The development of flexible procedural knowledge in equation solving

Jon R. Star
Michigan State University
jonstar@msu.edu

Presented at the AERA 2004 symposium:
“Finding balance: Re-visiting the relations between conceptual and procedural knowledge”
Bethany Rittle-Johnson and Jon R. Star, Chairs
The development of flexible procedural knowledge in equation solving

Doing mathematics involves the use of mathematical procedures. Throughout the K-16 school mathematics curriculum, particularly after the introduction of symbolic algebra, students learn and are expected to use procedures. However, as reports assessing US mathematics achievement can attest (e.g., Beaton et al., 1996; Schmidt, McKnight, Cogan, Jakwerth, & Houang, 1999), many US students merely develop rote knowledge of procedures. As an educational outcome, rote knowledge is problematic in that it is resistant to transfer, error-prone, and inflexible (e.g., Anderson & Lebiere, 1998). Rote knowledge has direct links to low academic achievement in mathematics – students with rote knowledge have great difficulty on both near- and far-transfer problems (Hiebert & Carpenter, 1992), particularly in the domain of algebra (Blume & Heckman, 1997; Kieran, 1992). Instead of rote knowledge, our educational goals for the use of mathematical procedures involve flexibility, including the ability to select appropriate procedures for particular problems and modify procedures when conditions warrant. However, such a non-rote outcome for procedural learning has not been adequately considered in prior research, despite its clear potential for improving student achievement in mathematics.

The research reported here investigates the development of students’ knowledge of procedures, with the primary goal of identifying instructional conditions that are more likely to lead to flexible knowledge of procedures.

Rationale

The acquisition of procedural skill has been extensively studied in cognitive psychology, particularly in the past 40 years. There is convincing evidence that skills are learned through practice (Anderson & Lebiere, 1998; Fitts, 1964). A learner initially executes a skill through a slow, methodical process of executing factual instructions, step by step. With practice, the learner is able to perform the skill more quickly and efficiently. With continued practice, a learner is able to perform the skill even more quickly and efficiently, but at the cost of increasing inflexibility and rigidity in behavior, with the endpoint being a state of automaticity (Anderson, 1992, 1996; Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977).

From an educational standpoint, while the rapid, efficient, and automatic execution of a procedure is certainly one desired outcome, it is not the only important goal. Automatic executors are often derailed when they encounter problems that require very slight modifications of well-practiced procedures. While students with rote knowledge may know a set of procedures very well, they have great difficulty applying these procedures in problems or situations that are not identical to the settings in which they were learned. If we seek to maximize student achievement in mathematics, we must foster the development of students who not only can execute a procedure automatically but also flexibly. Increased flexibility enables students to solve problems that automatic executors cannot, particularly near- and far-transfer problems, and thus has great potential to lead to gains in student achievement.

The recently issued “Adding It Up” report from the National Research Council (2001) echoes this sentiment by identifying procedural fluency as one strand of mathematical proficiency, which they define as, “knowledge of procedures, knowledge of when and how to use them appropriately, and skill in performing them flexibly, accurately, and efficiently” (p. 121). Automatic execution fails to address all of the goals set forth by the “Adding It Up” report, particularly with respect to appropriate use of procedures and flexible performance. Mathematically proficient students should be able to move in and out of automatic execution as
experts do (Ericsson & Charness, 1994; Ericsson, Krampe, & Tesch-Romer, 1993); they should know “what to do and why” (Skemp, 1978, 1987); they should know “why [the procedure] has the steps it has, why it has certain conditions on its applicability, and why it succeeds when those conditions are met and the steps are followed accurately” (VanLehn, 1990, p. 38).

Numerous interventions have been undertaken to foster the development of non-rote learning outcomes for procedures. A common assumption among these interventions is that procedural knowledge must be connected or associated with more principled, “conceptual” knowledge (e.g., Hiebert & Lefevre, 1986). Connections to conceptual knowledge are fostered through a variety of activities, including the use of concrete manipulatives (e.g., base-ten blocks in elementary school), the generation of non-standard solution algorithms by students, and the in-class sharing and comparing of student-generated solution strategies. Promising results have been obtained using such activities, particularly in the domain of arithmetic (Carpenter, Franke, Jacobs, Fennema, & Empson, 1997; Fennema et al., 1996; Fuson et al., 1997; Hiebert et al., 1996).

Surprisingly, despite the promise of these approaches, little or no effort has been devoted to applying them to the learning of algebra procedures in high school. Arguably, flexibility in the use of mathematical procedures is paramount in algebra, given that procedures play a particularly prominent role in mathematics beginning with algebra (Kieran, 1992; Lewis, 1981; Thorndike et al., 1923). Algebra historically has represented students’ first sustained exposure to the abstraction and symbolism that makes mathematics powerful (Kieran, 1992); the symbolic procedures in algebra 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). As such, algebra is a particularly important venue in which procedural flexibility should be developed. Regrettably, students’ failure to do so has 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 National Assessment of Educational Progress indicates that many 12th graders can solve only the most simple and routine algebra tasks (Blume & Heckman, 1997).

Clearly much work remains to be done in pursuit of procedural flexibility in algebra for all students. The present study begins this effort by seeking to identify instructional conditions that foster flexibility in the use of procedures in secondary school mathematics.

**Defining flexibility**

Fundamentally, executing a procedural skill requires that one have knowledge of its component steps and the order in which these steps should be applied. But not all performances of a skill are the same. In particular, skillful execution in mathematics can mean two very different things. On the one hand, skillful execution involves being able to use procedures rapidly, efficiently, with minimal error, and with minimal conscious attention; in other words, to execute a procedure automatically or by rote (Anderson & Lebiere, 1998). On the other hand, being “skilled” means being able to deliberatively select appropriate procedures for particular problems and modify procedures when conditions warrant – in other words, to execute procedures flexibly.

Colloquially, flexibility refers to the ability to change according to particular circumstances and to have multiple strategies for successfully navigating changing or uncertain
conditions. Although there is little research specifically on the flexible use of algebra procedures (cf Krutetskii, 1976; Lewis, 1981), a number of studies have examined flexible use of arithmetic procedures (Blöte, Klein, & Beishuizen, 2000; Blöte, Van der Burg, & Klein, 2001; Carroll, 2000; Hiebert & Wearne, 1996; Resnick, 1980; Resnick & Ford, 1981). In this literature, there is some variation in the ways that flexibility is operationalized and defined.

A common view of flexibility defines it as the ability to select appropriate procedures for completing particular kinds of problems. For example, in the work of Blöte and colleagues (Blöte et al., 2000; Blöte et al., 2001), who studied students’ knowledge of procedures for performing mental addition, a student is flexible in the use of mental addition procedures if she chooses a procedure that most closely matches the number characteristics of a problem, because such a matching procedure allows for the most efficient solution of the problem. According to Blöte and colleagues, when solving the problem 62 – 29, a flexible solver would choose to subtract 62 – 30 (the nearest multiple of 10 to 29) to get 32, and then add back 1 (the difference between 30 and 29) to yield 33. Alternatively, a different procedure would be used when subtracting 62 – 24: A flexible solver would choose to start at 24 and progressively add until the total reaches 62: \(24 + 6 = 30, 30 + 30 = 60, 60 + 2 = 62, \) and \(6 + 30 + 2 = 38.\) Although both procedures can be used to solve both problems, Blöte and colleagues (Blöte et al., 2000; Blöte et al., 2001) claim that some procedures are easier to perform on certain problems. A flexible solver has knowledge of multiple procedures, recognizes which procedure is easiest to perform on a problem, and chooses to execute it.

A similar definition of flexibility can be found in the work of Dowker (1992; 1996). Dowker studied the strategies used by novices and experts as they estimated the answers to arithmetic problems. She found that experts not only knew more strategies, but they also had greater knowledge of how certain strategies were more or less useful on certain problems. In addition, experts showed greater variation in their strategies, even to the point of using different strategies when attempting identical problems on different occasions.

In other studies, flexibility is defined as the ability to invent procedures when problem conditions warrant (Resnick, 1980; Resnick & Ford, 1981); such flexibility (as invention) has been linked to greater conceptual understanding (Blöte et al., 2001). For example, Carpenter and colleagues (Carpenter, Franke, Jacobs, Fennema, & Empson, 1998) studied children’s invention of procedures for performing multi-digit arithmetic calculations and found that inventiveness was linked to understanding of base-ten concepts and greater success on transfer problems. Similarly, Carroll (2000) found that students who used an invented algorithm for solving a multi-digit addition problem were more likely to be able to generate a second way to solve the same problem than students who relied upon the standard algorithm.

Finally, in Siegler’s work on strategy choice (Shrager & Siegler, 1998; Siegler, 1996; Siegler & Jenkins, 1989; Siegler & Shrager, 1984), flexibility is closely tied to adaptability. Children have the ability to use many different strategies in completing simple addition problems. With increased problem-solving experience, choices among multiple strategies are made by considering which strategies are judged to be the most appropriate for a given problem and/or which ones have proven to be successful or efficient in previous problem-solving experiences.

All of these conceptions of flexibility are plausible; being flexible involves selecting appropriate procedures for particular problems and inventing new procedures when faced with unfamiliar problems. Accordingly, the present work adopts the following definition of flexibility
in the use of algebra procedures. A flexible solver has knowledge of multiple solution procedures and demonstrates the abilities to select appropriate procedures for particular problems, modify procedures when problem conditions warrant, and invent new procedures when faced with unfamiliar problems (see also Star & Seifert, 2002).

**How does flexibility develop?**

What conditions can lead to the development of flexible knowledge of procedures? This study considers and evaluates two instructional conditions that may be related to flexible knowledge, both of which are informed by hypotheses from the literature about how flexibility develops.

One hypothesis for how flexible knowledge develops, which draws heavily from the work of Sweller and colleagues [Owen, 1989 #309; Owen, 1985 #1015; Ward, 1990 #1009; Sweller, 1983 #304; Sweller, 1982 #315; Mawer, 1982 #316], suggests that in order to use procedures flexibly, one must have well-developed schemata in the problem-solving domain. Experts, who are presumably more flexible than novices, have been found to have more richly developed schemata. When faced with a new problem, experts can use a schema to “encode and unify” (Owen & Sweller, 1989, p. 323) the elements of the problem to plan and implement problem solving actions. Novices, in the absence of relevant schema, have to memorize the individual elements of the problem state as they attempt to move toward a solution. This process imposes a high cognitive load, thus reducing the likelihood of schema acquisition and thus reducing flexibility. Sweller and colleagues have identified numerous interventions that improve novice’s schema acquisition, including the use of worked-out examples (Sweller & Cooper, 1985), various manipulation of problem goals and subgoals (Mawer & Sweller, 1982), and the use of self-explanations (Mwangi & Sweller, 1998).

This literature on the role of schemata in flexible procedure use is closely related to the view that the presence of conceptual knowledge that is linked to one’s knowledge of procedures can account for flexibility. Numerous studies have demonstrated that increased conceptual knowledge is related to improved procedural performance (e.g., Byrnes, 1992; Byrnes & Wasik, 1991; Rittle-Johnson & Alibali, 1999; Rittle-Johnson, Siegler, & Alibali, 2001).

A second and distinct hypothesis explaining the development of flexibility is procedural practice. Through practice one can develop fluency in the use of procedural operators, which in turn may allow one to use these operators to flexibly solve new problems. Among experts, fluency and flexibility co-exist (Hatano & Inagaki, 1986), and certainly procedural practice plays a significant role in the development of expertise (Ericsson & Charness, 1994; Gagné, 1983). So it is worth considering whether procedural practice may contribute to the development of flexibility. In support of this idea, Kotovsky, Hayes, and Simon (1985) found that when participants are presented with challenging problems, those who have automated rules for solving similar problems are more likely to be successful problem solvers. Owen and Sweller (1989) propose that the automation of problem solving operators allows experts to attend fully to aspects of a problem that are not familiar, allowing for the use of more complex and innovative solution methods. Similarly, MacKay (1982) found increased flexibility with skill practice, and Cumming (1999) found the absence of automaticity related to weak performance on more complex tasks (see also Mayer & Wittrock, 1996).

However, there is also reason to believe that automation may be detrimental to the development of flexibility. With procedural practice, knowledge underlying skill execution becomes compiled and increasing unavailable for conscious access (Anderson, 1992; Neves &
Flexible problem solving may require accessing this unavailable compiled knowledge. Some research indicates that automation of skills has little do with problem solving performance (Rabinowitz & Woolley, 1995). Similarly, Renkl and Helmke (1992) found a dissociation between tasks promoting problem solving performance and those promoting skill fluency. And, in the area of reading comprehension, Spring and colleagues (1981) did not find a positive transfer relationship between automaticity and reading comprehension.

These two hypotheses for the development of flexibility, relating to schema acquisition and automaticity of skills, are explored in the present study. Two instructional interventions that are related to these explanations are experimentally evaluated. The first uses the “alternative ordering paradigm” (Star, 2001; Star & Seifert, 2002), where students generating and comparing multiple solutions to procedural problems during learning. As described below, the alternative ordering paradigm is hypothesized to aid in schema acquisition, resulting in greater flexibility. A second instructional intervention, in which some students viewed a demonstration of problem solving strategies, is also evaluated. Both of these interventions are described more below, as well as in the method section.

**Alternative ordering paradigm.** In the alternative ordering paradigm, participants are asked to re-solve previously completed problems but using a different ordering of transformations. In this task, students are not merely practicing the same solution over and over again, but instead are generating, comparing, and evaluating the effectiveness and efficiency of different solution procedures. The alternative ordering task is similar to what Klein has called the “flexibility-on-demand” task or FDT (Klein, 1998). The FDT was used by Klein as a way to assess students’ flexibility (Blöte et al., 2001; Klein, 1998); arithmetic learners were asked to solve a problem two times in a row using two different procedures. If a student produced two different solution procedures, this was taken as an indication of her flexibility.

In addition to its potential use as a way to assess flexibility (as intended by Klein, 1998), there is reason to speculate that tasks such as the alternative ordering paradigm or FDT may lead to greater flexibility. Owen and Sweller (1989) suggest that novices often rely upon high-load, means-end strategies when faced with unfamiliar problems; they continually evaluate and attempt to reduce the distance between the current problem state and the goal state. By reducing the specificity of the goal, Mawer and Sweller (1982) demonstrated that novices could be guided toward more expert, forward-search strategies, thus enhancing schema acquisition. The key feature of the alternative ordering task is that learners solve previously completed problems a second time, but using a different ordering of steps. Such a task requires that a learner accomplish two goals at the same time in her second solution attempt: first, she should complete the problem correctly; second, she should do so using a different ordering of transformations. Accomplishing both these goals simultaneously is hypothesized to have the same effect as reducing the specificity of the goal. Rather than exclusively focus on reducing the distance between the current problem state and the goal state, the learner must simultaneously manage two goals. The work of Sweller and colleagues (e.g., Mawer & Sweller, 1982) suggests that this can aid in schema acquisition and thus the development of greater understanding of the problem domain, which may lead to flexibility.

In addition, support for the hypothesis that the alternative ordering task might lead to greater flexibility can be found in the literature on self-explanations. There is a great deal of evidence that the generation of self-explanations under certain conditions facilitates learning and transfer (Chi, 2000; Chi & Bassok, 1989; Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Chi,
DeLeeuw, Chiu, & LaVancher, 1994; Chi & VanLehn, 1991; Pirolli & Recker, 1994; Renkl, 1997; Renkl, Stark, Gruber, & Mandl, 1998). Particular qualities of self-explanations have been found to be more strongly related to learning gains – qualities that the alternative ordering tasks also exhibits. First, a self-explanation was found to be especially beneficial to learning when it mentioned a purpose or goal for a particular action (Chi et al., 1989; Renkl, 1997; Renkl, Stark, Gruber, & Mandl, 1998). In other words, self-explanations are powerful when they make explicit the often-implicit goals of particular actions. This quality of self-explanations is consistent with the work of Catrambone (1994; 1995; 1996; Catrambone & Holyoak, 1989, 1990), who found that explicit encoding of subgoal-operator connections facilitates transfer. Ostensibly, self-explanations relating to subgoal-operator connections provide meaning or rationale for actions encountered in the problem or text and thus address gaps in the student’s representation of the problem. Second, self-explanations that display anticipative reasoning are more strongly linked to learning (Reimann, Schult, & Wichmann, 1993; Renkl, 1997). Anticipative reasoning refers to the generating of solutions or answers to problems or questions before such solutions are presented in the problem or text. Self-explanations that include anticipatory reasoning show that a learner is thinking and working ahead, providing the opportunity to compare the anticipated steps with the ones presented subsequently in the problem or text.

These two features of self-explanations are also present in the alternative ordering task. As students re-solve previously competed problems, they may come to know the various subgoals that are applicable for a particular transformation on certain problems (e.g., subgoal-operator connections) and may subsequently identify and repair gaps in their existing knowledge of the domain, perhaps leading to greater flexibility. Furthermore, as students become familiar with the alternative ordering task, they are aware that they will be asked to re-solve the same equation again, using a different ordering of steps. As a result, it is possible that students’ initial equation-solving strategies will be selected and implemented with an eye toward how the same equation would be approached the second time (e.g., anticipatory reasoning).

**Strategy demonstration.** In addition to the alternative ordering paradigm treatment described above, some participants in the present study also received supplemental instruction in the use of particular problem solving strategies. This additional manipulation was included in the present study for two reasons.

First, the inclusion of a direct strategy instruction condition addresses a critique of prior work (Star, 2001; Star & Seifert, 2002) on procedural flexibility, which failed to investigate this manipulation. Direct instruction on procedural strategies continues to be the norm in US mathematics classrooms (Schmidt et al., 1999), despite recent reform efforts recommending more student-centered approaches (National Council of Teachers of Mathematics, 1989, 2000). Numerous studies have found conceptual knowledge gains amongst students who are allowed to discover, generate, and compare multiple solution strategies at the elementary school level (e.g., Carpenter et al., 1998), but critiques of student-centered discovery methods continue to make a case for more teacher-centered approaches (e.g., Mayer, 2004). The present study evaluates the effects of direct strategy instruction to pure discovery on students’ flexibility.

Second, recent evidence suggests that, despite the promise of student-centered discovery approaches, there may be a place for direct instruction. Schwartz and Bransford (1998) found that 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 and Bransford (1998) emphasize that the
initial comparison of contrasting cases guides students’ attention to important features in a domain, leading to the development of differentiated domain knowledge. Once learners have sufficient differentiated knowledge in a domain, they are in a position to deeply understand an expository explanation that is presented to them, and thus may benefit from direct instruction. Schwartz and Bransford’s (1998) work suggests that an instructional routine optimal for the development of flexibility would begin with discovery-oriented problem solving, allowing students to develop differentiated knowledge of possible solution methods in a domain. Once basic knowledge of strategies existed, students may benefit from more explicit instruction on how to use methods strategically.

The present study assesses this prediction. After a period of initial discovery-oriented problem solving, some students were provided a brief period of expository instruction in strategy use. It was hypothesized that such a brief period of direct instruction would have a significant impact on students’ flexibility.

Research questions

The present study addressed the following research questions. First, does the alternative ordering paradigm lead to greater flexibility in the use of mathematics procedures? Second, does supplemental and direct instruction in certain problem solving strategies, given after students have developed sufficient domain knowledge, result in greater flexibility? Third, to what extent are gains in procedural flexibility, conceptual knowledge, and equation solving competence (e.g., ability to solve equations correctly) correlated?

Method

Participants and procedure

Rising 6th grade students (e.g., those who had just completed the 6th grade) from several suburban, middle class public school districts were recruited to participated in the present study; 134 (83 girls, 51 boys) volunteered to participate. All students attended one-hour classes for five consecutive days (Monday through Friday) during the summer. Class size ranged from 8 to 15 students; two researchers supervised each class. Parents’ scheduling preferences were used to assign students to classes, with attention also given to keeping the ratio of girls to boys in all classes approximately the same. Students sat individually at tables for the duration of each class; classes were held in a large seminar room on the campus of a local university. Students were given a $50 gift certificate to a local bookstore in return for their participation.

A pre-test was administered to all students during the first experimental session on Monday. Students were given 10 minutes for the pre-test; all students finished in the allotted time. The pre-test assessed students’ knowledge of formal linear equation solving techniques. Immediately following the pre-test, students received a scripted 30-minute benchmark lesson on linear equation-solving transformations. In this lesson, students were introduced to the four basic transformations used to solve linear equations: combining like variable or constant terms (COMBINE), using the distributive property (EXPAND), adding or subtracting a constant or variable term to both sides of an equation (MOVE), and multiplying or dividing both sides of an equation by a constant (DIVIDE). Students were not given any strategic instruction on how transformations could be chained together to solve an equation. Rather, the focus in instruction was strictly on pattern recognition: identifying which transformation could be used for particular patterns of symbols, and how that transformation was correctly applied. For example, students were shown the symbols, $2x + 3x$, and instructed that the COMBINE step allowed this
expression to be rewritten as $5x$. The intent in this brief period of instruction was to provide novice algebra learners with sufficient or prerequisite (Zhu & Simon, 1987) knowledge to enable them to begin to solve very straightforward equations. The lesson provided students both guided and independent practice at using each of the four transformations. At the conclusion of instruction, students were provided with an 8.5” by 11” card that summarized the instructed material. Students were told that they could refer to the card at any time during problem solving.

For the next three one-hour experimental sessions (Tuesday, Wednesday, and Thursday), students worked individually and at their own pace through a booklet containing 31 linear equations to be solved (see Table 1, below). The booklet was divided into three sections, and students completed only the problems in one section during each class. On Tuesday, students could complete a maximum of 11 problems; on Wednesday and Thursday, students could complete a maximum of 10 problems each day. Problems in a section that were not completed in a class were not returned to on the next day; if a student only solved 8 of the 11 problems on Tuesday, he/she was instructed to start on problem 12 on Wednesday.

To provide a record of how long each student worked on each problem, students wrote the time, in minutes and seconds, at the bottom of each page, immediately prior to turning the page to work on the next problem. Digital clocks with large displays were positioned throughout the room. If a student finished the day’s problems before the end of the class, he/she was instructed to close his/her booklet and sit quietly.

If a student became stuck while attempting a problem, he/she raised his/her hand and was approached by an experimenter. The experimenter answered the student’s questions in a semi-standardized format. Specifically, the experimenter corrected the student’s arithmetic mistakes (e.g., if the student multiplied 2 by 3 and got 5, the experimenter pointed out this error) or reminded him/her of the four possible transformations and how each was used. Experimenters never gave strategic advice to students, such as suggesting which transformation to apply next or whether one method of solution was any better than another method. Students independently came up with their own choices for which transformations to apply.

During the final class on Friday, students completed a post-test. The post-test contained all of the problems on the pre-test, as well as additional problems designed to assess students’ flexibility and conceptual understanding (see Materials, below). Students were given 40 minutes to take the post-test; all students finished in the allotted time.

Six months after the conclusion of the study, students who had given consent to be re-contacted were approached about taking a delayed post-test. 46 students volunteered to take the delayed post-test and were compensated with a $10 gift certificate to a local bookstore. Students sat alone at tables in a local cafe as they were administered a delayed post-test, which was identical to the post-test. (However, the delayed post-test did not include the four encoding items.) Students were given 40 minutes to take the delayed post-test; all students finished in the allotted time. At the conclusion of the post-test, each student was individually asked several yes/no questions about their current math class in 7th grade, whether they had solved equations in their math class or in any extracurricular program, and whether they had studied or been tutored in preparation for the delayed post-test.

**Materials**

**Problem booklet.** One goal of the present research was to investigate whether students learned multiple ways to solve equations, and, if so, whether they became aware that some strategies are more effective than others on certain problems. Equations in the problem booklet...
were chosen to provide students with the opportunity to use either a standard algorithm or two problem-specific atypical strategies.

For many linear equations, the standard algorithm is the most efficient solution strategy. Several problems of this type (labeled as “Standard” in Table 1) were included in the problem booklet, particularly on the first day. “Divide before expanding” problems were those where a more efficient solution involved dividing both sides by a constant before distributing. Similarly, on “change in variable” problems, a more efficient solution could be arrived at by making a change in variable, as described above. Both “divide before expanding” and “change in variable” problems could be solved using a standard algorithm, but choosing a less typical approach could result in a more efficient solution method.

**Table 1**: Equations attempted during experimental sessions

<table>
<thead>
<tr>
<th>Day</th>
<th>#</th>
<th>Problem</th>
<th>Problem type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tuesday</td>
<td>1</td>
<td>(x + 4 = 7)</td>
<td>(Initial mastery)</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>(2x = 12)</td>
<td>(Initial mastery)</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>(2x + 4 = 10)</td>
<td>(Initial mastery)</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>(3x + 5x = 16)</td>
<td>(Initial mastery)</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>(5x + 3 = 2x + 6)</td>
<td>(Initial mastery)</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>(3(x + 1) = 15)</td>
<td>Divide before expanding</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>(4(x + 3) = 8x)</td>
<td>Standard</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>(3(x + 2) = 5x)</td>
<td>Standard</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>(4(x + 1) = 2(x + 1))</td>
<td>Change in variable</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>(2(x + 5) + 4x = 4 + 8x)</td>
<td>Standard</td>
</tr>
<tr>
<td></td>
<td>11*</td>
<td>(3(x + 2) + 9x = 3x + 15)</td>
<td>Standard</td>
</tr>
<tr>
<td>Wednesday</td>
<td>12</td>
<td>(2(x + 1) = 10)</td>
<td>Divide before expanding</td>
</tr>
<tr>
<td></td>
<td>13*</td>
<td>(3(x + 2) = 15)</td>
<td>Divide before expanding</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>(2(x + 3) + 4(x + 3) = 24)</td>
<td>Change in variable</td>
</tr>
<tr>
<td></td>
<td>15*</td>
<td>(6(x + 2) + 3(x + 2) = 27)</td>
<td>Change in variable</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>(2(x + 5) = 4(x + 5))</td>
<td>Change in variable</td>
</tr>
<tr>
<td></td>
<td>17*</td>
<td>(3(x + 3) = 5(x + 3))</td>
<td>Change in variable</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>(3(x + 2) + 6(x + 2) + 9(x + 2) = 30x + 6x)</td>
<td>Change in variable</td>
</tr>
<tr>
<td></td>
<td>19*</td>
<td>(5(x + 1) + 10(x + 1) + 5(x + 1) = 30x + 10x)</td>
<td>Change in variable</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>(2(x + 4) = 14)</td>
<td>Divide before expanding</td>
</tr>
<tr>
<td></td>
<td>21*</td>
<td>(4(x + 1) = 12)</td>
<td>Divide before expanding</td>
</tr>
<tr>
<td>Thursday</td>
<td>22</td>
<td>(3(x + 2) = 6(x + 2))</td>
<td>Change in variable</td>
</tr>
<tr>
<td></td>
<td>23*</td>
<td>(4(x + 3) = 2(x + 3))</td>
<td>Change in variable</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>(3(x + 1) = 15)</td>
<td>Divide before expanding</td>
</tr>
<tr>
<td></td>
<td>25*</td>
<td>(2(x + 3) = 12)</td>
<td>Divide before expanding</td>
</tr>
<tr>
<td></td>
<td>26</td>
<td>(9(x + 2) + 3(x + 2) + 3 = 18x + 9)</td>
<td>Change in variable</td>
</tr>
<tr>
<td></td>
<td>27*</td>
<td>(2(x + 3) + 4(x + 3) + 4 = 8x + 6)</td>
<td>Change in variable</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>(2(x + 1) + 6(x + 1) = 4(x + 1))</td>
<td>Change in variable</td>
</tr>
<tr>
<td></td>
<td>29*</td>
<td>(3(x + 3) + 6(x + 3) = 5(x + 3))</td>
<td>Change in variable</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>(2(x + 5) = 14)</td>
<td>Divide before expanding</td>
</tr>
<tr>
<td></td>
<td>31*</td>
<td>(5(x + 3) = 20)</td>
<td>Divide before expanding</td>
</tr>
</tbody>
</table>

*Indicates problems where treatment students engaged in the alternative ordering paradigm

In addition to these three types of problems (“standard,” “divide before expanding,” and “change in variable”), the first five problems were included to assist students in gaining initial mastery of the equation solving transformations in the context of solving equations. Prior work has indicated that students need an opportunity to develop basic fluency in the use of equation solving transformations before more sophisticated strategy discovery could occur (Star, 2001). Also, while completing the first five problems, it was intended that students would become familiar with other features of problem solving in this domain, including the fact that a problem is completely solved when one arrives at the line with the form \(x = a\).

Pre- and post-assessments. The pre-test and post-tests contained a number of types of problems. Most were modified versions of questions used on previous assessments of students’ flexibility (Star, 2001), while a few were new to the present study. The pre-test had 18 problems (problems 1 to 18); the post-tests contained these 18 problems, plus 12 additional problems (problems 19 to 30). (See Table 2 below.)

Table 2: Problems on the pre- and post-assessments

<table>
<thead>
<tr>
<th>Problem type</th>
<th>Problems</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Encoding</td>
<td>1, 3 (ill-formed), 2, 4 (well-formed)</td>
<td>1. (2+5x192) 2. (2x+15=29) 3. ((2x+15)(2x+15)=x6+(2) 4. (3(x + 2) + 9(x + 2) = 6(x + 2))</td>
</tr>
<tr>
<td>Equations to be solved: Familiar</td>
<td>5, 7, 8, 9</td>
<td>8. (4(x + 1) = 8(x + 1)) 9. (3(x + 2) + 9(x + 2) = 6(x + 2))</td>
</tr>
<tr>
<td>Equations to be solved: Transfer</td>
<td>6, 10, 11, 12, 13</td>
<td>6. (2(x + 1) + 3x + 4 = 8 + 3x + 4) 10. (0.2x + 0.3 = 0.9) 11. (3x + x + 5 = 13) 12. (2(3x + 2) + 2x = 20) 13. (2(2x + 3x) + 2 = 12)</td>
</tr>
<tr>
<td>Conceptual knowledge</td>
<td>16, 17, 18</td>
<td>16. Without using any of the solving steps, when is (31(x + 27) = 121?) 17. Here are two equations: (213x + 476 = 984) (213x + 476 + 4 = 984 + 4) Without using any of the solving steps, what can you say about the answers to these equations? 18. Solve this equation for (x): (2x + 3 + a = 9 + a)</td>
</tr>
<tr>
<td>Manipulation check</td>
<td>19, 20</td>
<td>19. Solve this equation in two different ways: (3(x + 1) + 6(x + 1) = 18)</td>
</tr>
<tr>
<td>Flexibility A:</td>
<td>14, 15</td>
<td>14. When I work on equations like the ones I did this week, I always try to begin all equations that I solve the same way, with the same first step. 15. When I work on equations like the ones I did on this week, I always try to solve each one the same way, using all of the same steps, in the same order.</td>
</tr>
<tr>
<td>Flexibility B:</td>
<td>21, 22, 23</td>
<td>22. Figure out ALL possible steps that the student could do in the NEXT step of this equation. (2(x + 3) + 6 + 4x + 8 = 4(x + 2) + 6x + 2x)</td>
</tr>
<tr>
<td>Flexibility C:</td>
<td>24, 25, 26, 27</td>
<td>25. Use the COMBINE step to combine the (2(x + 1)) and (5(x + 1)) terms in the equation below, and write what you get below.</td>
</tr>
</tbody>
</table>

2(x + 1) + 5(x + 1) = 14

Flexibility D: 28, 29, 30

28. Below are the beginning steps of how a student tried to solve several equations. Look at the way that this student started each equation and answer the questions below.
3(x + 2) = 12
x + 2 = 4
(a) In the part of the solution shown above, what solving step did the student use to get from the first line to the second line?
(b) Looking at the problem shown above, do you think that this way of starting to do this problem is a good idea?

Problems 1 to 4 were encoding items, designed to investigate whether increased domain knowledge of equation solving would lead to a greater ability to remember well-formed equations. Larkin (1989), drawing from earlier work by deGroot (1966; 1965) with chess players, predicts that increased domain knowledge results in an improved ability to ‘see’ a hierarchical and organized structure when one exists. More specifically, when shown both ill-formed and well-formed strings of algebra symbols, Larkin (1989) would predict that experts are more likely to recall the well-formed strings than the ill-formed strings. Novices, who can’t tell the difference between well-formed and ill-formed strings, would not show any differences in recall. In essence, the differential ability to recall well-formed strings as opposed to ill-formed strings is a measure of schema acquisition. Those with greater domain knowledge would be expected to have more sophisticated schemata.

To test this prediction, students were instructed that they would be shown some symbols on a screen for a few seconds, and they were to try to remember exactly what they saw. They were not allowed to begin writing until the symbols were erased. A string of symbols corresponding to a well-formed or an ill-formed equation was shown for four seconds, and students were given ten seconds to recall the string. Training occurred on one practice item, the string, “(I ❤ Math !!! ✓)”. Problems 1 and 2 were the equation 2x + 15 = 29 and a randomly generated, ill-formed arrangement of these symbols, “2+=5x192”. Problem 3 and 4 used the equation, 3(x + 2) + 9(x + 2) = 6(x + 2) and a randomly generated, ill-formed arrangement of these same symbols, “3(+2x(9x+)2+=)x6+(2)”. In both cases, the ill-formed string was presented first (problems 1 and 3).

Problems 5 to 13 were equations that students were asked to solve. Four of these problems, problems 5, 7, 8, and 9, were familiar to students, as they were isomorphic to equations in the problem booklet. Problem 5 was an initial mastery problem, 7 was a “divide before expanding” problem, and 8 and 9 were “change of variable” equations.

Problems 6, 10, 11, 12, and 13 were unlike any booklet problems and served as a measure of equation solving transfer. All five problems could be solved using the standard algorithm, but each problem either had an unfamiliar feature and/or could be solved using an atypical solution strategy (other than “divide before expanding” or “change of variable”). Problem 6 (see Table 1 above) was taken from a previous study (Star, 2001); a particularly efficient strategy for solving this equation is to MOVE or “cancel” the common 3x + 4 terms on both sides of the equation as a first step. Problem 10 was the only instance where students saw decimal coefficients; an efficient strategy is to multiply (using the DIVIDE transformation) the equation by 10 as a first step. Problem 11 has a variable term with coefficient 1 that needs to be combined, (COMBINE 3x + x to yield 4x), a feature not previously seen by students on booklet problems. Problem 12
has a variable term with coefficient other than 1 that needs to be distributed (EXPAND 2(3x + 2) to yield 6x + 4), which was also not previously seen. And problem 13 has two variable terms that may need to be distributed or combined, (EXPAND 2(2x + 3x) to yield 4x + 6x, or COMBINE 2(2x + 3x) to yield 2(5x)), which again was not previously seen on booklet problems.

Problems 16, 17, and 18 were designed to assess students’ conceptual knowledge of equations and equation solving, including the concept of an equation (that an equation is true for some but not all values of a variable) and the nature of the equals sign (that the quantities on either side of an equal sign are equal and remain so when the same number is added to both sides).

Problems 19 and 20 served a manipulation check; students were given an equation and asked to solve it in two different ways. Students who had participated in the alternative ordering paradigm had been asked to complete similar tasks numerous times, while control class students had not.

The remaining 12 problems assessed students’ flexibility in various ways. Problems 14 and 15 were multiple choice questions that assessed whether students believed (mistakenly) that all solution strategies used to solve linear equations begin in exactly the same way or are identical. Problems 21 to 23 gave students an equation and asked them to identify all possible transformations that could be done as a next step. For example, given the equation 15(x + 3) + 5(x + 3) = 10(x + 3), students were asked to indicate which of the COMBINE, MOVE, DIVIDE, and/or EXPAND transformations could be done as a next step. (For this problem, all four transformations can be done next.) Problems 24 to 27 gave students an equation and asked them to perform a transformation in an atypical way. For example, students were given the equation 2(x + 1) + 5(x + 1) = 14 and asked to use the COMBINE step to combine 2(x + 1) and 5(x + 1). Finally, problems 28 to 30 showed a student using a transformation atypically and asked the student to identify what transformation had been used. For example, students were asked which transformation had been applied to change the equation 3(x + 2) + 4(x + 2) = 14, to 7(x + 2) = 14. After selecting the transformation, students were asked whether this strategy seemed to be a good strategy for solving an equation or not.

It is possible that some problems on the post-test could help students to complete other problems. To address this possibility, the order of the flexibility problems was carefully chosen. In addition, students were only allowed to work forwards on the post-test; once a page was turned, they were prohibited from turning back to that page and looking at or reworking previous problems.

Design

The design of this study was 2 (Treatment: alternative ordering paradigm vs. control) by 2 (Strategy instruction: demonstration provided vs. no demonstration provided). Students were randomly assigned to each condition, as described below.

Treatment. At the conclusion of the first experimental session on Monday, each class was randomly assigned as either a treatment class or a control class. The treatment and control classes differed only in that students in the treatment class completed alternative ordering tasks, while students in the control class did not. On alternative ordering tasks, students were given problems that they had previously solved and asked to re-solve them, but using a different ordering of transformations. On problems where treatment students were asked to provide an alternative solution, control students completed a different but isomorphic problem. For example, both treatment and control students were asked to solve the problem, 4x + 10 = 2x + 16. Students in
the treatment class were given this same problem again and were asked to solve it using a different ordering of transformations. The students in the control class were given a structurally equivalent problem, $6x + 9 = 3x + 12$, instead.

Treatment students engaged in the alternative ordering paradigm regularly. Tuesday’s problems included only one problem where students re-solved a previously completed equation; problem 11 was a repeat of problem 10. However, beginning on Wednesday and problem 12, every odd-numbered problem was a repeat of the previous even-numbered problem, for treatment students. Treatment students were given no special instruction or examples in preparation for completing alternative ordering problems; they were only provided with the following instruction prior to starting work each day: “On some problems, you will be asked to work on a problem that you’ve already done before, and you’ll be asked to try to solve it using a different ordering of steps.”

Strategy instruction. At the conclusion of Tuesday’s experimental session, classes were randomly assigned to one of two strategy instruction conditions: strategy demonstration provided vs. no strategy demonstration provided. In classes where a strategy demonstration was provided, during the first eight minutes of Wednesday’s class, an experimenter showed students three worked-out examples of equations. To introduce his presentation, the experimenter noted that everyone had done well on the previous day’s problems, but as the problems got harder, he suggested that it might help some students to see how he (the experimenter) solved some equations.

The experimenter then solved three equations on a blackboard, two of which were taken from problems previously seen during Tuesday’s session: $2(x + 5) + 4x = 4 + 8x$ (which was problem 10 in students’ booklets and a “standard” equation), $3(x + 1) = 15$ (which was problem 6 in students’ booklets and a “divide before expanding” equation), and $3(x + 1) + 2(x + 1) = 20$ (which is isomorphic to problem 14, which students had not yet seen, and which is a “change in variable” equation). During this demonstration, the experimenter solved each problem using the strategy that matched each problem class (e.g., solving the “change in variable” equation by implementing a change in variable) but did not refer to any strategies by name.

Students observed but were not allowed to take notes during the presentation of worked-out examples. The experimenter framed his solutions by saying, “This is the way that I solve this equation.” Students were not offered any qualitative judgments about the experimenter’s solution methods (e.g., that they were good, efficient, or better than other strategies). The experimenter ended his presentation by telling the class that there were many different ways that an equation could be solved, and that, “Maybe you’ll find it helpful to see how I solved each of these problems.”

Analysis

Students’ solutions to the pre- and post-assessments were graded by two independent analysts, who subsequently met to resolve all disagreements.

Results

Prior knowledge and assignment to condition

Pre-test scores were examined to confirm that participants began the study with little or no prior knowledge of formal equation-solving procedures. Mean pre-test scores were low ($M = 12\%$, $SD = .167$); most students who were able to correctly solve one or more equations did so using informal methods such as guess-and-check or unwinding.
To investigate possible bias in random assignment to condition, a 2 (Treatment: alternative ordering paradigm vs. control) by 2 (Strategy instruction: demonstration provided vs. no demonstration provided) ANOVA was conducted on the overall pre-test score, as well as pre-test measures of encoding, equation solving, and conceptual knowledge. There were no significant main or interactions effects on any of the measures, indicating that random assignment to condition was done without bias.

**Encoding**

A 2 (Test administration: pre-test vs. post-test) by 2 (Treatment: alternative ordering paradigm vs. control) by 2 (Strategy instruction: demonstration provided vs. no demonstration provided) repeated measures ANOVA was conducted on the encoding items to investigate whether students’ ability to recall ill-formed and well-formed symbols strings improved from pre-test to post-test as a result of increased domain knowledge. There was a significant main effect for test administration, F(1,128) = 66.815, p < .001, indicating that mean scores on post-test encoding items were higher than on the pre-test (see Table 3). The data indicates that post-test gains came as a result of improved performance on the well-formed items, while memory for ill-formed items was unchanged. As a result of increased knowledge in the domain of equation solving, students were more successful in remembering strings of symbols that looked like an equation than they were random arrangements of symbols.

There were no other significant main effects or interactions for the encoding items, suggesting that neither the treatment nor the strategy instruction impacted students’ ability to encode ill-formed and well-formed symbol strings.

<table>
<thead>
<tr>
<th>Table 3: Mean percentage correct (and SD) on encoding items</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>No strategy demonstration provided</td>
</tr>
<tr>
<td>All (n = 63)</td>
</tr>
<tr>
<td>Treatment (n = 31)</td>
</tr>
<tr>
<td>Control (n = 32)</td>
</tr>
<tr>
<td>Strategy demonstration provided</td>
</tr>
<tr>
<td>All (n = 69)</td>
</tr>
<tr>
<td>Treatment (n = 35)</td>
</tr>
<tr>
<td>Control (n = 34)</td>
</tr>
<tr>
<td>All Treatment (n = 66)</td>
</tr>
<tr>
<td>All Control (n = 66)</td>
</tr>
<tr>
<td>All conditions (n = 132)</td>
</tr>
</tbody>
</table>

Equation solving

To investigate students’ equation solving performance, a 2 (Test administration: pre-test vs. post-test) by 2 (Treatment: alternative ordering paradigm vs. control) by 2 (Strategy instruction: demonstration provided vs. no demonstration provided) repeated measures ANOVA was conducted on all equation solving variables. Similar 2 x 2 x 2 repeated measures ANOVAs were conducted for the delayed post-test, comparing post-test scores with delayed post-test scores and pre-test scores with delayed post-test scores.

Performance on all equations. Combining both familiar and transfer equation solving, there was a significant main effect for test administration, $F(1,128) = 448.27, p < .001$, indicating that students were more successful solving equations on the post-test as compared to the pre-test (see Table 4). There were no other significant main effects or interactions. Mean percentage correct jumped across all conditions from 13% on pre-test to 65% on post-test.

Table 4: Mean percentage correct (and SD) on equation solving measures

<table>
<thead>
<tr>
<th></th>
<th>Equation solving – All</th>
<th>Equation solving – Familiar</th>
<th>Equation solving – Transfer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre-test</td>
<td>Post-test</td>
<td>*Delayed Post-test</td>
</tr>
<tr>
<td>No strategy demonstration provided</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All (n = 63)</td>
<td>15 (0.24)</td>
<td>68 (0.3)</td>
<td>53 (0.31)</td>
</tr>
<tr>
<td>Treatment (n = 31)</td>
<td>14 (0.25)</td>
<td>64 (0.3)</td>
<td>47 (0.35)</td>
</tr>
<tr>
<td>Control (n = 32)</td>
<td>17 (0.24)</td>
<td>72 (0.3)</td>
<td>58 (0.27)</td>
</tr>
<tr>
<td>Strategy demonstration provided</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All (n = 69)</td>
<td>12 (0.21)</td>
<td>63 (0.3)</td>
<td>47 (0.32)</td>
</tr>
<tr>
<td>Treatment (n = 35)</td>
<td>11 (0.2)</td>
<td>57 (0.33)</td>
<td>42 (0.35)</td>
</tr>
<tr>
<td>Control (n = 34)</td>
<td>12 (0.22)</td>
<td>69 (0.26)</td>
<td>52 (0.3)</td>
</tr>
<tr>
<td>All Treatment (n = 66)</td>
<td>12 (0.22)</td>
<td>61 (0.31)</td>
<td>44 (0.35)</td>
</tr>
<tr>
<td>All Control (n = 66)</td>
<td>14 (0.23)</td>
<td>70 (0.28)</td>
<td>54 (0.28)</td>
</tr>
<tr>
<td>All conditions (n = 132)</td>
<td>13 (0.23)</td>
<td>65 (0.3)</td>
<td>50 (0.31)</td>
</tr>
</tbody>
</table>

* 46 students took the delayed post-test: 21 in treatment classes (9 in no strategy demonstration classes, and 12 in strategy demonstration classes), and 25 in control classes (11 in no strategy demonstration classes, and 14 in strategy demonstration classes).

These gains in equation solving performance dropped from the post-test to the delayed post-test (but remained significantly higher than at pre-test, $F(1,142) = 82.745, p < .001$), with a significant main effect for test administration, $F(1,41) = 27.28, p < .001$. There were no other
main effects or interactions on the delayed post-test analysis. Students’ mean percentage correct dropped from 65% on the post-test to 50% on the delayed post-test.

Familiar equations. Results indicate that students showed significant improvement in their performance on familiar equations, $F(1,128) = 343.57, p < .001$. Students correctly solved an average of 73% of familiar equations correctly, an improvement from 18% on the pre-test.

In addition, there was a significant interaction between test administration and treatment, $F(1,128) = 4.086, p < .05$. While both treatment and control class students scored the same at pre-test (17% and 18% correct, respectively), control class students showed greater gains at post-test (79% correct, as compared to 67% correct for treatment class students).

Most of these gains in familiar equation solving persisted six months later; scores on the delayed post-test were significantly higher than the pre-test ($F(1,42) = 48.006, p < .001$), but significantly lower than the post-test, $F(1,41) = 38.371, p < .001$. Control class students’ scores on familiar equations on the delayed post-test were only marginally higher than treatment class students, $F(1,41) = 3.350, p = .074$.

Transfer equations. Similar results were found on transfer equation performance. Students showed improvement from pre-test to post-test in their ability to solve transfer equations, $F(1,128) = 339.387, p < .001$. Overall performance improved from 10% correct on the pre-test to 59% correct on the post-test. Students lost some of these gains six months later, dropping to 51% correct on the delayed post-test. Students’ delayed post-test scores was significantly lower than the post-test scores, $F(1,41) = 8.265, p < .01$, but remained significantly higher than pre-test scores, $F(1,42) = 73.401, p < .001$. There were no other significant main effects or interactions.

Summary, equation solving. Students became better equation solvers as a result of their participation in the present study. On both familiar and transfer equations, problem-solving performance improved significantly from pre-test to post-test, with some but not all gains lost six months after the end of the study. Control class students did better on familiar equations on the post-test, but this difference was reduced by the time of the delayed post-test. Providing a demonstration of strategies had no impact on students’ ability to successfully solve equations. Also, re-solving previously completed problems had no effect on students’ performance on transfer equations.

Conceptual knowledge

Recall that conceptual knowledge of equations was assessed through problems 16, 17, 18 on the pre- and post-assessments. To investigate whether students’ conceptual knowledge in this domain improved as a result of participation in the study, a 2 (Test administration: pre-test vs. post-test) by 2 (Treatment: alternative ordering paradigm vs. control) by 2 (Strategy instruction: demonstration provided vs. no demonstration provided) repeated measures ANOVA was conducted. Similar 2 x 2 x 2 repeated measures ANOVAs were conducted for the delayed post-test, comparing post-test scores with delayed post-test scores and pre-test scores with delayed post-test scores.

There was a significant main effect for test administration; students did show large gains in conceptual knowledge from the pre-test ($M = 6\%, SD = 0.16$) to the post-test ($M = 45\%, SD = 0.31$), $F(1,128) = 205.07, p < .001$. (See Table 5.) Interestingly, conceptual knowledge scores not only held six months after the study but also actually improved. On the delayed post-test, students’ scores improved on average to 55% correct, a difference that was significantly higher than at post-test, $F(1,41) = 6.850, p < .05$.  

Table 5: Mean percentage correct (and SD) on conceptual knowledge items

<table>
<thead>
<tr>
<th></th>
<th>Pre-test</th>
<th>Post-test</th>
<th>*Delayed Post-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>All (n = 63)</td>
<td>8 (0.2)</td>
<td>49 (0.3)</td>
<td>60 (0.34)</td>
</tr>
<tr>
<td>Treatment (n = 31)</td>
<td>8 (0.21)</td>
<td>48 (0.24)</td>
<td>70 (0.35)</td>
</tr>
<tr>
<td>Control (n = 32)</td>
<td>9 (0.19)</td>
<td>49 (0.35)</td>
<td>52 (0.31)</td>
</tr>
<tr>
<td>All (n = 69)</td>
<td>3 (0.12)</td>
<td>42 (0.32)</td>
<td>51 (0.34)</td>
</tr>
<tr>
<td>Treatment (n = 35)</td>
<td>2 (0.08)</td>
<td>38 (0.33)</td>
<td>33 (0.25)</td>
</tr>
<tr>
<td>Control (n = 34)</td>
<td>5 (0.15)</td>
<td>46 (0.31)</td>
<td>67 (0.35)</td>
</tr>
<tr>
<td>All Treatment (n = 66)</td>
<td>5 (0.15)</td>
<td>43 (0.3)</td>
<td>49 (0.34)</td>
</tr>
<tr>
<td>All Control (n = 66)</td>
<td>7 (0.17)</td>
<td>47 (0.33)</td>
<td>60 (0.33)</td>
</tr>
<tr>
<td>All conditions (n = 132)</td>
<td>6 (0.16)</td>
<td>45 (0.31)</td>
<td>55 (0.34)</td>
</tr>
</tbody>
</table>

* 46 students took the delayed post-test: 21 in treatment classes (9 in no strategy demonstration classes, and 12 in strategy demonstration classes), and 25 in control classes (11 in no strategy demonstration classes, and 14 in strategy demonstration classes).

There were no other significant main effects of interactions from pre-test to post-test. However, in examining the differences between students’ scores on the post-test and the delayed post-test, there was a significant Treatment by Strategy instruction interaction, F(1,41) = 6.555, p < .05. When no strategy demonstration was provided, treatment class students showed relatively large gains in conceptual knowledge from post-test (M = 48%, SD = 0.24) to delayed post-test (M = 70%, SD = 0.35). In contrast, when strategy demonstration was provided, control class students showed relatively large gains in conceptual knowledge from post-test (M = 46%, SD = 0.31) to delayed post-test (M = 67%, SD = 0.35). This interesting interaction should be interpreted with caution, consider the low alpha reliability of the three-item conceptual knowledge measure, which was 0.27 for the post-test and 0.42 for the delayed post-test.

Flexibility

Fourteen questions were included on the post-test and delayed post-test to assess students’ flexibility; two of these questions, problems 14 and 15, also appeared on the pre-test. It was assumed that students with little or no prior knowledge of linear equation solving strategies would not be flexible at pre-test; as evidence supporting this assumption, students’ scores on problems 14 and 15 were not significantly above chance. As a result, instead of testing for pre/post learning gains, ANCOVA was used to assess differences by condition on the post-test and delayed post-test. To control for pre-existing differences in knowledge of equation solving strategies, the pre-test measure of percent of all equations solved correctly was used as a covariate in all analyses.
With pre-test equation solving as a covariate, a 2 (Treatment: alternative ordering paradigm vs. control) by 2 (Strategy instruction: demonstration provided vs. no demonstration provided) ANCOVA was conducted on post-test scores to investigate condition differences in students’ flexibility. There were significant main effects for treatment, F(1,132) = 8.821, p < .01, and for strategy instruction, F(1,132) = 7.965, p < .01 (see Table 6). Treatment students were more flexible than control class students, with mean scores on flexibility items of 54% and 45%, respectively. And students who were provided a strategy demonstrated became more flexible than those who were not, with mean scores of 53% and 45% respectively on flexibility items. But the effects of the treatment and the strategy demonstration on students’ flexibility did not appear to cumulative, as there was no significant interaction.

Table 6: Mean percentage correct (and SD) on flexibility items

<table>
<thead>
<tr>
<th></th>
<th>All (n = 63)</th>
<th>Post-test</th>
<th>*Delayed Post-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>No strategy demonstration provided</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment (n = 31)</td>
<td></td>
<td>51 (0.17)</td>
<td>39 (0.22)</td>
</tr>
<tr>
<td>Control (n = 32)</td>
<td></td>
<td>39 (0.18)</td>
<td>46 (0.24)</td>
</tr>
<tr>
<td>Strategy demonstration provided</td>
<td>All (n = 69)</td>
<td>53 (0.25)</td>
<td>42 (0.25)</td>
</tr>
<tr>
<td>Treatment (n = 35)</td>
<td></td>
<td>57 (0.27)</td>
<td>42 (0.31)</td>
</tr>
<tr>
<td>Control (n = 34)</td>
<td></td>
<td>50 (0.23)</td>
<td>42 (0.2)</td>
</tr>
<tr>
<td>All Treatment (n = 66)</td>
<td></td>
<td>54 (0.23)</td>
<td>41 (0.27)</td>
</tr>
<tr>
<td>All Control (n = 66)</td>
<td></td>
<td>45 (0.21)</td>
<td>43 (0.21)</td>
</tr>
<tr>
<td>All conditions (n = 132)</td>
<td></td>
<td>49 (0.23)</td>
<td>42 (0.24)</td>
</tr>
</tbody>
</table>

* 46 students took the delayed post-test: 21 in treatment classes (9 in no strategy demonstration classes, and 12 in strategy demonstration classes), and 25 in control classes (11 in no strategy demonstration classes, and 14 in strategy demonstration classes).

However, on the delayed post-test, these gains both for treatment and strategy instruction largely disappeared. A 2 (Treatment: alternative ordering paradigm vs. control) by 2 (Strategy instruction: demonstration provided vs. no demonstration provided) ANCOVA, with pre-test equation solving as a covariate, showed no significant main effects or interactions. When post-test flexibility was included as an additional covariate, there was a marginally significant treatment main effect, F(1,46) = 3.888, p = .056, indicating that treatment class students showed a greater drop from the post-test to the delayed post-test.

The 14 flexibility items, taken together, were a reliable measure; alpha reliability was 0.76 on the post-test and 0.80 on the delayed post-test.
Effects of outside domain exposure on delayed post-test scores

In the six months following the study (and between the post-test and delayed post-test administration), the then-7th grade participants may have received algebra instruction in their mathematics class and/or received extracurricular mathematics tutoring (e.g., from parents or after school classes); any of these outside influences could have influenced results of the delayed post-test analysis. To investigate this possibility, students were asked a series of questions immediately after taking the post-test to ascertain whether they had exposure to linear equation solving since the end of the summer study. 67% of students who took the delayed post-test indicated that they did have some exposure to problem solving in this domain since the summer.

To investigate whether students who had outside exposure to linear equation solving performed better on the delayed post-test, a 2 × 2 × 2 ANCOVA was conducted on all delayed post-test variables (equation solving, familiar equation solving, transfer equation solving, conceptual knowledge, and flexibility), with pre-test equation solving score as a covariate. There were several significant main effects for outside exposure. Students who had been exposed to equation solving outside of the study context scored significantly higher on delayed post-test measures of equation solving, F(1, 46) = 6.365, p < .05; transfer equation solving, F(1, 46) = 9.409, p < .01; and flexibility, F(1, 46) = 5.539, p < .05. (See Table 7, below.) On measures for familiar equation solving and conceptual knowledge, while those with outside exposure scored higher than those without, the difference was not significant in either case.

Table 7: Mean percentage correct (and SD), delayed post-test measures, by outside exposure

<table>
<thead>
<tr>
<th>Measure</th>
<th>Outside exposure to equation solving</th>
<th>No outside exposure to equation solving</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equation solving – All *</td>
<td>59 (.30)</td>
<td>30 (.28)</td>
</tr>
<tr>
<td>Equation solving – Familiar</td>
<td>54 (.32)</td>
<td>33 (.26)</td>
</tr>
<tr>
<td>Equation solving – Transfer **</td>
<td>63 (.32)</td>
<td>28 (.32)</td>
</tr>
<tr>
<td>Conceptual knowledge</td>
<td>61 (.33)</td>
<td>42 (.32)</td>
</tr>
<tr>
<td>Flexibility *</td>
<td>49 (.23)</td>
<td>28 (.18)</td>
</tr>
</tbody>
</table>

* p < .05; ** p < .01; n = 46

Despite this clear effect for those students with outside exposure to equation solving, there were no significant interactions involving the effect of outside influences. In other words, while students who were exposed to equation solving between the post-test and the delayed post-test did perform better on many sections of the delayed post-test, these gains were not concentrated in any one treatment group or strategy instruction group, which could have potentially biased the delayed post-test results.

(Note that there was a significant two-way interaction between strategy instruction and treatment on conceptual knowledge, F(2, 46) = 6.246, p < .01. This is a similar interaction as the one described above for conceptual knowledge. When no strategy demonstration was provided,

treatment class students outperformed control class students on the delayed post-test. When strategy demonstrated was provided, control class students did better on this measure.)

**Summary of results**

The present study found that students became better equation solvers as a result of their participation. On both familiar and transfer equations, problem-solving performance improved significantly from pre-test to post-test, with some but not all gains lost six months after the end of the study. Control class students did better on familiar equations on the post-test, but this difference was reduced by the time of the delayed post-test. Providing a demonstration of strategies had no impact on students’ ability to successfully solve equations. Also, the alternative ordering paradigm had no effect on students’ performance on solving equations.

Students’ ability to encode symbolic strings of well-formed equations (as a measure of their domain schemata) improved as a result of equation solving experience. In addition, students’ conceptual knowledge of equations improved from the pretest to the posttest. Interestingly, further gains in conceptual knowledge occurred in the six months prior to the study, with delayed posttest scores higher than posttest scores. And both the alternative ordering tasks and the strategy demonstration improved students’ flexibility. However, these differences largely disappeared in the six months after the study and did not show up on the delayed posttest.

About two-thirds of students had additional exposure to equation solving problems between the post-test and the delayed post-test administration. As would be expected, those with additional exposure outperformed those with no exposure on measures of equation solving and flexibility. However, students with additional exposure were evenly distributed amongst all conditions of the study, thus avoiding possible biasing of the delayed post-test results.

**Discussion**

The present research explored the following issues. First, does the alternative ordering paradigm lead to greater flexibility? Second, does well-timed direct instruction on problem solving strategies result in greater flexibility? And third, what is the relationship between gains in flexibility, conceptual knowledge, and equation solving competence?

With respect to the first issue, this study contributed to a growing body of evidence (Star, 2001; Star & Seifert, 2002) that the alternative ordering paradigm significantly improves students’ flexibility in equation solving. By re-solving previously completed problems but using a different ordering of steps, treatment students became knowledgeable about multiple ways that linear equations could be solved and demonstrated the ability to select appropriate procedures for particular problems.

With respect to the second research question, an extremely short (eight minute) demonstration of problem solving methods, provided after students had the opportunity to develop basic proficiency, had significant effects on students’ flexibility. Students who received this strategy instruction became more flexible than those who did not. In fact, this kind of well-timed direct instruction was a much more efficient means of improving students’ flexibility, as those receiving eight minutes of strategy instruction became as flexible as those participating in the alternative ordering paradigm during the entire study. These results are consistent with those of Schwartz and Bransford (Schwartz & Bransford, 1998).

The present study suggests that the relationship between flexibility, conceptual knowledge, and equation solving competence is quite complex. On average, students in all conditions showed gains in conceptual knowledge, as evidenced by improved scores on the conceptual knowledge measures as well as the encoding items. Neither treatment enhanced these
conceptual knowledge gains, despite the fact that both treatments improved students’ flexibility. It appears that improvements in equation solving competence were correlated with gains in conceptual knowledge, but neither of these necessarily led to greater flexibility. In other words, this study did not provide evidence of flexibility being tightly correlated or linked with conceptual knowledge.

Further evidence of the complex relationship between solving competence, conceptual knowledge, and flexibility was found in two other aspects of the present results. First, consider students’ post-test scores six months after the study ended. For students who claimed to have had no outside exposure to linear equation solving since the study, measures of solving competence and flexibility dropped, while conceptual knowledge measures held constant. For those with outside exposure, solving competence and flexibility measures held constant, while conceptual knowledge actually rose.

Second, consider the finding that control class students, who ostensibly received more procedural practice than those students who solved in the alternative ordering paradigm, became better equation solvers on familiar equations. But this advantage did not appear on transfer equations, and improvement on solving familiar equations did not result in improved flexibility nor improved conceptual knowledge.

(Additional paragraphs need to be written here about the implications of this work and future directions. Suggestions are welcome!)
References


---


