DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD

It pays to compare: An experimental study on computational estimation

Citation	Star, Jon R., and Bethany Rittle-Johnson. 2009. It pays to compare: An experimental study on computational estimation. Journal of Experimental Child Psychology, 102, no. 4: 408-426.
Published Version	doi:10.1016/j.jecp.2008.11.004
Accessed	February 19, 2015 8:36:16 AM EST
Citable Link	http://nrs.harvard.edu/urn-3:HUL.InstRepos:4889485
Terms of Use	This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

(Article begins on next page)

Running Head: IT PAYS TO COMPARE

It Pays to Compare: An Experimental Study on Computational Estimation

Abstract

Comparing and contrasting examples is a core cognitive process that supports learning in children and adults across a variety of topics. In this experimental study, we evaluated the benefits of supporting comparison in a classroom context for children learning about computational estimation. Fifth- and sixth-grade students ($n = 157$) learned about estimation either by comparing alternative solution strategies or by reflecting on the strategies one at a time. At posttest and retention test, students who compared were more flexible problem solvers on a variety of measures. Comparison also supported greater conceptual knowledge, but only for students who already knew some estimation strategies. These findings indicate that comparison is an effective learning and instructional practice in a domain with multiple acceptable answers.

KEYWORDS: learning processes; computational estimation; mathematics education; comparing solution strategies; flexibility; conceptual and procedural knowledge

It Pays to Compare: An Experimental Study on Computational Estimation

There is currently a push to make psychological research more educationally relevant by applying established results from cognitive science toward the improvement of pressing educational problems (National Research Council, 2000). Typically this process begins with the identification of a body of literature from cognitive science that has the potential to inform educational practice; researchers then build upon existing laboratory studies of the phenomena by conducting studies in school settings, using rigorous experimental designs. The present study is an example of this approach. We evaluated whether supporting a core cognitive process – comparison – in a classroom context supported children's learning about computational estimation.

Comparison

A robust literature in cognitive science makes a strong case that comparison – identifying similarities and differences in multiple examples – is a critical and fundamental pathway to flexible, transferable knowledge (Gentner, Loewenstein, & Thompson, 2003; Kurtz, Miao, & Gentner, 2001; Loewenstein & Gentner, 2001; Namy & Gentner, 2002; Oakes & Ribar, 2005; Schwartz & Bransford, 1998). For example, college students who were prompted to compare two business cases by reflecting on their similarities were much more likely to transfer the solution strategy to a new case than were students who read and reflected on the cases independently (Gentner et al., 2003).

Much of the existing research on comparison has not been done with K-12 students, or in classroom settings. Nevertheless, having students compare and contrast alternative solution strategies is one of the core principles in current reform pedagogy in mathematics (Silver, Ghousseini, Gosen, Charalambous, & Strawhun, 2005). Case studies of expert mathematics

teachers emphasize the importance of students actively comparing solution strategies (Ball, 1993; Fraivillig, Murphy, & Fuson, 1999; Hufferd-Ackles, Fuson, & Sherin, 2004; Lampert, 1990; Silver et al., 2005). Furthermore, teachers in high-performing countries such as Japan and Hong Kong often have students produce and discuss multiple solution strategies (Richland, Zur, & Holyoak, 2007; Stigler & Hiebert, 1999). This emphasis on sharing and comparing solution strategies was formalized in the National Council of Teachers of Mathematics (NCTM) *Standards* (NCTM, 1989, 2000, 2006). However, little empirical evidence directly links this teaching practice to student learning.

Recently, Rittle-Johnson and Star (2007) provided initial evidence that the benefits of comparison as demonstrated in laboratory tasks are also applicable to students' learning of algebra in classrooms. Seventy seventh-grade students were randomly assigned to learn about algebra equation solving by either 1) comparing and contrasting alternative solution strategies or 2) reflecting on the same solution strategies one at a time. At posttest, students in the compare group had made greater gains in procedural knowledge and flexibility and comparable gains in conceptual knowledge.

Despite the success of this study, there is a compelling need to replicate the findings from Rittle-Johnson and Star (2007), for several reasons. First, *no* prior studies could be found that assessed the causal influence of comparing contrasting strategies on student learning in mathematics. Additional studies are needed to confirm this finding. Second, there was no retention test to evaluate whether the benefits of comparison persisted over a delay. Third, while Rittle-Johnson and Star found comparison to be effective at improving students' procedural knowledge and flexibility, comparison was not found to differentially impact conceptual knowledge. Given the critical importance of conceptual knowledge to students' learning of

mathematics (Hiebert & Carpenter, 1992), additional studies are needed to demonstrate that comparing multiple strategies improves both procedural and conceptual knowledge. Finally, comparing solution strategies may only facilitate learning in rule-based domains such as algebra equation solving. Many mathematical domains are rule-based, but some areas of mathematics, such as estimation, are less rule-driven and there are multiple correct answers for a given problem. Is comparison effective in less constrained domains such as computational estimation?

Computational Estimation

Estimation is a critically useful skill in everyday life and in mathematics. 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 may 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 "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). Unfortunately, current instructional methods have not been particularly effective at supporting estimation knowledge. It is well documented that a large majority of students have difficulty estimating the answers to problems in their heads (e.g., Case & Sowder, 1990; Hope & Sherrill, 1987; Reys, Bestgen, Rybolt, & Wyatt, 1980; Sowder, 1992).

Estimation is also a domain in which comparing multiple strategies is thought to be beneficial. According to the recent US National Mathematics Advisory Panel (2008) recommendation:

Textbooks need to explicitly explain that the purpose of estimation is to produce an appropriate approximation. Illustrating multiple useful estimation procedures for a single problem, and explaining how each procedure achieves the goal of accurate estimation, is a useful means for achieving this goal. Contrasting these procedures with others that produce less appropriate estimates is also likely to be helpful. (p. 27)

In this study, we focus on computational estimation, which is defined as the process of mentally generating an approximate calculation for a given arithmetic problem (Rubenstein, 1985). Computational estimation is an interesting domain in which to extend the work of Rittle-Johnson and Star (2007) for several reasons. First, as noted above, estimation is less constrained than other mathematical domains such as equation solving. Second, there are a wide variety of estimation strategies that can lead to accurate estimates, and good estimators know and use many estimation strategies (Dowker, 1992, 1997; Dowker et al., 1996).

Third, computational estimation problems do not have a single correct answer; rather, the correctness or 'goodness' of an estimate depends on two sometimes-competing goals. The first, simplicity, refers to how easy it is to compute an estimate (Reys & Bestgen, 1981; LeFevre, Greenham & Waheed, 1993). For example, to compute an estimate for 31 x 46, students may round both numbers to the nearest ten (*round both*; i.e., 30 x 50) or round one number to the nearest ten (*round one*; e.g., 30 x 46). For many elementary school students, it seems plausible that the first strategy is easier. The second goal, proximity, refers to how close the estimate is to the exact answer (Reys & Bestgen, 1981; LeFevre, Greenham & Waheed, 1993). In this

example, *round one* leads to an estimate that is closer to the exact value than *round both*. Note these two goals often compete with each other, in that an easy-to-compute estimate is often not very proximal to the exact value, or conversely, the strategy leading to most proximal answer is not the easiest to compute (Lemaire, Lecacheur, & Farioli, 2000).

These features of estimation make it an ideal domain to extend the work of Rittle-Johnson and Star (2007) because of the many ways in which this domain is different from algebra equation solving (the content used in Rittle-Johnson and Star). In fact, a case can be made that comparison is *less* likely to be effective in computational estimation than it was in algebra equation solving, for at least two reasons. First, when comparing estimation strategies, learners need to look at both the strategy and the estimate in order to evaluate the relative effectiveness of a strategy. In contrast, when comparing equation solving strategies, a learner can essentially ignore the answer and instead focus on similarities and differences between strategies. Second, the efficiency of solution strategies, which is a key criterion on which multiple strategies can be compared, is less obvious in estimation as compared to equation solving. One advantage to using equation solving is that it is relatively easy (and visually apparent) to judge the relative efficiency of two strategies for solving an equation. In contrast, when computing an estimate, efficiency and ease of computation are often individual and subjective judgments.

Overall, comparing solution strategies is much more complex for computational estimation than for algebra, and thus it seems plausible that learners will find it more difficult to learn from comparing multiple strategies in estimation than in algebra equation solving. As a result, computational estimation is an interesting and important domain in which to replicate and extend the results of Rittle-Johnson and Star (2007).

Target Outcomes

Our target outcomes were three critical components of mathematical competence: procedural knowledge, procedural flexibility, and conceptual knowledge (Hiebert, 1986; National Research Council, 2001). Procedural knowledge is the ability to execute action sequences to solve problems, including the ability to adapt known procedures to novel problems (the later ability is sometimes labeled "transfer") (Rittle-Johnson, Siegler, & Alibali, 2001). Procedural flexibility incorporates knowledge of multiple ways to solve problems and when to use them (National Research Council, 2001; Star, 2005, 2007) and is an important component of mathematical competence (Beishuizen et al., 1997; Blöte, Van der Burg, & Klein, 2001; Dowker, 1992; Star & Rittle-Johnson, 2008; Star & Seifert, 2006). To disentangle *knowledge* from *use*, we included an independent measure of flexibility knowledge as well as coded for flexible use of strategies on the procedural knowledge assessment. Finally, conceptual knowledge is "an integrated and functional grasp of mathematical ideas" (National Research Council, 2001, p. 118). This knowledge is flexible and not tied to specific problem types, and is therefore generalizable (although it may not be verbalizable).

Current Study

We compared learning from comparing multiple solutions (*compare condition*) to learning from studying sequentially presented solutions (*sequential condition)* for fifth- and sixth-grade students learning how to compute estimates for multi-digit multiplication problems. Students in both conditions studied worked examples of hypothetical students' estimation strategies and answered questions about the strategies with a partner.

Three features of our study design merit a brief justification. First, we chose to provide students with worked examples because doing so insured exposure to multiple strategies for all

students and facilitated side-by-side comparison of these strategies for students in the *compare* condition. Many 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 on their own (see Atkinson, Derry, Renkl, & Wortham, 2000 for a review). Second, we chose to have students work with a partner because past research indicates that students who collaborate with a partner tend to learn more than those who work alone (e.g., Johnson & Johnson, 1994; Webb, 1991) and teaching students to generate conceptual explanations for their partners improves their own learning (e.g., Cobb & Bauersfeld, 1995; Fuchs et al., 1997). And third, we chose to prompt students to generate explanations when studying worked examples because there is a great deal of evidence that doing so leads to greater learning, as compared to cases when students are not asked to provide explanations (e.g., Bielaczyc, Pirolli, & Brown, 1995; Chi, de Leeuw, Chiu, & LaVancher, 1994).

We hypothesized that students in the compare group would show greater improvements from pretest to posttest, with gains persisting on a retention test, on three outcome measures 1) procedural knowledge (particularly transfer), 2) procedural flexibility, and 3) conceptual knowledge – than students in the sequential group. We expected these differences to emerge as a result of students making more explicit comparisons between strategies and answers, which should highlight the ease and efficiency of multiple estimation strategies and illuminate relationships between estimation strategies, problem types, and attainment of estimation goals (simplicity and proximity).

Method

Participants

Students from two schools participated in the study. School A is a private urban school where 69 fifth-grade students participated (32 female). There were four fifth-grade mathematics classes (all taught by the same teacher) at the school. Students' mean age was 10.6 years (range: 10.0 years to 11.4 years); a majority were Caucasian (13% minority, with 13% African-American). Approximately 10% of students at School A received financial aid. School B is a small rural school where 45 fifth graders and 46 sixth graders participated. At School B, 5th grade students' mean age was 10.7 years (range: 10.0 years to 11.8) while sixth grade students' mean age was 11.8 years (range: 11.0 years to 13.1 years) There were two fifth grade classes (taught by the same teacher) and two sixth grade classes (taught by the same teacher). A majority of participating students were Caucasian. Approximately 36% of students at School B received financial aid. Across the schools, three students were dropped from the study because they were absent from school and missed more than one intervention session. Thus the analysis below includes data from a total of 157 students.

Design

We used a pretest-intervention-posttest design, including a retention test. For the intervention, students were randomly paired with another student in their class, and then pairs of students were randomly assigned to condition, with approximately equal numbers of pairs in each condition within each class. Students in the *compare* condition (n = 82) studied sets of two worked examples for the same problem and answered questions encouraging comparison of the two examples. Students in the *sequential* condition (n = 75) studied the same two worked

examples on two isomorphic problems and answered questions encouraging reflection on a single example.

Materials

    Intervention. The intervention focused on three estimation strategies for multiplying one, two, and three-digit integers (see Table 1). In addition to *round one* and *round both*, the third strategy was to truncate (or *trunc*) each multiplicand, covering up or ignoring the ones digits and multiplying the tens digits, and subsequently adding two zeros to the resulting product (for 13 x 27, 1■ x 2■ yields 2, and then adding two zeros yields an estimate of 200). This strategy is relatively easy and fast and has been advocated for by researchers on computational estimation (Sowder & Wheeler, 1989). The strategies were presented to students by way of worked examples. Worked examples in the compare condition typically illustrated two different estimation strategies for each problem, with the two strategies differing either in terms of proximity to the exact value, ease of computation, or both. On some worked examples, three different estimation strategies were presented for the same problem; on other examples, the same strategy was used to solve two different problems.

    Two packets of worked examples were created for each condition (see Figure 1). In the *compare* packets, there were 15 sets of worked examples (13 pairs and two groups of 3) presented side-by-side on the same page for a total of 32 worked examples. At the bottom of each page were questions prompting students to compare and contrast the worked examples. During the first day of problem solving, the questions that accompanied the worked examples focused on ease of computation, such as "Whose way is easiest? Why?" and "If the number problem were changed [from 13 x 88] to 47 x 88, would that student's way still be easiest? Why or why not?". On the second and third day of problem solving, the questions focused on

proximity to the exact answer, such as "Without knowing the exact value, whose estimate is closer to the exact value of her number problem?" and "Look at the two ways shown above. Do you think one way will always give a closer estimate than the other way on any multiplication problem? Why or why not?".

In the *sequential* packets, there were also 32 worked examples. The same estimation strategies were presented, in some cases with identical problems and in some cases with isomorphic problems (e.g., 27 x 63 and 57 x 43 are isomorphic), but with each worked example presented on a separate sheet. Thus, exposure to multiple strategies of estimation was equivalent across the two conditions. At the bottom of each page was one question prompting students to reflect on that estimation strategy, with an equal number of prompts in the two conditions. The initial questions focused on the ease of a single strategy, such as "If the number problem were changed [from 38 x 63] to 234 x 71, would Casey's way be easy to do? Why or why not?". The later questions focused on closeness to exact value, such as "Without calculating the exact value, how far is your estimate from the exact value?".

Practice problems were integrated into each packet. Each practice problem set asked students to estimate the solution to two problems and then answer one question about their strategy(s) of estimation. In the compare packet, students were asked to estimate the solution to the same problem in two different ways. In the sequential packet, students were asked to estimate the solution to two problems, one of which was identical to the problem in the compare packet and the second of which was isomorphic to the first.

Assessment. The same assessment was used as an individual pretest, posttest, and retention test (see Table 2). It was designed to assess procedural knowledge, flexibility, and conceptual knowledge. *Procedural knowledge* measures assessed knowledge of how to estimate,

using both familiar (six problems, such as 12 x 24 and 113 x 27) and transfer problems (six

problems, such as 1.19 x 2.39 and 102 ÷ 9). In addition, three mental estimation problems

assessed students' ability to compute an estimate quickly and mentally. *Flexibility* knowledge

measures assessed students' ability to recognize, implement, and evaluate multiple strategies for

computing estimates. Flexibility items fell into three categories: (a) Knowledge of multiple

strategies, where two questions asked students to compute estimates in multiple ways; (b)

Recognize and evaluate ease of use, where two questions determined whether students knew

which strategies were computationally easier to implement; and (c) Recognize and evaluate

closeness of estimate, where five questions determined whether students knew which strategies

resulted in an estimate that was most proximal to the exact value. Ten *conceptual knowledge*

items assessed students' knowledge of core concepts related to estimation. Conceptual

knowledge items were modified from past research (Sowder, 1992; Sowder & Wheeler, 1989;

Dowker, 2005) and focused on: (a) definitions of estimation, including an open-ended item

(What does "estimate" mean?) and multiple choice rating of definitions of estimation, (b) the

acceptance of multiple strategies of estimation and multiple values of estimates, and (c) the

impact of estimation strategies on distance from the correct answer (e.g., under- vs.

overestimate).

Procedure

The study occurred during one week of students' regular mathematics classes and

replaced the students' regular instruction on computational estimation. On *Monday*, students

completed a 30-minute written pretest and then were provided with a 10-minute introduction

lesson by a member of the research team. The goals of the introduction lesson were to introduce

students to the idea of estimation as getting an approximate answer and to show students *trunc*, an estimation strategy that they may not be familiar with.

On *Tuesday***,** students were divided into pairs to begin work on the intervention packet. During the partner work, the pairs of students were asked to first explain their answers verbally to one another and then write down a summary of their answer on the packet. We recorded the verbal interactions (using an audiotape recorder and microphones for each pair). During the partner work, the regular classroom teacher and members of the project team circulated and provided 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* to do next or *why* you might use a particular strategy). At the conclusion of each class, students were given a brief homework assignment to practice completing estimation problems.

On *Wednesday*, there was a brief scripted lesson on proximity. The lesson introduced students to the use of the number line and the idea of proximity or closeness to the exact value as a means to evaluate estimates. Then, pairs were given the day's intervention packet. On *Thursday*, students spent the first 30 minutes completing the packet focused on proximity that they had begun on Wednesday. At the end of the class session, a member of the research team provided a scripted 10-minute integrative lesson, providing some points of closure about estimation. The lesson reminded students that estimation is a way to get an approximate answer, that there are many ways to arrive at an estimate, and that different ways of estimating give different estimates. In addition, the lesson pointed out two criteria that may be used in evaluating whether one estimate is better than another (simplicity and proximity). On *Friday*, students

completed the posttest. Two weeks later, children completed the assessment again to assess retention.

Coding

Assessment. The 15 problems on the procedural knowledge assessment were scored for accuracy of the answer; an accurate estimate was defined as one within 30% of the exact value (Rubenstein, 1985). In addition to scoring accuracy, students' solution strategies were coded into categories based on the strategy of estimation used (*trunc*, *round both*, and *round one*). (Some students used a variety of other, idiosyncratic estimation strategies; in rare cases, students calculated the exact value rather than computing an estimate. In all such cases, these strategies were coded as "other".) Inter-rater reliability for coding strategies of estimation (based on 20% of the sample) was 92% (exact agreement).

On the conceptual knowledge assessment, students received one point for correctly answering each of the objective questions. In addition, students explained their reasoning on three items, and these explanations were scored on a 2-point scale. These explanation scores were added to students' conceptual knowledge totals. A conceptual knowledge score was calculated as a percentage of possible points. Inter-rater reliability for the three explanation items (based on 20% of the sample) was 93% (exact agreement).

For the flexibility assessment's three components, the percentage of possible points on each component was calculated, and the three percentages were averaged to yield an overall flexibility score. Inter-rater reliability on 20% of the sample was calculated for the items on the flexibility assessment that were not objective, and exact agreement was 96% for all subjective flexibility items.

Strategy optimization. As an additional measure of flexibility, students' strategies on the six familiar procedural knowledge items were coded for the selection of the optimal estimation strategy, both in terms of ease and for proximity, for each problem.

The optimal strategy in terms of proximity was the strategy of estimation that yielded an estimate that was closest to the exact value. On four of the six problems, *round both* was optimal; on one problem (37 x 17), *round one* was optimal; and on one problem (23 x 52), *round both* and *trunc* were equally optimal in terms of proximity.

The optimal strategy in terms of ease was the strategy of estimation that was the fastest to execute. On four of the six problems, we expected *trunc* to be fastest, on one problem (8 x 76) *round one* should be fastest, and on one problem (12 x 24), *round one* or *trunc* was expected to be fastest. In order to verify which strategies were optimal in terms of ease, we conducted a reaction time study with a sub-sample of 26 of the students who agreed to participate several months after the conclusion of the main study. Students were reminded of the solution strategies and were asked to use a given strategy on a block of 8 problems. Two blocks of problems incorporated a range of two two-digit numbers and students were asked to use *trunc* on one block and *round both* on the other. Two other blocks of problems involved numbers where one multiplicand was near 10 (e.g., 13 x 58) and students were asked to use *round one* on one block and *round both* on the other. Order of presentation of the blocks and which strategy was specified to use first were counterbalanced. Problems were administered using E-Prime and shown to participants on a laptop. As expected, for problems with one multiplicand near 10, *round one* (M = 3.9 s, SD = 1.4) was faster than *round both* (M = 5.0 s, SD = 1.5), $t(25) = 4.517$, $p < .001$. For the other problems, *trunc* (M = 5.1 s, SD = 1.1) was faster than *round both* (M = 6.8 s, SD = 1.7), $t(25) = 5.883$, $p < .001$.

Intervention. Students' answers to intervention questions were coded for the following features: mention of multiple ways to compute an estimate, and comparison of steps, simplicity, proximity, or no comparison. Inter-rater reliability for coding of students' intervention question responses (based on 20% of the sample) was 87% (exact agreement).

Data Analysis

Given that random assignment occurred at the dyad, not the individual, level, and that knowledge was assessed multiple times, we ran a 3-level unconditional means models in HLM (Raudenbush, Bryk, & Congdon, 2003) to evaluate non-independence in dyads (Raudenbush & Bryk, 2002; Singer & Willett, 2003). After controlling for school, no more than 3% of the variance was between-dyad, and chi-square tests confirmed that this variance was not significantly greater than 0 (all *p*'s above .17, with most above .5). There was not sufficient variation at the dyad level to model this level, so we ignored dyad and conducted our analyses at the individual level using repeated-measures ANCOVAs.

Some students were absent on an assessment day. Three students did not complete the pretest, three did not complete the posttest, and two did not complete the retention test, and no student missed more than one assessment. Statisticians strongly recommend the use of imputation, rather than the traditional approach of omitting participants with missing data, because it leads to more precise and unbiased conclusions (Peugh & Enders, 2004; Schafer & Graham, 2002). When the data is missing at random and no more than 5% of the data is missing (as in this study), simulation studies indicate that imputation leads to the same conclusions as when there is no missing data (e.g., Barzi & Woodward, 2004). As recommended by Schafer and Graham (2002), we used the expectation-maximization (EM) algorithm for Maximum Likelihood Estimation via the missing value analysis module of SPSS. The students' missing

scores were estimated from all non-missing values on the variables that were included in the analyses presented below. Findings were the same when we deleted students with missing data from the analyses.

Results

We begin by describing students' results at pretest. We then report the effect of condition on gains in students' knowledge from pretest to posttest and retention test. Finally, we examine the effects of the manipulation during the intervention; in particular, we report on solution strategies and explanation quality during the intervention.

Pretest Knowledge

Many students began the study with some knowledge of estimation strategies and concepts. As shown in Table 3, students on average were able to generate accurate estimates for 3 or 4 of the 12 pretest procedural knowledge items and also had some success on measures of conceptual knowledge and flexibility. *Round both* was the most commonly used strategy on the pretest (see Table 4). Also, at pretest, there were no significant differences between conditions on measures of procedural knowledge, conceptual knowledge, or flexibility, $F(1,155) = 0.360$, 0.728, and 0.006, respectively.

Knowledge Gains from Pretest to Posttest and Retention

Students in the compare condition were expected to have higher procedural knowledge, procedural flexibility and conceptual knowledge at posttest and retention test. Separate repeated-measures ANCOVAs were conducted for each outcome, with time of assessment as a within-subject factor (posttest and retention test) and with condition as the between-subjects factor. Pretest scores on each measure, school, and grade level were included as covariates to control for

prior knowledge differences. Unless otherwise noted, condition did not interact with time, in line with our expectations that the effect of condition would persist at the retention test.

Procedural knowledge. Although means were in the expected direction, there was no significant main effect for condition on the procedural knowledge assessment, $F(1, 150) = 2.514$, $p = .115$, $\eta^2 = .016$, indicating that students in the sequential condition produced accurate estimates as frequently as students in the compare condition (see Table 3). The only significant predictors of accuracy were pretest procedural knowledge, $F(1, 150) = 31.170$, $p < .001$, $\eta^2 = .172$, and school, $F(1, 150) = 7.409$, $p = .007$, $\eta^2 = .047$. Compare and sequential students also did not differ in the number of problems attempted, $F(1, 150) = .794$, $p = .374$, $\eta^2 = .005$ (see Table 4). The only significant predictors of how many problems were attempted were pretest procedural knowledge, $F(1, 150) = 6.694$, $p = .011$, $\eta^2 = .043$, and pretest flexibility knowledge, $F(1, 150) = 4.037$, $p = .046$, $\eta^2 = .026$.

Flexibility. As expected, students in the compare condition became more flexible estimators. Evidence for this result comes from several sources. First, compare group students outperformed sequential students on the flexibility assessment, $F(1, 150) = 14.058$, $p < .001$, $\eta^2 = .086$ (see Table 3). Pretest procedural and flexibility knowledge also predicted flexibility knowledge, $F(1, 150) = 9.895$, $p = .002$, $\eta^2 = .062$ and $F(1, 150) = 41.481$, $p < .001$, $\eta^2 = .217$, respectively. The effect was strongest on the subscale assessing knowledge of multiple strategies, such as problems where students were given a problem and asked to compute an estimate in three different ways (see Table 2); compare students significantly outperformed sequential students on this subscale, $F(1, 150) = 22.155$, $p < .001$, $\eta^2 = .129$. For example, 29% of compare group students were able to produce an estimate in three different ways on both problems in the

knowledge of multiple strategies subscale at posttest, as compared to only 13% of sequential students.

Evidence for compare students' greater flexibility was also found in students' estimation strategies; students in the compare group were more likely to select the easiest strategy for computing estimates on the familiar procedural knowledge items. Recall that we determined if participants selected the optimal strategy for each problem in terms of ease (based on the results of the reaction time study described above) and proximity (based on which strategy led to the closest estimate). Compare group students were significantly more likely to optimize strategy choice for ease, $F(1, 150) = 7.500$, $p = .007$, $\eta^2 = .048$, (Estimated Marginal Mean (EMM) = 1.1 out of 6 problems, SE = 0.2 vs. EMM = 0.6, SE = 0.2), although not for proximity, $F(1, 150) = 1.324$, $p = .252$, $\eta^2 = .009$ (EMM = 2.9 out of 6 problems, SE = 0.2 vs. EMM = 3.1, SE = 0.2).

Compare students' ability to optimize for ease was driven by their greater use of the *trunc* strategy, which was often the fastest strategy to implement. Although *round both* was the most frequently used strategy both before and after the intervention, compare students used *trunc* more often than students in the sequential condition, $F(1, 150) = 8.928$, $p = .003$, $\eta^2 = .056$ (see Table 4). To explore the benefits of using the different strategies, we examined the correlation between frequency of using each strategy on a given assessment and performance on that assessment; frequency of using *round one* and *round both* correlated with procedural, flexibility and conceptual knowledge at posttest and retention test ($r(155)$'s ranging from .185 to .663), whereas frequency of using *trunc* did not.

Conceptual knowledge. There was no main effect for condition on students' conceptual knowledge, $F(1, 150) = .130$, $p = .719$, $\eta^2 = .001$. Only pretest conceptual and procedural

knowledge predicted overall conceptual knowledge, $F(1, 150) = 21.674$, $p < .001$, $\eta^2 = .126$ and $F(1, 150) = 14.069$, $p < .001$, $\eta^2 = .089$, respectively.

However, comparing solution strategies may be most helpful for students who are already familiar with at least one solution strategy (Rittle-Johnson & Star, in press). Based on this hypothesis, we explored whether condition interacted with pretest procedural knowledge, and there was a tendency for such an interaction, $F(1, 149) = 2.908$, $p = .090$, $\eta^2 = .019$, as well as a three way interaction between time, condition and pretest procedural knowledge, $F(1, 149) = 4.214$, $p = .042$, $\eta^2 = .028$. To interpret these interactions, we first categorized students as having low or moderate procedural knowledge at pretest, using a median split (median score at pretest was 20% correct). Then, we conducted separate analyses on posttest and retention test scores. At posttest, there was no main effect of condition or interaction with pretest procedural knowledge category, p's > .3. In contrast, at retention test, there was a condition by pretest category interaction, $F(1, 149) = 5.877$, $p = .017$, $\eta^2 = .038$. As shown in Figure 2, comparison did not impact conceptual knowledge for students who had low knowledge of estimation strategies at pretest. In contrast, compare students with modest knowledge of estimation strategies at pretest (i.e., at least 20% correct) had better maintenance of their conceptual knowledge than sequential students.

Effects of the Condition Manipulation on Intervention Activities

To better understand how condition impacted knowledge gains, we explored the effects of the condition manipulation on intervention activities. Before reporting these effects, it is important to note that the manipulation did *not* impact the amount of material covered during the intervention; on average, students in the compare and sequential conditions studied approximately 25 of the 32 available worked examples ($M = 25.4$ ($SD = 6.41$) vs. $M = 24.7$ (SD

= 6.15), respectively) and solved 10 of the available 15 practice problems ($M = 10.84$ ($SD = 3.63$) vs. $M = 9.81$ ($SD = 3.50$), respectively). We expected the compare condition to support more explicit comparisons between multiple estimation strategies.

Procedural knowledge during intervention activities. On the practice problems, students in the compare condition were more likely to compute accurate estimates than sequential students, $F(1, 150) = 4.828$, $p = .030$, $\eta^2 = .031$. Compare students generated accurate estimates on 93% of practice problems, while sequential students' estimates were accurate on only 88% of problems. Compare students were also more likely to use *trunc* (23% vs. 11% of problems) and *round one* (11% vs. 3% of problems), and less likely to use *round both* (32% vs. 48% of problems), than sequential students, $F(1, 150) = 19.064$, $p < .001$, $\eta^2 = .113$, $F(1, 150) = 27.178$, $p < .001$, $\eta^2 = .153$, and $F(1, 150) = 41.776$, $p < .001$, $\eta^2 = .218$, respectively.

Explanation quality. Student pairs provided written explanations to reflection questions when studying worked examples. Our explanation coding schemes were designed to indicate whether our condition manipulation had its intended effects.

The first coding scheme focused on whether students' explanations referenced multiple estimation strategies, as might be expected in the compare condition. 97% of compare group students' explanations referenced multiple strategies, while only 10% of sequential students' explanations did so. A representative explanation from a student pair in the compare group that illustrates this focus on multiple strategies is, "Annette didn't round up. Claire did, which makes it bigger."

Our second coding scheme investigated the characteristics of strategies that students compared. Of particular interest was the extent to which comparisons of multiple estimation strategies focused on the proximity of estimates to the exact answer, the ease of computing

estimates, comparison of specific solution steps, and/or other characteristics of estimation strategies (see Table 5). Students in the compare group were more likely to make explicit comparisons than those in the sequential group. Specifically, students in the compare group were significantly more likely to compare two estimates based on their respective proximity to the exact answer, the ease with which an estimate could be computed, and the specific steps involved in computing an estimate.

Overall, the intended effect of the intervention was manifest in students' explanations. Compare students, who repeatedly viewed side-by-side worked examples illustrating multiple estimation strategies, were more likely to reference multiple strategies in their explanations. Furthermore, these explanations frequently included comparisons of salient features of the estimates and estimation strategies, including proximity, simplicity, and the particular steps used in an estimation strategy.

We also explored whether individual differences in the frequency of making explicit comparisons during the intervention predicted outcomes at posttest and retention. In this model, frequency of generating comparisons during the intervention, rather than condition, was used as a predictor. Making more comparisons during the intervention was predictive of gains in flexibility $F(1, 150) = 15.554$, $p < .001$, $\eta^2 = .094$. However, it did not reliably predict procedural or conceptual knowledge gain (p's $> .2$). Frequency of comparing solution steps in particular was somewhat predictive of procedural knowledge, $F(1, 150) = 3.634$, $p = .059$, $\eta^2 = .024$.

Partner interaction. The discussions from a pair of high-learning and a pair of low-learning students in the compare condition were transcribed to better understand how comparison could support learning. Their discussions on two identical worked examples are presented in Table 6. It is evident that the high-learning pair consistently outperformed the low-learning pair

in noticing key differences of estimating (via rounding up or down), and how and when each strategy should be used. The high-learning pair easily synthesized or reconciled their knowledge from the past with the current example, compared solution steps, and analyzed accuracy, efficiency, and constraints of each strategy. In contrast, the low-learning pair had little perception of when and how to use different estimation strategies (rounding up versus down). The low-learning pair had difficulty synthesizing knowledge gained from multiple strategies, rarely compared solution steps, and did not consider efficiency and constraints of strategies.

Discussion

The goal of the present study was to evaluate whether comparing solution strategies is more effective than sequential study of strategies for learning about computational estimation. Despite a large literature in cognitive science demonstrating the benefits of comparison and frequent calls for teachers to compare and contrast multiple strategies during mathematics instruction, we could find only one study, Rittle-Johnson and Star (2007), that provided experimental evidence in mathematics classrooms for the benefits of comparison. Are the benefits of comparing solution strategies found for seventh-graders learning about equation solving generalizable to other domains, especially one in which there are multiple correct answers to a single problem?

One finding that appears to generalize across domains is that comparing solution strategies led to greater flexibility. The present study, taken together with Rittle-Johnson and Star (2007), gives compelling evidence that providing students with worked examples placed side by side on the same page with accompanying prompts for self-explanation leads to greater flexibility, as compared to presentation of the same examples one per page. Across the two studies, comparing solution strategies led to greater knowledge of multiple strategies and the

ability to adaptively select the most appropriate strategies for given problems or goals (in this case, optimization for ease of computation).

The fact that comparing solution strategies led to greater flexibility in the present study is particularly noteworthy, given our focus on the mathematical domain of computational estimation. Estimation is different from algebra equation solving in several ways that have potentially serious implications for the possible benefits of comparison in promoting flexibility. In particular, unlike linear equations, estimation problems do not have a single correct answer; rather, the goodness of an estimate depends on two often-competing goals: how easy the estimate is to compute, and how close the estimate is to the exact value of the problem. In addition, the relationship between a strategy and either of these goals is quite complex. Whether or not a strategy such as *round both* provides the most proximal answer depends on the problem; whether or not a strategy such as *round one* provides an easy estimate depends on whether a multiplicand is near 10 and on the computational resources of the person generating the estimates. We began the study with reasonable skepticism that comparing solution strategies would be effective in this new and more complex domain: It was not clear that a side-by-side comparison of two estimation strategies that yield different estimates would be productive. Thus, the present results, which indicated that comparison did help students become more flexible in their knowledge of estimation, are particularly noteworthy.

However, our findings in the domain of estimation diverged from prior work on comparing solution strategies in equation solving in two ways. First, there were different effects for comparison on the conceptual knowledge assessments. We found that comparing solution strategies helped students retain their conceptual knowledge of estimation, at least if they began the study with modest procedural knowledge. No benefit of comparison for conceptual

knowledge were detected in Rittle-Johnson and Star (2007), but a retention test was not included and interactions with prior knowledge were not evaluated. The present study suggests that students may need familiarity and fluency with a limited range of strategies before comparison of additional strategies aids knowledge of related concepts and that comparison may be most important for remembering the concepts after a delay.

Other research supports the idea that familiarity in a domain improves the effects of comparison. In particular, children often have difficulty learning from the comparison of two examples if they do not have prior experience within the domain, but providing children with relevant experience allows them to benefit from the comparisons (Gentner, Loewenstein & Hung, 2007; Kotovsky & Gentner, 1996). This does not mean that people need to be well versed in one example before comparing it to a different example (Gentner, 2005); modest amounts of prior knowledge or exposure seem to be sufficient. The current findings highlight the potential importance of familiarity for mathematics learning.

A second key difference between the findings of Rittle-Johnson and Star (2007) in equation solving and the present study concerns procedural knowledge. In Rittle-Johnson and Star, comparing solution strategies helped students solve more equations correctly, yet in the present study, the intervention did not lead students to estimate more multiplication problems reasonably. These divergent findings may have arisen from a key difference in the effectiveness of different strategies between the domains of estimation and equation solving. In particular, learning new and more sophisticated strategies for solving equations (e.g., improved flexibility) was related to gains in procedural knowledge; frequency of adopting new strategies predicted procedural knowledge scores at posttest and partially mediated the effect of condition on procedural knowledge in Rittle-Johnson and Star. In other words, in the domain of equation

solving, comparison was quite instrumental in introducing students to new strategies that made more complex problems easier to solve.

In contrast, in the present study, learning new strategies for computing estimates was not necessarily related to gains in procedural knowledge. Rather, mastering a single strategy – namely *round both* – was sufficient for solving both familiar and transfer problems. Comparison did encourage students to adopt the easy *trunc* strategy, but use of this strategy was not related to performance on any measure. Mathematics education researchers have advocated teaching *trunc* as a quick and easy way to estimate and, more generally, that instruction should focus on multiple strategies for estimation (Sowder, 1992; Sowder & Wheeler, 1989; Reys, et al., 1980; Reys & Bestgen, 1981). Similarly, the US National Mathematics Advisory Panel (2008) recommended that "Teachers should broaden instruction in computational estimation beyond rounding. They should insure that students understand that the purpose of estimation is to approximate the exact value and that rounding is only one estimation strategy." (p. 27). We agree that learning and comparing multiple estimation strategies is important. However, the present study also suggests that careful consideration must be given to *when* new strategies should be introduced. For example, our results indicate that *trunc* may not help students on problems where they are already able to execute *round both* accurately. However, on harder problems on which *round both* is difficult to execute correctly, or for younger students who are not able to implement *round both*, introducing alternative and easier strategies such as *trunc* seems warranted.

Implications for Research on Computational Estimation

Although the primary focus of the present study was on comparison, our results also contribute to the literature on computational estimation in at least two ways.

First, the results of the present study inform research on how students balance multiple goals in generating estimates. As noted above, one interesting aspect of computational estimation as a problem solving domain is that it requires consideration and balancing of two, sometimes competing, goals - simplicity and proximity (Lemaire, Lecacheur, & Farioli, 2000). Compare group students' adoption of the *trunc* strategy, which is very easy to implement but does not guarantee an accurate estimate, suggests that students in the present study tended to prioritize simplicity over proximity when computing estimates. This emphasis on simplicity may arise for at least two reasons. First, proximity may be more challenging to determine for a given problem and strategy, perhaps due to processing limitations, memory capacity, or knowledge of multiplication facts (Case & Sowder, 1990; Dowker, 2005). However, it may also be the case that students grasp the principle of simplicity before the principle of proximity. This latter interpretation would be consistent with the results of LeFevre and colleagues (LeFevre, Greenham, & Waheed, 1993), who found that knowledge of simplicity preceded proximity in students' strategy choices for estimates in grades 4 and 8, but that adults were able to use and balance both simplicity and proximity in their estimates (see also Levine, 1982).

Second and related, this transition to balancing simplicity and proximity when computing estimates seems to indicate greater adaptive expertise in adults than in children (Hatano, Miyake, & Binks, 1977; Baroody & Dowker, 2003). Children may develop routine expertise with estimation strategies, which might entail the adoption of and use of a new strategy such as truncation on a set of problems. But children with routine expertise in estimating would likely fail to adaptively use the *trunc* strategy, such as restricting its use to problems or contexts where this strategy is particularly appropriate. Over time, older children and adults may develop

adaptive expertise, where they flexibly coordinate competing goals for estimates with the characteristics of problems and problem-solving contexts.

Implications for Instruction

The current findings provide much needed evidence in support of reform efforts in mathematics education that advocate for comparison of solution strategies. Our unique use of random assignment of students to condition within their regular classroom context, along with maintenance of a fairly typical classroom environment, provided causal evidence for the benefit of comparing solution strategies while maintaining fairly good external validity. US teachers commonly use comparison in their lessons, but frequently *not* in ways that seem most conducive to the development of mathematical understanding (Richland et al., 2004; Richland et al., 2007). Experimental research on comparison, including our own, provides several suggestions for using comparison effectively in mathematics classrooms.

First, teachers must choose problems and solution strategies carefully. The problems should highlight important and meaningful concepts for students to learn and to be solvable using multiple strategies. In addition, students may need some familiarity with one of the strategies before comparing two different strategies.

Second, comparison requires careful support to be effective. Our materials were carefully designed to support effective comparison. Past research suggests that five features of our intervention may have been particularly important. As noted in Rittle-Johnson and Star (2007, in press), these features are 1) a written record of all to-be-compared solution strategies, with the solution steps aligned (Fraivillig et al., 1999; Richland et al., 2004; Richland et al., 2007), 2) explicit opportunities to identify similarities and differences in strategies (Fraivillig et al., 1999; Gentner et al., 2003; Silver et al., 2005), 3) instructional prompts to encourage students to

consider the efficiency of the strategies (Fraivillig et al., 1999; Lampert, 1990), 4) using common labels, such as labeling strategies, to invite comparison and help alignment (Namy & Gentner, 2002) and 5) providing some direct instruction to supplement learners' comparisons (e.g., Schwartz & Bransford, 1998). In the current study, scaffolds for effective comparison were embedded in the instructional material and seemed to support productive explanation during partner work in the classroom. We caution that poorly planned or implemented comparison is unlikely to facilitate learning.

Future Directions

This study is an important initial step in applying established results from cognitive science about the benefits of comparison toward improvements in pressing educational problems. However, there are several areas where future work should considering focusing. First, it is critical that follow-up studies on comparison include longer instructional interventions. The present study, as well as prior work by Rittle-Johnson and Star (2007, in press) involved very short one-week-long interventions. In order to convince teachers and schools to implement pedagogical approaches using comparison, research examining the feasibility and effectiveness of longer interventions is critical.

Second, future work should continue to investigate how and when comparison facilitates learning. Are some forms of comparison more effective than others? Rittle-Johnson and Star (in press) recently explored whether some forms of comparison (e.g., comparing solution strategies, as was done in the present study) are more conducive to learning than others (e.g., comparing two different problems, both solved with the same strategy, or comparing two equivalent problems, both solved with the same strategy). Results suggest that conceptual knowledge and procedural flexibility were best supported by comparison of solution strategies, but these results

merit replication in mathematical domains other than algebra equation solving (the focus of Rittle-Johnson and Star (in press)).

Finally, given the challenges associated with teachers' effective implementation of comparison (Richland et al., 2004; Richland et al., 2007), an important direction for future research is to explore other ways that classroom instruction can incorporate comparison of multiple strategies. In the present study, as well as Rittle-Johnson and Star (2007), students were able to realize the benefits of comparison using only written instructional materials and without teacher-led whole class discussions of similarities and differences between multiple strategies. A natural extension of this work would be to examine students' mathematics textbooks: To what extent do texts provide students with opportunities to compare multiple strategies? Greater incorporation of side-by-side comparisons of multiple strategies is a simple, and potentially very effective, way to improve mathematics textbooks.

Conclusion

This study contributes to a growing body of research demonstrating that comparing multiple strategies to the same problem facilitates learning. The focus here is on estimation, which is both a critically important real-world skill and a mathematical domain that is significantly more complex than equation solving, which has been the target of prior work. Comparison helped students develop a larger repertoire of estimation strategies, improved students' ability to select the most appropriate strategies for computing an easy estimate, and increased retention of conceptual knowledge for some students. When learning how to estimate, the present results provide experimental evidence that it pays to compare.

References

Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. (2000). Learning from examples: Instructional principles from the worked examples research. *Review of Educational Research, 70*(2), 181-214.

Ball, D. L. (1993). With an eye on the mathematical horizon: Dilemmas of teaching elementary school mathematics. *Elementary School Journal, 93*(4), 373-397.

Baroody, A. J., & Dowker, A. (Eds.). (2003). *The development of arithmetic concepts and skills: Constructing adaptive expertise*. Mahwah, NJ: Lawrence Erlbaum.

Barzi, F., & Woodward, M. (2004). Imputations of missing values in practice: Results from imputations of serum cholesterol in 28 cohort studies. *American Journal of Epidemiology, 160*, 34-45.

Beishuizen, M., van Putten, C. M., & van Mulken, F. (1997). Mental arithmetic and strategy use with indirect number problems up to one hundred. *Learning and Instruction, 7*(1), 87-106.

Bielaczyc, K., Pirolli, P. L., & Brown, A. L. (1995). Training in self-explanation and self-regulation strategies: Investigating the effects of knowledge acquisition activities on problem solving. *Cognition and Instruction, 13*(2), 221-252.

Blöte, A. W., Van der Burg, E., & Klein, A. S. (2001). Students' flexibility in solving two-digit addition and subtraction problems: Instruction effects. *Journal of Educational Psychology, 93*, 627-638.

Case, R., & Sowder, J. T. (1990). The development of computational estimation: A neo-Piagetian analysis. *Cognition and Instruction, 7*, 79-104.

Chi, M. T. H., de Leeuw, N., Chiu, M. H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18*(3), 439-477.

Cobb, P., & Bauersfeld, H. (Eds.). (1995). *The emergence of mathematical meaning: Interaction in classroom cultures*. Hillsdale, NJ: Lawrence Erlbaum.

Dowker, A. (1992). Computational estimation strategies of professional mathematicians. *Journal for Research in Mathematics Education, 23*(1), 45-55.

Dowker, A. (1997). Young children's addition estimates. *Mathematical Cognition, 3*, 141-154.

Dowker, A., Flood, A., Griffiths, H., Harriss, L., & Hook, L. (1996). Estimation strategies of four groups. *Mathematical Cognition, 2*(2), 113-135.

Dowker, A. (2005). *Individual differences in arithmetic: Implications for psychology, neuroscience and education*. New York: Psychology Press.

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

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

Gentner, D. (2005). The development of relational category knowledge. In D. H. Rakison & L. Gershkoff-Stowe (Eds.), *Building object categories in developmental time* (pp. 245-275). Mahwah, NH: Lawrence Erlbaum.

Gentner, D., Loewenstein, J., & Hung, B. (2007). Comparison facilitates children's learning of names for parts. *Journal of Cognition and Development, 8*(3), 285-307.

Gentner, D., Loewenstein, J., & Thompson. (2003). Learning and transfer: A general role for

   analogical encoding. *Journal of Educational Psychology, 95*(2), 393-405.

Hatano, G., Miyake, Y., & Binks, M. (1977). Performance of expert abacus operators. *Cognition,*

   5, 57-71.

Hiebert, J., & Carpenter, T. (1992). Learning and teaching with understanding. In D. Grouws

   (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 65-97). New

   York: Simon & Schuster Macmillan.

Hope, J. A., & Sherrill, J. M. (1987). Characteristics of unskilled and skilled mental calculators.

   Journal for Research in Mathematics Education, 18(2), 98-111.

Hufferd-Ackles, K., Fuson, K., & Sherin, M. G. (2004). Describing levels and components of a

   math-talk learning community. *Journal for Research in Mathematics Education, 35*(2),

   81-116.

Johnson, D. W., & Johnson, R. T. (1994). *Learning together and alone: Cooperative,*

   competitive and individualistic learning (4th ed.). Boston, MA: Allyn and Bacon.

Kotovsky, L., & Gentner, D. (1996). Comparison and categorization in the development of

   relational similarity. *Child Development, 67*(6), 2797-2822.

Kurtz, K., Miao, C.-H., & Gentner, D. (2001). Learning by analogical bootstrapping. *The*

   Journal of the Learning Sciences, 10, 417-446.

Lampert, M. (1990). When the problem is not the question and the solution is not the answer:

   Mathematical knowing and teaching. *American Educational Research Journal, 27*(1), 29-

   63.

LeFevre, J. A., Greenham, S. L., & Waheed, N. (1993). The development of procedural and

   conceptual knowledge in computational estimation. *Cognition and Instruction, 11*, 95-

   132.

Lemaire, P., Lecacheur, M., & Farioli, F. (2000). Children's strategy use in computational

   estimation. *Canadian Journal of Experimental Psychology, 54*(2), 141-148.

Levine, D. R. (1982). Strategy use and estimation ability of college students. *Journal for

   Research in Mathematics Education, 13*(5), 350-359.

Loewenstein, J., & Gentner, D. (2001). Spatial mapping in preschoolers: Close comparisons

   facilitate far mappings. *Journal of Cognition and Development, 2*(2), 189-219.

Namy, L. L., & Gentner, D. (2002). Making a silk purse out of two sow's ears: Young children's

   use of comparison in category learning. *Journal of Experimental Psychology: General,

   131*(1), 5-15.

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for

   school mathematics*. Reston, VA: Author.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school

   mathematics*. Reston, VA: Author.

National Council of Teachers of Mathematics. (2006). *Curriculum Focal Points for

   prekindergarten through grade 8 mathematics: A quest for coherence*. Reston, VA:

   Author.

National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the

   National Mathematics Advisory Panel*, U.S. Department of Education: Washington, DC.

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

   Washington, DC: National Academy Press.

National Research Council. (2001). *Adding it up: Helping children learn mathematics*. Washington, DC: National Academy Press.

Oakes, L. M., & Ribar, R. J. (2005). A comparison of infants' categorization in paired and successive presentation familiarization tasks. *Infancy, 7*(1), 85-98.

Peugh, J. L., & Enders, C. K. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement. *Review of Educational Research, 74*(4), 525-556.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (Second ed.). Thousand Oaks, CA: Sage Publications, Inc.

Raudenbush, S. W., Bryk, A. S. & Congdon, R. (2004). *HLM-6*. Scientific Software International.

Reys, R. W., Bestgen, B., Rybolt, J. F., & Wyatt, J. W. (1980). *Identification and characterization of computational estimation processes used by in-school pupils and out-of-school adults* (No. ED 197963). Washington, DC: National Institute of Education.

Reys, R. W., & Bestgen, B. J. (1981). Teaching and assessing computational estimation skills. *The Elementary School Journal*, *82*(2), 116-127.

Richland, L. E., Zur, O., & Holyoak, K. J. (2007). Cognitive supports for analogies in the mathematics classroom. *Science, 316*, 1128-1129.

Richland, L. E., Holyoak, K. J., & Stigler, J. W. (2004). Analogy use In eighth-grade mathematics classrooms. *Cognition and Instruction, 22*, 37-60.

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

Rittle-Johnson, B., & Star, J. R. (2007). Does comparing solution methods facilitate conceptual

   and procedural knowledge? An experimental study on learning to solve equations.

   Journal of Educational Psychology, 99(3), 561-574.

Rittle-Johnson, B., & Star, J. R. (in press). The effects of different comparisons on conceptual

   knowledge and procedural flexibility for equation solving. *Journal of Educational

   Psychology.*

Rubenstein, R. N. (1985). Computational estimation and related mathematical skills. *Journal for

   Research in Mathematics Education, 16*, 106-119.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art.

   Psychological Methods, 7(2), 147-177.

Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. *Cognition and Instruction, 16*(4),

   475-522.

Silver, E. A., Ghousseini, H., Gosen, D., Charalambous, C., & Strawhun, B. (2005). Moving

   from rhetoric to praxis: Issues faced by teachers in having students consider multiple

   solutions for problems in the mathematics classroom. *Journal of Mathematical Behavior,

   24*, 287-301.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and

   event occurrence*. New York: Oxford University Press.

Sowder, J. T. (1992). Estimation and related topics. In D. Grouws (Ed.), *Handbook of research

   on mathematics teaching and learning* (pp. 371-389). New York: Macmillan.

Sowder, J. T., & Wheeler, M. M. (1989). The development of concepts and strategies used in

   computational estimation. *Journal for Research in Mathematics Education, 20*, 130-146.

Star, J. R. (2005). Reconceptualizing procedural knowledge. *Journal for Research in Mathematics Education, 36*(5), 404-411.

Star, J. R. (2007). Foregrounding procedural knowledge. *Journal for Research in Mathematics Education, 38*(2), 132-135.

Star, J. R., & Rittle-Johnson, B. (2008). The development of flexible knowledge: The case of equation solving. *Learning and Instruction, 18*, 565-579.

Star, J. R., & Seifert, C. (2006). The development of flexibility in equation solving. *Contemporary Educational Psychology, 31*, 280-300.

Stigler, J. W., & Hiebert, J. (1999). *The teaching gap: Best ideas from the world's teachers for improving education in the classroom*. New York: Free Press.

Webb, N. M. (1991). Task-related verbal interaction and mathematics learning in small groups. *Journal for Research in Mathematics Education, 22*(5), 366-389.

Table 1

Strategies for Estimating 13 x 27

Strategy name	How to compute	Estimate	Distance from exact value (351)
Round both	10 x 30	300	51 (14.5%)
Round one	10 x 27	270	81 (23.0%)
	13 x 30	390	39 (11.1%)
Trunc	1█ x 2█ and append two zeros	200	151 (43.0%)

Table 2

Sample Items for Assessing Procedural Knowledge, Flexibility, and Conceptual Knowledge

Problem Type	Sample Items	Scoring and Scale Alpha at post-test
I. Procedural Knowledge		$\alpha = .89$
a. Mental (n = 3)	12 x 34; 23 x 49	1 pt for each estimate within 30% of exact value
b. Familiar (n = 6)	12 x 24; 113 x 27	1 pt for each estimate within 30% of exact value
c. Transfer (n = 6)	1.19 x 2.39; 102 ÷ 9	1 pt for each estimate within 30% of exact value
II. Flexibility		$\alpha = .80$
a. Recognize and evaluate ease of use (n = 2)	Luther computes an estimate for 27 x 39 by rounding both numbers and multiplying 30 x 40. Riley arrives at an estimate by rounding one number, multiplying 27 x 40. Which way to estimate is easier?	1 pt for each correct choice
b. Knowledge of multiple strategies (n = 2)	Estimate 12 x 36 in 3 different ways	1 pt for two correct, unique solutions; 2 pts for three correct, unique solutions

| c. Recognize and evaluate closeness of estimate (n = 5) | Carmine computes an estimate for 9 x 48 by rounding both numbers and multiplying 10 x 50. Radika arrives at an estimate by rounding one number, multiplying 10 x 48. Without calculating the exact value, which estimate is closer to the exact value? | 1 pt for each correct choice |

| III. Conceptual Knowledge | | $\alpha = .61$ |
| (n = 10) | 1) (Define) What does "estimate" mean? 2) (Recognize definition) Rate each definition of estimation as not so good, kind of good, or very good: "Estimation is making math problems easy and quick," [*kind of good*] "Estimation is guessing," [*not so good*] "Estimation is using easier numbers and getting close to the true value." [*very* | 1) 2 pts if mention both simplicity and proximity (e.g., "It is an easy way to get close to the exact answer"); 1 pt for simplicity or proximity 2) 1 pt for each correct choice |

good]

3) (Accept multiple estimates) Kim and Ahmad were asked to estimate 81 x 15. Kim estimated 80 x 10 = 800; Ahmad estimated 80 x 20 = 1600. Is Kim's answer (800) an OK estimate? Is Ahmad's answer (1600) an OK estimate?

3) 1 pt each if identify that an OK estimate

4) (Over vs. underestimate) To estimate 21 * 39, Steven multiplies the tens digits, 2█ * 3█ and adds two zeros. Will Steven's estimate be bigger or smaller than the exact value?

4) 1 pt for choosing "smaller"

Table 3

Student Performance By Condition

	Pretest		Posttest		Retention	
	M	*SD*	*M*	*SD*	*M*	*SD*
Compare						
Procedural	31	30	71	21	65	28
Flexibility	35	20	59	24	60	23
Conceptual	36	15	48	16	47	18
Sequential						
Procedural	28	31	64	27	60	30
Flexibility	34	24	47	27	52	24
Conceptual	34	15	47	17	45	15

Table 4

Estimation Strategy By Condition (Proportion of Trials)

Strategy	Pretest		Posttest		Retention	
	Compare	Sequential	Compare	Sequential	Compare	Sequential
Round both	28	23	51	55	52	55
Round one	1	2	2	3	3	3
Trunc	3	2	14	5	11	6
Other	15	21	14	14	16	18
Blank	53	52	19	23	18	19

Table 5

Percent of Intervention Explanations Containing Comparisons, by Condition

Explanation Characteristic	Sample Explanations	Compare	Sequential
Compare proximity	"Keith's estimate is closer because Keith was 25 off. And Jeffrey's 100 off"	31%*	3%
Compare simplicity	"Ronny's way is easier because he rounded two digits rather than 1 digit."	13%*	4%
Compare steps	"Gretchen didn't round her numbers and Vanessa rounded."	22%*	2%
Compare other	"Both answers have the same number of 0s."	4%	1%
No compare	"Marquan's way was an estimate."	37%*	91%

Note: Students could receive more than one code for a single explanation

*Difference between conditions is significant at $p < .01$.

Table 6

Sample Dialogue of a High-Learning and a Low-Learning Pair During the Intervention

High-Learners: Alice and Paul	Low-Learners: Jane and Don
Paul: [reading question] Why is Annette's estimate smaller than the exact value while Claire's estimate is bigger than the exact value?	Jane: [reading question] Why is Annette's estimate smaller than the exact value while Claire's estimate is bigger than the exact value?
Alice: Well, I think that Annette's answer is smaller because she rounded both of them down. Claire's way is bigger because she rounded both of them up.	Don: Because she rounded the ones digit…
	Jane: She didn't round
	Don: Yes she did… see she rounded 48 up to 50.
Paul: Okay want to read what we said?	Jane: Because Claire rounded up.
Alice: Okay well I said that Annette's way is smaller because she rounded down, 29 down to 20 and 48 down to 40, and that Claire's way is bigger and more exact because she rounded up. We talked about it yesterday, rounding not to the closest one.	Jane: [reading question] Do you think one way will always give an estimate closer than another way on any multiplication problem?
	Don: Yeah because if you round down, it's not …
Paul: I used to think it was no big deal, but you really notice.	Jane: If you round down, then the solution…

Alice: [reading question] Do you think that one

way will always give a closer estimate than the

other way on any multiplication problem?

Paul: Absolutely! Because Claire's way, she

estimated the closest, up for both of them, and

she got the closest answer, but so I think that is

a much better answer. So yes I think it does

make a difference.

[each student writes his or her response…]

Alice: Okay so let me read mine. I think

Claire's way will give a better answer because

she rounded to the closest number and Annette

rounded to the furthest number, she rounded

them both down and I don't think she should

have done that.

Figure Captions

Figure 1. Sample packet page for A) the compare condition and B) the sequential condition.

Figure 2. Conceptual knowledge by condition. (Estimated marginal mean across posttest and

retention test. Error bars are standard errors.)

Figure 1

A. Compare condition

About how much is 27 * 43?

Allie's way:	Claire's way:
27 * 43	27 * 43
My estimate is 800.	My estimate is 1200.
I covered up the ones digits and then multiplied the tens digit like this: 2■ * 4■ = 8 Then I added two zeros because I covered up two digits and got 800.	I rounded both numbers. I rounded 27 up to **30**. I rounded 43 down to **40**. Then I multiplied **30 * 40** and got 1200.

3. How is **Allie**'s way similar to **Claire**'s way?

4a. Use **Allie**'s way to estimate 21 * 43.

4b. Would **Claire**'s way give a different estimate for 21 * 43 than **Allie**'s way?

B. Sequential condition

About how much is 57 * 23?

Claire's way:

57 * 23

My estimate is 1200.

I rounded both numbers.

I rounded 57 up to **60**.
I rounded 23 down to **20**.

Then I multiplied **60 * 20** and got 1200.

2a. Is this an OK estimate?   _____Yes   _____No

2b. How do you decide if an estimate is OK?

-------------------------------------- *NEXT PAGE* --------------------------------------

About how much is 27 * 43?

Allie's way:

27 * 43

My estimate is 800.

I covered up the ones digits and then multiplied the tens digits like this:

2█ * 4█ = 8

Then I added two zeros because I covered up two digits and got 800.

4a. Use **Allie's** way to estimate the answer to 21 * 43.

4b. Did Allie get the same estimate for these two number problems?