individual solutions. See Table 3 for a list of sample questions that students in the compare and sequential groups will answer in conjunction with the worked examples. After studying a set of worked examples and answering these questions, all students will try to solve a new problem using what they have noticed from the worked examples. All studies will be conducted in students’ mathematics classrooms, led by their regular mathematics teacher.
Table 3: Sample questions to accompany worked examples in all studies

<table>
<thead>
<tr>
<th>Compare group (Two worked examples presented simultaneously)</th>
<th>Sequential (control) group (Two worked examples presented separately)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Sammy and James did different first steps. Is it OK to do either step first in this problem? Explain your reasoning.</td>
<td>1. Do you think that the solution method used on this problem is a good one? Why?</td>
</tr>
<tr>
<td>2. Why might it be helpful to know two different ways to solve problems like this one?</td>
<td>2. When James multiplied, what number did he multiply by? Why did he multiply by that number?</td>
</tr>
</tbody>
</table>

In Studies 1 and 2, students will work in pairs, alternating between studying two worked examples for a single problem and solving a similar practice problem. This partner work will occur during two consecutive classroom periods. Examining students’ solutions to the practice problems during these two days will allow us to examine learning microgenetically over the course of the intervention. On the day following the partner work, students will receive a brief whole-class lecture wrap-up of key ideas (without use of contrasting examples), based on the findings from Schwartz and Bransford (1998) on the importance of an integrative lecture after exploratory experiences.

In Studies 3 and 4, the intervention used in Studies 1 and 2 will be adapted to a more typical classroom environment of combining partner activities with whole-class discussion of the activities and key ideas. The classroom intervention will occur over three consecutive classroom periods. The intervention will involve both the partner activities used in Studies 1 and 2, as well as whole-class problem solving (students solving problems, presenting their solutions to the class, and the teacher leading a whole-class discussion of the problem solutions and procedures used). In classrooms in the compare group, the activities and whole-class discussion will focus on contrasting solution procedures; in classrooms in the sequential group, the activities and discussion will be on sequentially discussing (but not comparing) the solution procedures. All classroom instruction will be led by the regular classroom teacher.

In Study 5, the classroom intervention will be “scaled up” to a much larger number of classrooms in public schools as a first step towards assessing the generalizability of this teaching approach. In all studies, students’ knowledge will be assessed on a pretest, immediate posttest and delayed posttest.

Hypotheses

In all five studies, we hypothesize that students in the compare group will show greater improvements in problem-solving ability (particularly transfer), procedural flexibility, and conceptual understanding than students in the sequential group. Improvements in problem-solving ability will be indicated by students solving both familiar and novel problems correctly. Improvements in procedural flexibility will be indicated by students using multiple solution procedures and by students recognizing and explaining the strengths and weaknesses of different procedures. Improvements in conceptual understanding will be indicated by students’ ability to articulate and apply key concepts in the domain, and integration of conceptual and procedural knowledge will be indicated by students’ ability to justify whether procedures make sense. Finally, we hypothesize that the rate of learning correct problem-solving procedures will be faster for students in the compare group, as
measured by how soon students begin to consistently solve the practice problems correctly during the intervention itself. Overall, we will use multiple assessments to provide converging evidence on the benefits of comparing alternative solution procedures.

Participants and Partner Schools

All students at the target grade level in each participating schools will be eligible to participate. The first four studies will be conducted with fifth- and seventh-grade students at an independent private school in an urban location—the University School of Nashville. (Contrary to its name, the school has not been affiliated with a university since 1974.) The school seeks to enroll students with racial, cultural, and economic diversity. Twelve percent of the student body is African American, and 10% of students are originally from other countries. Approximately 10% receive financial aid. Each grade has approximately 80 students divided into four classrooms. There is one mathematics teacher within each grade. The fifth-grade mathematics teacher has 35 years of teaching experience, and the seventh-grade teacher has five years of teaching experience.

The fifth study will be conducted with seventh-grade students at public suburban and rural middle schools in Tennessee and Michigan. We have confirmed the participation of one school district, Cheatham County Schools, in Tennessee. This district has approximately 550 seventh-grade students in three middle schools. It is a rural school district; 97% of the students are white and 27% qualify for free or reduced lunch. Assuming a class size of 28 per class, there are approximately 20 math classes across the three schools in this district. Our power analysis (see Study 5 description below) indicates that 18 classrooms are sufficient to detect moderate effects, so Study 5 could be run entirely within the Cheatham schools.

However, in the event of possible attribution or other unforeseen difficulties, we are attempting to secure the participation of one additional district in Michigan. We are currently negotiating with the Redford Union School District, which has approximately 350 seventh-grade students in one middle school (approximately 13 classes). It is a suburban school district; 88% of students are white and 26% qualify for free or reduced lunch. Should our attempts in Redford be unsuccessful, we have also begun preliminary conversations with two other districts in Michigan. We are confident that we will be able to identify a participating district in Michigan well in advance of the start of Study 5 in Fall 2007.

With regard to our selection of sites, note that poor mathematics achievement is not only a problem in poor, urban public schools, but rather is endemic at all types of schools, including the urban private school that is the focus of Studies 1 through 4 and the suburban and rural public schools that are the focus of Study 5. Studies 1 through 4 will evaluate the promise of the contrasting case approach, and Study 5 will evaluate the effectiveness of this intervention under more diverse classroom conditions.

Assignment to Groups

In Studies 1 and 2, students will be randomly paired in their mathematics class; each pair will then be randomly assigned to condition, with an approximately equal number of pairs in each condition within each mathematics class. In Studies 3 and 4, we will randomly assign the four classrooms to condition, with two classrooms participating in each condition. In Study 5, we will randomly assign classroom to condition, with an approximately equal number of classrooms in each condition within each school.
The set of studies proposed here is an important step in bridging from the cognitive science research that typically occurs in lab settings and with college students to real K-12 classrooms. Furthermore, the proposed program has important implications for basic cognitive science, and thus will advance both basic and applied research goals. An overview of the data collection plan for the four studies is presented in Table 4.

<table>
<thead>
<tr>
<th>Year 1</th>
<th>Fall 2005</th>
<th>Study 1</th>
<th>Topic: Equation Solving</th>
<th>Study 2</th>
<th>Topic: Computational Estimation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Spring 2006</td>
<td>-----</td>
<td></td>
<td>Study 2</td>
<td></td>
</tr>
<tr>
<td>Year 2</td>
<td>Fall 2006</td>
<td>Study 3</td>
<td>Topic: Equation Solving</td>
<td>Study 4</td>
<td>Topic: Computational Estimation</td>
</tr>
<tr>
<td></td>
<td>Spring 2007</td>
<td>-----</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year 3</td>
<td>Fall 2007</td>
<td>Study 5</td>
<td>Topic: Equation Solving</td>
<td>Study 4</td>
<td>Topic: Computational Estimation</td>
</tr>
<tr>
<td></td>
<td>Spring 2008</td>
<td>-----</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Study 1**

*Participants.* Approximately 80 seventh-grade students will participate during the fall semester of Year 1.

*Measures.* Students will complete an individual pretest, immediate posttest, and delayed posttest designed to assess their problem-solving knowledge, flexibility, and conceptual understanding of early algebra. Measures will be based on previous research conducted by both authors on equation solving (Rittle-Johnson & McMullen, 2003, 2004; Star, 2001, 2002, 2004; Star & Rittle-Johnson, 2004), as well as current piloting of the assessment at the participating school (described below). There will be four categories of measures:

1) Problem-solving knowledge (solve both familiar and novel (transfer) equations)
2) Flexibility (recognize, implement, and evaluate multiple solution procedures for the same problem)
3) Conceptual knowledge (answer questions designed to tap key concepts, such as maintaining equivalence)
4) Integration of problem solving and conceptual knowledge (justify whether procedures make sense)

See Table 1 in Appendix B for sample questions for each category and Table 2 in Appendix B for sample scoring rubrics for explanation items on the flexibility and conceptual knowledge assessments.

To provide more detailed, descriptive data on student learning and understanding, 10% of children in each condition (evenly distributed over each classroom) will also complete a similar assessment one-on-one with an experimenter, who will ask for think-aloud protocols of their methods and reasoning. In addition to using these experimenter-designed measures, we will obtain students’ mathematics grades on the unit test and for the term to evaluate whether there is a more general impact on children’s algebraic knowledge.

*Procedure.* The study will occur during one week of students’ regular mathematics classes and will be timed to replace the students’ regular instruction on solving relatively complex equations (and thus occur immediately after instruction on solving basic 2-step
equations). On Monday, students will complete the written pretest (approximately 30 minutes). If needed, the regular classroom teacher will then provide a brief review of needed skills from previous lessons (e.g., subtracting from both sides in one-step equations, using the distributive property). Next, the teacher will show students a new equation (similar to the ones to be used in partner work) and ask them to attempt to solve the problem. This will provide students with an opportunity to invent a procedure for solving the problem and/or with motivation to learn ways to solve the problems. Then, the teacher will emphasize that there are multiple ways to solve the problem correctly and discuss one way to solve the problem with the whole class (discussing two ways would contaminate the groups). Finally, the partner work (intervention) for the following two days will be introduced to the students. The objective on Monday is to prepare students for the following two days of the intervention.

On Tuesday and Wednesday, students will be broken into pairs to study sets of worked examples and solve similar practice problems. For each problem set, students in the compare group will see two alternative solutions to the same problem simultaneously and be asked to compare and contrast the two solutions, using questions like those in Table 3 (see also Table 3, Appendix B, which includes a sample pair of worked examples and student responses). Students in the sequential group will see the same two alternative solutions, but they will see them individually on different, isomorphic problems (e.g., $3(x + 1) = 15$ and $5(x + 2) = 20$) and be asked to think about the solutions individually, using questions like those in Table 3 (see also Table 4 in Appendix B). In both conditions, after studying each problem set, students will solve a similar practice problem on their own and check their answer (but not their solution procedure) with the teacher. They will then move on to the next worked example/practice problem set. Interleaving practice problems with worked examples helps students to monitor their understanding of solving the problems (and provides us with evidence of gradual changes in problem solving approaches; Atkinson et al., 2000). The problems used during the intervention will be based on those used by Star and Rittle-Johnson in prior studies (Star, 2001; Star & Rittle-Johnson, 2004; Star & Seifert, 2004) and will be selected to match the students’ existing curriculum as much as possible (e.g., $1/2(x + 3) = 10$, $2(t - 1) + 3(t - 1) = 25$, $4(y + 5) + 12 = 7(y + 5)$). Students will study and solve 4-6 sets of problems a day, for a total of 8-12 problem sets. During the partner work, the pairs of students will be asked to first explain their answers verbally to one another, then write down a summary of their answer on the worksheet. The written explanation serves to push students to summarize their ideas and come to a consensus. We will also record the verbal interactions (using an audiotape recorder and microphones for each pair), as we are concerned that students of this age may give richer explanations verbally than they articulate in writing. Tapes of the verbal interaction will be transcribed and used as an alternate source for coding students’ explanation quality. It will also provide us with an index of their on-task behavior. During the partner work, the regular classroom teacher and members of the project team will circulate and provide help when requested (e.g., by re-phrasing and breaking down questions, by providing general encouragement and by helping students implement steps during problem solving, without providing any guidance on what step to do next or why you might do a particular step). Each evening, students will all be given the same brief homework assignment to practice solving equations.

On Thursday, the classroom teacher will provide a scripted 10-minute integrative lesson, reviewing different procedures (but with no contrasting of methods) for solving equations and highlighting key ideas such as how one knows if a step is OK to do (e.g., must
keep both sides equal) and factors that influence choice among multiple solution paths (e.g.,
gets you closer to solving the equation, less likely to make mistakes). This integrative lesson
is motivated by findings of Schwartz and Bransford (1998) that contrasting cases are most
effective when followed by an integrative lesson. Students will then complete the posttest,
which is identical to the pretest. Selected students will also be interviewed on Thursday or
Friday, as noted above. **Two weeks later**, children will complete the posttest a second time to
assess retention.

**Analysis plan.** For each pair of students, we will average the two scores on each of the
measures at pretest and posttest and use these average partner scores in the analyses (rather
than scores from each individual). To ensure random assignment was effective, we will use
MANOVA to verify that the two groups did not differ in average partner score on the pretest
across the different measures (see Appendix B). To examine the effects of condition on the
posttest, we will use a combination of MANOVA to test for an overall effect of condition on
the four different measures, as well as follow-up one-way ANOVA tests to evaluate the effect
of condition on each of the measures. We will also use ANOVA to compare the number of
practice problems solved correctly during the intervention between the two conditions.
Finally, we will code the quality of students’ spoken and written explanations during the
intervention, using existing categories for guidance (e.g., state purpose or goal for a particular
solution step; integrate two examples). Again using ANOVA, we will then compare the
frequency of different explanation types between the two conditions. Power was calculated
for these one-way ANOVAs. Assuming a small-to-moderate effect size of three-quarters of a
standard deviation, with alpha = 0.05 and \( n = 20 \) pairs per condition, the power is 0.8. Results
from our pilot study below also suggest adequate power.

**Pilot of Study 1 (completed).** A pilot of Study 1 was conducted in November 2004,
with 70 seventh-grade students (18 pairs of students in the compare condition and 17 pairs in
the sequential condition). The pilot indicated that the proposed methodology is feasible and
implementable. The students understood and engaged in the desired tasks. The pilot is
providing useful insights, such as the need for using a larger variety of reflection questions
during the intervention to reduce repetitive student answers.

Preliminary findings from this pilot study are very encouraging. Pairs in the compare
condition scored higher at posttest on the problem-solving and flexibility assessments (there
were no reliable condition differences at pretest, and pretest performance was included as a
covariate in all posttest analyses). First, consider problem solving, which assessed posttest
performance on four *learning* problems (structurally identical to those presented during the
intervention) and four *transfer* problems (see Table 1, Appendix B). A repeated-measures
ANCOVA, with problem type (learning or transfer) as a within-subject factor and condition
(compare or sequential) as a between-subject factor, indicated a main effect for condition,
\[ F(1, 31) = 4.78, p = .04, \text{ partial } \eta^2 = .13, \] and no main effect or interaction with problem type.
On both learning and transfer problems, pairs in the compare group were more successful at
solving the problems correctly. Next, consider flexibility, which was a composite measure
based on the three item types outlined in Table 1, Appendix B. A one-way ANCOVA
indicated that pairs in the compare condition were more flexible than pairs in the sequential
group at posttest, \[ F(1, 32) = 4.32, p = .05, \text{ partial } \eta^2 = .12. \] For the conceptual knowledge
assessment, we are still in the process of developing and implementing scoring rubrics to
evaluate the quality of students’ explanations. Unfortunately, the items we used that did not
require explanation were often too easy, and students were near ceiling at pretest (e.g., 94%
knew that the expression $k + 6$ can have infinitely many values, which was an item taken from the 1996 NAEP). Outcomes on the conceptual knowledge assessment allowed us to revise proposed assessment items to a more appropriate difficulty level and to develop scoring rubrics for explanation items (see Tables 1 and 2 in Appendix B).

**Study 2**

*Participants.* Approximately 80 fifth-grade students will participate during the spring semester of Year 1.

*Measures.* As in Study 1, students will complete an individual pretest, immediate posttest, and delayed posttest designed to assess their problem-solving ability, procedural flexibility and conceptual understanding. The assessment will be based on items used in previous research on computational estimation (LeFevre et al., 1993; Reys et al., 1980; Rubenstein, 1985; Sowder & Wheeler, 1989), as well as on national and international assessments such as NAEP and TIMSS, and will cover the same problem categories as in Study 1. See Table 1 in Appendix B for sample questions from each category.

*Procedure and analysis plan.* The procedure and analysis plan for Study 2 will be identical to what was described above for Study 1, but with computational estimation as the target domain. Again, the timing of the intervention will be coordinated with the students’ regular mathematics curriculum. The problems used during the intervention will be based on those used by LeFevre and colleagues (LeFevre et al., 1993) and will be matched to the students’ existing curriculum as much as possible. Tables 5 and 6 in Appendix B provide examples of a pair of worked examples students in the compare group and sequential group would see in Study 2.

**Studies 3, 4, and 5 Overview**

Studies 1 and 2 will provide critical data on the effectiveness of students comparing alternative solutions, rather than thinking about solution procedures individually. Studies 3, 4, and 5 will expand these studies to more typical classroom conditions, which should also strengthen the manipulations. Students will begin by solving an equation independently; a few students will then be asked to share their (different) solution procedures. Thus, the different solution procedures will often be generated by students within the class, rather than provided through worked examples, which is more typical in reform classrooms (e.g., Hiebert et al., 1996; Lampert, 1990) and which may benefit student understanding (although hypothetical student solutions—i.e., worked examples—will be used if no one in the class generates an important alternative solution). Next, the teacher will lead the students in a discussion of the procedures, using questions similar to those used in Studies 1 and 2. By having the teacher lead this discussion, we can ensure that all students are at least exposed to critical issues and comparisons (which cannot be ensured in partner work). Exposure to the ideas in whole-class discussion should then facilitate discussion of the ideas in partner work. Finally, this discussion will be led by the regular classroom teacher to evaluate the effectiveness under normal classroom conditions. Students will also discuss solutions in pairs, as in Studies 1 and 2.

**Study 3**

*Participants.* Four seventh-grade mathematics classes, with approximately 20 students each, will participate in Study 3 during the fall of Year 2.
Measures. Identical measures to those used in Study 1 will be used in Study 3.

Teacher training procedure. Participating teachers will attend a one-day seminar on the goals of the study and how to lead the discussions. The seminar will be conducted by Star, who has previous experience with teacher training. The training materials will include 1) target alternative procedures to be looking for, and how to introduce these procedures if they are not spontaneously generated by the students; and 2) target big ideas to elicit during the discussion, such as how to know if a step is OK to do (legitimate) and factors that influence choice among alternative procedures (e.g., get you closer to the solution or make you less likely to make errors).

Procedure. The study will occur during one week of students’ regular mathematics classes and will be led by the regular mathematics teacher.

On Monday, the procedure will be very similar to that used in Study 1. Students will complete the written pretest (approximately 30 minutes). If needed, the teacher will provide a brief review of needed skills from previous lessons. Finally, the teacher will lead an activity to introduce the intervention to be used during the following three classroom periods. Students will be given one algebra equation to solve on their own. The teacher will have several students share their (different) solution methods, then engage the class in a discussion of the solution procedures. In the compare condition, the teacher will have multiple students share their solution procedures and then ask students to compare and contrast the solution methods, using questions similar to those used in compare group pairs in Study 1. In the sequential condition, the teacher will have students share solution procedures one at a time and have students discuss each solution procedure individually, using questions similar to those used by the sequential group pairs in Study 1.

On Tuesday, the teacher will repeat the solve-individually-then-share-and-discuss cycle introduced on Monday with two additional problems. Next, students will work with a partner, using the same procedure as in Study 1 for partner work. Towards the end of the class period, the teacher will lead a whole-class discussion of two of the problems from the partner work. Whole-class discussions will allow the teacher to ensure that students consider factors each individual pair of students may not spontaneously address. The teacher will have a set of ideas that should be discussed, either spontaneously by students or introduced by the teacher. For example, the teacher may need to guide attention to how you decide if a step is legitimate or to noticing problem features that influence choices between solution procedures (see Teacher Training Procedure section). The teacher will also encourage students to expand and build upon one another’s ideas, since students tend to provide a basic response and move on without thinking more deeply about the issue.

The critical difference between the discussion in the compare classrooms and the sequential classrooms will be whether alternative solutions are contrasted with one another and students are prompted to think about these comparisons (using prompts from Study 1). In the compare classrooms, children will be probed to think about why two solution methods lead to the same answer and why you might chose one method over another on a particular problem. In the sequential classrooms, children will be probed to think about the legitimacy of individual steps and procedures and whether individual methods are good to use, but will never directly compare different solution methods.

Wednesday and Thursday will have a similar format as Tuesday, with alternation between whole-class discussion of solution procedures generated by the students and partner activities studying worked examples. (Students will not keep the same partner during the
three-day period.) Increasingly sophisticated equation solving procedures will be illustrated in the worked examples, and increasingly difficult problems will be introduced over the three days.

On Friday, the teacher will give a brief integrative lesson similar to that used in Study 1, and students will complete the posttest. Two weeks later, students will complete a delayed posttest identical to the immediate posttest to assess retention.

To ensure fidelity of the treatment, a member of the project team will observe each classroom lesson and have a checklist for monitoring whether key procedures and ideas were discussed. If needed, the project member will remind the teacher to introduce a key idea that has not been covered. All classroom sessions will be videotaped in order to assess fidelity of the treatment.

Analysis plan. For our current purposes, we will use individual performance on the pretest and posttest as the unit of analysis, with classroom membership included as a control variable. We recognize that there are some confounds in using individual student performance as the unit of analysis when instruction occurs in a classroom context, but this approach will provide crucial supportive evidence needed before scaling up to a more rigorous, multi-site, many-classroom evaluation of our instructional approach. We will compare student performance under the two instructional conditions on the immediate and delayed posttests, using analyses parallel to those used in Study 1. Power was calculated for the primary one-way ANOVAs, assuming an effect size of three-quarters of a standard deviation, with alpha = 0.05 and $n = 40$ per condition, yielding a power of 0.952.

We will also explore differences in performance between students in the two classrooms assigned to the same condition. This will provide some index of the impact of differences in whole-class discussion that may emerge between classrooms in the same condition and allow us to explore aspects of classroom dialogue that may be particularly important for promoting learning (if differences in learning do exist). In addition, videotapes of classroom sessions will provide qualitative data on differences that emerge both within and between the compare and sequential conditions.

Study 4
Participants. Four fifth-grade mathematics classes, with approximately 20 students each, will participate in Study 4 during the spring of Year 2, along with their mathematics teachers.

Measures. Identical measures to those used in Study 2 will be used in Study 4.

Teacher training procedure. As in Study 3, participating teachers will attend a one-day seminar on the goals of the study and how to lead discussions in the different classrooms.

Procedure and analysis plan. The procedure and analysis plan for Study 4 will be identical to what was described for Study 3, but with estimation as the target domain. Target big ideas that the teacher will ensure students discuss include why different estimation procedures yield different estimates that are both “correct” and features for noticing when a particular procedure is most appropriate (e.g., the presence of a 9 in an important place value makes rounding up an easy procedure that yields a relatively close estimate).

Study 5
Study 5 will evaluate the generalizability of the approach to a wider range of students and teachers, using the seventh-grade lessons on algebra as the sample content.
Participants. At least 20 seventh-grade mathematics classes from schools in two public school districts (in Michigan and Tennessee) will participate, along with their regular mathematics teacher. Random assignment will occur at the classroom level.

Measures. Identical measures to those used in Studies 1 and 3 will be used in Study 5.

Teacher training procedure. As in Studies 3 and 4, teachers will participate in an all-day workshop led by Star. The training will focus on principles of the study and target content for coverage, as well as techniques for facilitating productive discussion (as outlined above). A member of the project team will also meet with each teacher individually several times before the implementation to address individual concerns.

Procedure. The procedure used in Study 5 will be identical to that used in Study 3.

Analysis plan. Given that random assignment will occur at the classroom, not the individual, level, and that knowledge is assessed multiple times, we will use a three-level hierarchical linear regression model (Bryk & Raudenbush, 1992). The first level of the model measures individual knowledge growth over time for each student. We are interested in the growth rate of students and will have only three data points, so the model will assume the growth rate is linear. The second level of the model measures differences between students within classrooms. The third level estimates differences between classrooms. Because our treatment occurs at the classroom level, the treatment effect will be defined at this third level. For supplemental analyses, such as comparing performance during the intervention (e.g., number of problems solved or quality of explanations when working alone or with a partner), we will use a two-level hierarchical linear regression model, with differences between students within classrooms as Level One and differences between classrooms as Level Two.

Power was conducted using procedures recommended for HLM by Raudenbush and Liu (2000), using their Optimal Design power program (available at www.umich.edu/~rauden). Assuming an alpha of .05, 28 students per classroom, an intra-class correlation of .05, and a small-to-medium standardized effect size of .40 (Cohen, 1988), 18 classrooms will yield a detectable contrast with 80% confidence.

Summary

The current studies combine and extend basic research in cognitive science to help solve a critical goal in education—how to inculcate flexible, integrated knowledge rather than inert knowledge. Having students explain contrasting examples is a promising, domain-general instructional approach for reaching this goal. The current studies have important implications for basic cognitive science and for education.

First, they evaluate whether analogical encoding with contrasting cases is beneficial to a new class of tasks (problem solving) and to new outcomes (problem-solving ability, procedural flexibility, and conceptual understanding). Second, they evaluate the use of contrasting examples as a method for improving both the quality of students’ explanations in a social context and students’ subsequent learning. Thus, they provide important extensions of basic research in cognitive science on analogical encoding with contrasting cases and on learning from self-explaining worked examples.

Finally, the five studies evaluate the effectiveness of explaining contrasting examples for teaching mathematics in classroom settings. They provide rigorous evaluation of components of current educational reforms in mathematics. These reforms advocate that teachers act more as facilitators, encouraging students to explain their own thinking and problem-solving procedures to other students. However, the optimal content for these
discussions is unclear, and the research basis for this reform effort is inconclusive. The current studies will provide valuable evidence towards evaluating this claim. They will also provide specific examples of curriculum materials for two foundational components of mathematics instruction—computational estimation and equation solving.

**Personnel**

**Dr. Bethany Rittle-Johnson,** Vanderbilt University, will serve as co-principal investigator on the proposed studies. Rittle-Johnson is an assistant professor in the Psychology and Human Development Department at Peabody College, Vanderbilt University. She completed her Ph.D. in Developmental Psychology at Carnegie Mellon University under the direction of Drs. Robert Siegler and Martha Alibali. Her dissertation examined the mutually supportive, iterative relations in the development of conceptual and procedural knowledge of mathematics. IES Director Grover Whitehurst described this research in his address at the third education summit as an example of rigorous research that informs the Math Wars. During her post-doc, Rittle-Johnson collaborated with Dr. Kenneth Koedinger in designing and piloting a Cognitive Tutors middle-school mathematics curriculum, which included a paper curriculum for use three days a week in the classroom and an intelligent tutoring system for use two days a week in the computer lab. Rittle-Johnson has published in *Journal of Educational Psychology* and *Child Development* and presented her work at numerous psychology and educational conferences.

**Dr. Jon Star,** Michigan State University, will serve as co-principal investigator on the proposed studies. Star is an assistant professor in the College of Education at Michigan State University, with a joint appointment in the Department of Teacher Education and the Department of Counseling, Educational Psychology, and Special Education. He completed his Ph.D. in Education and Psychology at the University of Michigan in 2001 under the direction of Dr. Colleen Seifert. His dissertation explored the development of students’ knowledge of equation-solving procedures. Star has presented his thesis research at several conferences; a paper from his thesis is currently under review. Star has published in *Journal for Research in Mathematics Education, Developmental Psychology, Journal of Cognition and Development,* and *Cognition and Instruction,* and he has collaborated with Dr. Susan Gelman, Dr. Jack Smith, and Dr. Edward Silver, among others. Star currently is co-PI on an NSF-funded project that investigates the design, implementation, and effectiveness of professional development for college mathematics instructors. Prior to graduate school, Star worked for six years as a middle- and secondary-school mathematics teacher. He currently teaches courses for both pre-service and in-service mathematics teachers and has also conducted numerous teacher professional development workshops for inservice middle-school mathematics teachers.

Together, Drs. Rittle-Johnson and Star are among the few new psychologists conducting research in education—a rarity, according to Dr. Whitehurst’s recent interview in the *APA Monitor* (October, 2003).

**Dr. Daniel Schwartz,** Stanford University, will serve as a project consultant. He is an associate professor of Learning, Design & Technology in the School of Education at Stanford University. Before completing his Ph.D. at Teachers College, Columbia University, in 1992, Dr. Schwartz worked for eight years as a secondary-school teacher. His research examines how people move from untutored mental models to more formal and verbal understanding in the domains of mathematics, mechanics, and biology. His work employs laboratory and computer-modeling methodologies, as well as classroom interventions. He is one of the
leading researchers in the use of contrasting cases to prepare students for future learning. He worked extensively on the NSF-funded Jasper Woodbury and Scientist in Action series. He is the PI of NSF-REC grants to develop innovative ways to prepare students to learn statistics and to explore the use of teachable agents to evaluate the wisdom that people learn best when they teach. He is also PI on NSF-BCS grants examining the role of action in spatial cognition and the significance of external representations for child development. His work appears in books and journals including *Cognitive Psychology; Cognitive Science; Cognition and Instruction; Educational Technology Research and Development; Mind, Culture & Activity; Review of Research in Education; Journal of the Learning Sciences; Journal of Research in Science Teaching; Journal of Applied Psychology; Journal of Experimental Psychology: Learning, Cognition, & Memory; and Memory and Cognition.*

**Dr. Ken Koedinger,** Carnegie Mellon University, will also serve as a project consultant. He is an associate professor in the Human-Computer Interaction Institute at Carnegie Mellon University. His background includes a B.S. in Mathematics, an M.S. in Computer Science, a Ph.D. in Cognitive Psychology, and experience teaching in an urban high school. This multi-disciplinary preparation has been critical to his research goal of creating educational technologies that dramatically increase student achievement. Toward this goal, he creates “cognitive models,” computer simulations of student thinking and learning, that are used to guide the design of educational materials, practices, and technologies. These cognitive models provide the basis for an approach to educational technology called “Cognitive Tutors,” which create rich problem-solving environments for students to work in and provide just-in-time learning assistance much like a good human tutor does. Koedinger has developed Cognitive Tutors for mathematics and science and has tested them in the laboratory and the classroom. The U.S. Department of Education designated the Cognitive Tutor Algebra course as one of five "Exemplary Curriculum" for K12 mathematics. A co-founder of Carnegie Learning Inc., a company marketing these technology-enhanced learning solutions to schools and colleges across the country, Koedinger’s research has contributed new principles and techniques for the design of educational software and has produced basic cognitive science research results on the nature of mathematical thinking and learning. He has authored 47 peer-reviewed publications, 2 textbooks, 6 book chapters, and 41 other papers and has been a project investigator on 15 major grants, including a current grant from the Institute of Educational Sciences.

**Dr. Kimberly Maier,** Michigan State University, will serve as a project statistical consultant. Maier is an assistant professor of Measurement and Quantitative Methods in the Department of Counseling, Educational Psychology, and Special Education in the College of Education, Michigan State University. She received her Ph.D. in Education at the University of Chicago in 2000 under the direction of Larry Hedges and Barbara Schneider. Maier has published three recent articles in the *Journal of Educational & Behavioral Statistics.* Her current research focuses on the application of hierarchical measurement models to multilevel item response theory. She has developed the Hierarchical Measurement Model (HMM), a statistical model that enables accurate estimation of the effects of multilevel covariates on latent trait measures. Other areas of interest include Bayesian data analysis methods for educational research and the study of gender differences in science and mathematics education.
Resources

Both the Department of Counseling, Educational Psychology and Special Education at Michigan State University and the Department of Psychology and Human Development at Vanderbilt University would provide excellent facilities to support the design and analysis phases of this project. Resources include all necessary office space (for PIs and research staff), office equipment, phones, copying facilities, basic secretarial supplies, and extensive networking and computer backup systems. The departmental research management staff and computer support staff, and the college technology support staff, would support the smooth completion of this research. The Institute for Research on Teaching and Learning at Michigan State and the Learning Sciences Institute—including the new IES-sponsored Experimental Education Research Training (ExpERT) program—at Vanderbilt University will provide additional professional resources and services as needed. Michigan State and Vanderbilt University both have top-ranked colleges of education, a rich variety of faculty expertise in psychological and educational research, and are leaders in the development of educational interventions.

Studies 1, 2, 3, and 4 will be conducted at the University School of Nashville (it is no longer affiliated with a university). The head of the middle school is a mathematics teacher himself and is enthusiastic about making all necessary arrangements for the smooth completion of this project. Rittle-Johnson has ongoing relations with several teachers at the school that will also facilitate the execution of this project. Study 5 will be conducted at several public rural and suburban schools in Tennessee and Michigan, as described above.

All research and data collection will abide by the strict guidelines for human subjects through the Institutional Review Boards at Vanderbilt University and Michigan State University.
# APPENDIX B

Table 1: Sample items for assessing problem solving, flexibility, and conceptual understanding

<table>
<thead>
<tr>
<th>Problem Type</th>
<th>Linear equation solving (Studies 1, 3, and 5)</th>
<th>Computational estimation (Studies 2 and 4)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PROBLEM SOLVING</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| a. Procedural learning (solve problems similar to those used during the intervention) | 1. Solve this equation:  
   \[ -\frac{1}{4} (x - 3) = 10 \]  
   2. Solve this equation:  
   \[ 4(x + 1) = 8(x + 1) \]  
   *(from Star, 2001, and pilot of proposed work)* | 1. Estimate the answer to 5 \times 144.  
   2. If you collected 82 pop bottles a day for 12 days, approximately how many pop bottles would you have?  
   *(from LeFevre et al., 1993)* |
| b. Procedural transfer (solve novel problems) | 3. Solve this equation:  
   \[ 2(2x + 3x) + 2 = 12 \]  
   4. Solve this equation:  
   \[ -3(x + 5 + 3x) - 5(x + 5 + 3x) = 24 \]  
   *(from Star, 2004, and pilot of proposed work)* | 3. Raymond must buy enough paper to print 28 copies of a report that contains 64 sheets of paper. Paper is only available in packages of 500 sheets. How many whole packages of paper will he need to buy to do the printing?  
   *(from NAEP, 1992)* |
| **FLEXIBILITY**                       |                                                                                                              |                                                                                                                                                                        |
| a. Prompted flexibility (prompt to solve a problem in two different ways) | 1. Solve this equation in two different ways:  
   \[ 3(x + 1) + 6(x + 1) = 18 \]  
   *(from Star, 2001 and pilot of proposed work)* | 1. Estimate the answer to 5 \times 144 using two different ways. |
| b. Recognize multiple correct ways    | 2. For the equation 4(x + 3) + 2 = 8, identify all possible steps that could be done next.  
   *(Response choices are: combine like terms, distribute, add/subtract on both sides, and multiply/divide on both sides)*  
   *(from Star, 2004, and pilot of proposed work)* | 2. You need 28 copies of a 19-page paper. Identify all reasonable estimates of how many sheets of paper you will need?  
   *(Response choices are: 30 \times 20; 30 \times 20 – (2 \times 20); 2 \times 1 = 2, so 200; 30 \times 15)* |
| c. Evaluate & justify alternative/atypical approaches | 3. Here is the first step of a student’s solution to an equation:  
   \[ 3(x + 2) = 12 \]  
   \[ x + 2 = 4 \]  
   Do you think that this way of starting this problem is (a) a very good way; (b) OK to do, but not a very good way; (c) Not OK to do? Explain your reasoning.  
   *(from Star, 2004, and pilot of proposed work)* | 3. Mark’s garden has 84 rows of cabbages. There are 57 cabbages in each row. Which of these gives the BEST way to estimate how many cabbages there are altogether?  
   a. 100 \times 50 = 5000  
   b. 90 \times 60 = 5400  
   c. 80 \times 60 = 4800  
   d. 80 \times 50 = 4000  
   Explain your reasoning:  
   *(from TIMSS)* |
<table>
<thead>
<tr>
<th>CONCEPTUAL KNOWLEDGE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Answer questions about domain concepts</strong></td>
</tr>
<tr>
<td>1. Here are two equations:</td>
</tr>
<tr>
<td>[ 213x + 476 = 984 ]</td>
</tr>
<tr>
<td>[ 213x + 476 + 4 = 984 + 4 ]</td>
</tr>
<tr>
<td>Without solving either equation, what can you say about the answers to these equations? (Responses are: both answers are the same; both answers are different; I can’t tell without doing the math). Explain your reasoning</td>
</tr>
<tr>
<td>2. What does the equals sign mean?</td>
</tr>
<tr>
<td>3. If ( m ) is a positive number, which of these is equivalent to ( m + m + m + m )? (Responses are: ( 4m ); ( m^4 ); ( 4(m + 1) ); ( m + 4 ))</td>
</tr>
<tr>
<td><em>(from NAEP, 1992)</em></td>
</tr>
<tr>
<td>4. Without solving, which of the following equations is equivalent to the equation ( 32(x - 12) = 96 )? Circle all that apply.</td>
</tr>
<tr>
<td>(Responses are: ( 32x - 12 = 96 ); ( x - 12 = 96 - 32 ); ( 16x - 16 \cdot 12 = 48 ); ( 16x - 6 = 48 ); ( 32(x - 12)/32 = 96/32 )). Explain your reasoning.</td>
</tr>
<tr>
<td>5. Encoding task: Recall algebra equations from memory (These items serve as an index of organized knowledge – e.g., Chase and Simon, 1973)</td>
</tr>
<tr>
<td><em>(from Star, 2004; Rittle-Johnson &amp; McMullen, 2003, 2004; pilot of proposed work)</em></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>INTEGRATED KNOWLEDGE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>(justify procedures)</strong></td>
</tr>
<tr>
<td>1. Here is how Sally solved this equation: (show worked example) Does Sally’s solution make sense? Explain why or why not.</td>
</tr>
<tr>
<td>1. Here is how Sally estimated the answer to this problem: (show worked example) Does Sally’s way make sense? Explain why or why not.</td>
</tr>
</tbody>
</table>
## Table 2: Sample scoring rubrics for students’ explanation to sample linear equation-solving items in Table 1, Appendix B (based on Pilot)

<table>
<thead>
<tr>
<th>Item</th>
<th>Quality</th>
<th>Description</th>
<th>Sample Explanation from Pilot</th>
</tr>
</thead>
<tbody>
<tr>
<td>FLEXIBILITY: Here is the first step of a student’s solution to an equation: [3(x + 2) = 12] [x + 2 = 4] Do you think that this way of starting this problem is (a) a very good way; (b) OK to do, but not a very good way; (c) Not OK to do? Explain your reasoning.</td>
<td>Excellent</td>
<td>Justify why step is OK (e.g., do the same thing to both sides) and evaluate its efficiency or simplicity</td>
<td>“They did the same thing to both entire sides of the problem using the properties of equality. It is also a quicker way to get to the answer than other ways to do it.”</td>
</tr>
<tr>
<td></td>
<td>Good</td>
<td>Accurately evaluate its efficiency or simplicity</td>
<td>“It works well on this problem because it simplifies it down and now there is only one step left.”</td>
</tr>
<tr>
<td></td>
<td>Partial</td>
<td>Express preference for an alternative method</td>
<td>“For me it is easier to distribute across parentheses.”</td>
</tr>
<tr>
<td></td>
<td>Poor/Incorrect</td>
<td>Incorrect or no explanation</td>
<td>“You have to always do parenthesis first, so they should have distributed across parentheses first before multiplying or dividing it.”</td>
</tr>
<tr>
<td>CONCEPTUAL: 1 Here are two equations: [213x + 476 = 984] [213x + 476 + 4 = 984 + 4] Without solving either equation, what can you say about the answers to these equations? (a) Both answers are the same, (b) both answers are different, (c) can’t tell without doing the math. Explain your reasoning.</td>
<td>Excellent</td>
<td>Choose (a) and justify that doing same thing to both sides doesn’t change the value of (x)</td>
<td>“If you add a number, such as four, to both sides (of the equal sign) you end up with the same answer as if you didn’t add it.”</td>
</tr>
<tr>
<td></td>
<td>Good</td>
<td>Choose (a) and reference numbers canceling out in the second equation</td>
<td>“Because both of the fours cancel out.”</td>
</tr>
<tr>
<td></td>
<td>Partial</td>
<td>Choose (a) but cannot justify</td>
<td>“Can’t explain”</td>
</tr>
<tr>
<td></td>
<td>Poor/Incorrect</td>
<td>Choose (b) or (c) and/or give incorrect or no explanation</td>
<td>“The answer has to be different because an extra 8 is added in the second equation.”</td>
</tr>
<tr>
<td>2. Without solving, which of the following equations is equivalent to the equation [32(x – 12) = 96]? Circle all that apply: (a) [32x – 12 = 96]; (b) [x – 12 = 96 – 32]; (c) [16x – 16\times12 = 48]; (d) [16x – 6 = 48]; (e) [32(x – 12)/32 = 96/32]. Explain your reasoning</td>
<td>Excellent</td>
<td>Justify why (c) and (e) are correct, typically by referencing doing the same thing to both sides</td>
<td>“E, because all that changed was both sides were divided by the same number (division property of equality) and C, because both sides were divided by two, and the rest of the distributing will take place (16 x 2).”</td>
</tr>
<tr>
<td></td>
<td>Good</td>
<td>Specific justification of at least one correct choice</td>
<td>“If you multiply both sides of equation [e] by 32, both would read the same.”</td>
</tr>
<tr>
<td></td>
<td>Partial</td>
<td>Vague justification of at least one correct choice</td>
<td>“Because that’s one way you could solve the problem. It is a different way to state it.”</td>
</tr>
<tr>
<td></td>
<td>Poor/Incorrect</td>
<td>Incorrect choice(s) and/or give incorrect or no explanation</td>
<td>“I chose this one [a] because it is the only one that is written down showing it has an answer of 96.”</td>
</tr>
</tbody>
</table>
Table 3: Sample pair of worked examples to an algebra equation for compare group
(Studies 1, 3, and 5).

Note: Sample student explanation from pilot is included to illustrate expected student response.

1. Here are two students’ solutions to the algebra equation, $2(x - 3) = 8$. Describe each student’s solution to your partner and then answer the questions below.

<table>
<thead>
<tr>
<th>Sammy’s Solution:</th>
<th>James’ Solution:</th>
</tr>
</thead>
<tbody>
<tr>
<td>$2(x - 3) = 8$</td>
<td>$2(x - 3) = 8$</td>
</tr>
<tr>
<td>$2x - 6 = 8$</td>
<td>$x - 3 = 4$</td>
</tr>
<tr>
<td>$2x = 14$</td>
<td>$Add$ or $Add$</td>
</tr>
<tr>
<td>$x = 7$</td>
<td>$7$</td>
</tr>
</tbody>
</table>

1. Sammy and James did different first steps. Is it OK to do either step first in this problem? Explain your reasoning.

Yes, it is. James’ solution (first step) was correct, like Sammy’s, but James realized that if $2(x-3) = 8$, then $x-3$ must equal $1/2$ of 8. He did more of it in his head.

2. Why might it be helpful to know two different ways to solve equations like this one?

If you knew two different ways you can check your work and if one way is easier, use it. If one way uses easier numbers to divide by, that way would be easier.
Table 4: Sample worked examples to two isomorphic algebra equations for sequential group (Studies 1, 3, and 5).

Note: Sample student explanation from pilot is included to illustrate expected student response.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Here is one student’s solution to the algebra equation 2(x – 3) = 8. Describe the student’s solution to your partner, and then answer the questions below</td>
<td></td>
</tr>
<tr>
<td>Sammy’s Solution:</td>
<td></td>
</tr>
<tr>
<td>2(x - 3) = 8</td>
<td></td>
</tr>
<tr>
<td>2x - 6 = 8</td>
<td>Distribute</td>
</tr>
<tr>
<td>2x = 14</td>
<td>Add on both</td>
</tr>
<tr>
<td>x = 7</td>
<td>Divide on both</td>
</tr>
<tr>
<td>1. Do you think the solution method used on this problem is a good one? Why?</td>
<td></td>
</tr>
<tr>
<td>Yes, because he distributed in the right way and he added and divided on both sides correctly.</td>
<td></td>
</tr>
</tbody>
</table>

----- (next page) -----  

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2. Here is one student’s solution to the algebra equation, 4(x - 5) = 12. Describe the student’s solution to your partner, and then answer the questions below</td>
<td></td>
</tr>
<tr>
<td>James’ Solution:</td>
<td></td>
</tr>
<tr>
<td>4(x - 5) = 12</td>
<td></td>
</tr>
<tr>
<td>x - 5 = 3</td>
<td>Divide on both</td>
</tr>
<tr>
<td>x = 8</td>
<td>Add on both</td>
</tr>
<tr>
<td>1. When James added on both sides, what number did he add? Why did he add that number?</td>
<td></td>
</tr>
<tr>
<td>He added 5 on both sides to get x by itself.</td>
<td></td>
</tr>
</tbody>
</table>
Table 5: Sample pair of worked examples to a computational estimation problem for compare group (Studies 2 and 4)

1. Here are two ways that students estimated the answer to 25 x 9. Describe each student’s solution to your partner and then answer the questions below.

<table>
<thead>
<tr>
<th>Sammy’s Solution (Rounding):</th>
<th>James’ Solution (Front-end multiplying):</th>
</tr>
</thead>
<tbody>
<tr>
<td>25 x 10 = 250</td>
<td>2 x 9 = 18, so 180</td>
</tr>
<tr>
<td>“I rounded the 9 to a 10 and then multiplied 25 x 10, which is 250.”</td>
<td>“I multiplied 2 times 9 and got 18. That’s 18 tens, so my estimate is 180.”</td>
</tr>
</tbody>
</table>

1. Sammy and James estimated the answers differently. Is it OK to estimate either way for this problem? Explain your reasoning.

2. Why might it be helpful to know two different ways to estimate answers to problems like this one?
### Table 6: Sample worked examples to two isomorphic computational estimation problems for sequential group (Studies 2 and 4)

1. Here is how one student estimated the answer to $25 \times 9$. Describe the student’s solution to your partner, and then answer the question below.

   **Sammy’s Solution (Rounding):**
   
   $25 \times 10 = 250$
   
   “I rounded the 9 to a 10 and then multiplied $25 \times 10$, which is 250.”

1. Do you think the estimation method used on this problem is a good one? Why?

2. Here is how one student estimated the answer to $9 \times 55$. Describe the student’s solution to your partner, and then answer the question below.

   **James’ Solution (Front-end multiplying):**
   
   $9 \times 5 = 45$, so 450
   
   “I multiplied 9 times 5 and got 45. That’s 45 tens, so my estimate is 450.”

1. When James multiplied, what number did he multiply 9 by? Why did he multiply by that number?