Conceptual Knowledge, Procedural Knowledge, and Metacognition in Routine and Nonroutine Problem Solving

David W. Braithwaite

Lauren Sprague

Florida State University

Correspondence concerning this article should be addressed to David W. Braithwaite, Department of Psychology, Florida State University, 1107 W. Call Street, Tallahassee FL 32306, braithwaite@psy.fsu.edu.

This research was supported by National Science Foundation under Grant No. 1844140 to the first author. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

The authors are grateful to K. Anders Ericsson for his helpful comments on the design of the studies reported here and on the manuscript for this article.

Abstract

When, how, and why students use conceptual knowledge during math problem solving is not well understood. We propose that when solving routine problems, students are more likely to recruit conceptual knowledge if their procedural knowledge is weak than if it is strong, and that in this context metacognitive processes, specifically feelings of doubt, mediate interactions between procedural and conceptual knowledge. To test these hypotheses, in two studies (*N*s = 64 and 138), university students solved fraction and decimal arithmetic problems while thinking aloud; verbal protocols and written work were coded for overt uses of conceptual knowledge and displays of doubt. Consistent with the hypotheses, use of conceptual knowledge during calculation was not significantly positively associated with accuracy, but was positively associated with displays of doubt, which were negatively associated with accuracy. In Study 1, participants also explained solutions to rational arithmetic problems; using conceptual knowledge in this context was positively correlated with calculation accuracy, but only among participants who did not use conceptual knowledge during calculation, suggesting that the correlation did not reflect "online" effects of using conceptual knowledge. In Study 2, participants also completed a nonroutine problem solving task; displays of doubt on this task were positively associated with accuracy, suggesting that metacognitive processes play different roles when solving routine and nonroutine problems. We discuss implications of the results regarding interactions between procedural knowledge, conceptual knowledge, and metacognitive processes in math problem solving.

1. Introduction

Research on math cognition, learning, and development has traditionally distinguished between procedural and conceptual knowledge (Hiebert & Lefevre, 1986). Procedural knowledge refers to knowledge of procedures for solving problems, such as the step-by-step algorithms that children are taught in school. Conceptual knowledge is a multifaceted construct that includes knowledge of categories, relationships, principles, and representations.

Problems for which students have been taught and have practiced appropriate solution procedures, which we term "routine problems," can in principle be solved using procedural knowledge alone. Yet, performance on routine problem solving tasks is correlated with individual differences in conceptual knowledge (Bailey, Hansen, & Jordan, 2017; Fuchs et al., 2010; Hecht & Vagi, 2010; M. Schneider, Rittle-Johnson, & Star, 2011) and is improved by interventions that focus on conceptual knowledge (Fuchs et al., 2013; Fyfe, DeCaro, & Rittle-Johnson, 2014; Rittle-Johnson, Siegler, & Alibali, 2001; Siegler & Ramani, 2009). These findings suggest that even when solving routine problems, at least some students use conceptual knowledge at least some of the time[1].

This conclusion in turn suggests several questions. Which students use conceptual knowledge when solving routine problems? Under what circumstances do they do so? What mechanisms govern interactions between procedural and conceptual knowledge in this context? We propose two related hypotheses: that when solving routine problems, students are more likely to use conceptual knowledge (if they have it) when their procedural knowledge is weak than

[1] Students may also rely on procedural knowledge to help understand concepts and perform conceptual tasks (Rittle-Johnson, 2017). Our focus on effects of conceptual knowledge on performance of procedural tasks in the present studies does not imply an absence of effects in the other direction.

when it is strong; and that in this context metacognitive processes, specifically feelings of doubt, mediate interactions between procedural and conceptual knowledge.

In the next two sections, we provide rationales for these hypotheses based on prior research regarding the roles of conceptual knowledge and metacognition in math problem solving. We also detail several empirical predictions implied by our hypotheses. Then, we discuss the domain in which we tested the predictions: fraction and decimal arithmetic. Last, we describe the present studies in more detail.

1.1. Using Conceptual Knowledge During Problem Solving

One way in which students might use conceptual knowledge during problem solving is to help determine solution strategies (Rittle-Johnson, 2017). Specifically, when presented problems that cannot be solved through straightforward application of known procedures, students can use conceptual knowledge to adapt previously-learned procedures or invent new ones (Baroody, 2003; Canobi, 2009; Hiebert & Lefevre, 1986; Perry, 1991; Shrager & Siegler, 1998). For example, in Rittle-Johnson and Alibali (1999), fourth and fifth graders who received a conceptual intervention regarding one type of math equivalence problem (a+b+c = a+__) subsequently invented strategies to solve different types of equivalence problems (e.g., a+b+c = d+__), whereas students who received a procedural intervention usually failed to do so.

The above function of conceptual knowledge has most often been emphasized in the context of nonroutine problems, but conceptual knowledge could play a similar role for routine problems when a student cannot remember an appropriate procedure or is unsure which of multiple candidate procedures is correct. To illustrate, a student asked to calculate 3/5+1/5 might wonder whether to pass the common denominator of the operands into the answer, yielding 4/5, or to add the denominators of the operands, yielding 4/10. A student who understands that 3/5 and 1/5 represent 3 parts and 1 part of a whole divided into 5 parts might use this understanding to reason that the former procedure is the correct one.

A second way in which conceptual knowledge might be used during problem solving is to detect errors (Ohlsson & Rees, 1991). Using inappropriate procedures, or incorrectly executing appropriate procedures, often leads to answers that violate conceptual constraints. For example, $3/5+1/5 = 4/10$ violates the constraint that a sum of positive numbers is larger than the addends, because $4/10 < 3/5$. If a student commits an error that violates a conceptual constraint, conceptual knowledge might enable the student to detect the error, which might in turn lead the student to correct their mistake and obtain a correct answer (Ohlsson, 1996; Siegler, Thompson, & Schneider, 2011). For example, consistent with the possibility of this mechanism, Wong and Odic (2021) recently demonstrated that adults can use their sense of numerical magnitudes—a form of conceptual knowledge—to make rapid, intuitive judgments about the direction of arithmetic errors.

In both of the mechanisms just described, conceptual knowledge becomes involved when procedural knowledge fails, such as when a student does not know or is uncertain of an appropriate procedure, or commits an error due to using an incorrect procedure or executing a procedure incorrectly. In contrast, a student who can confidently retrieve and correctly execute an appropriate procedure for a given problem may do so “on autopilot,” without using conceptual knowledge. Thus, we hypothesized that in the context of routine problems, use of conceptual knowledge is more likely when procedural knowledge is weak than when it is strong.

This hypothesis implies that relations between conceptual knowledge and performance on procedural tasks may be more complex than previously documented. If a student’s procedural knowledge relevant to a given problem is weak, the student may use conceptual knowledge to compensate. However, in doing so, the student is unlikely to perform better than they would have if their procedural knowledge had been strong enough that they did not need to use conceptual knowledge. Thus, despite conceptual and procedural knowledge being positively correlated when assessed separately, we predicted that using conceptual knowledge when solving routine problems would not be positively associated with accuracy (Table 1, Prediction 1).

Table 1. Predictions Tested in the Present Study.

Number	Prediction
1	Use of conceptual knowledge will not be positively associated with accuracy when solving routine math problems
2	Doubt will be positively associated with using conceptual knowledge when solving routine math problems
3	Doubt will be negatively associated with accuracy when solving routine math problems

1.2. Metacognitive Processes During Problem Solving

Like conceptual knowledge, metacognitive processes—that is, processes that regulate and monitor cognitive processes—are critical for solving math problems (Alibali, Brown, & Menendez, 2019; Crowley, Shrager, & Siegler, 1997; Garofalo & Lester, 1985). Successful problem solvers self-monitor more than less successful problem solvers do (Schoenfeld, 1992; Stillman & Galbraith, 1998), and individuals with stronger metacognitive knowledge perform better in math (Carr, Alexander, & Folds-Bennett, 1994; Nelson & Fyfe, 2019; W. Schneider & Artelt, 2010). Further, metacognitive interventions can improve problem solving performance (e.g., Desoete et al., 2003; Hacker et al., 2019).

We propose that during routine problem solving, metacognitive processes—specifically, feelings of uncertainty and feelings of error, which we refer to jointly as “doubt”—mediate interactions between procedural knowledge and conceptual knowledge. This may occur in at least two ways, corresponding to the two mechanisms described in the previous section. First, when selecting a strategy to solve a problem, a student may feel doubt if no procedure is sufficiently activated in memory, or if multiple procedures are highly activated. This feeling of doubt could prompt use of conceptual knowledge to help generate or select an appropriate procedure. Second, if a student commits a procedural error and thereby generates an answer that violates a conceptual constraint, knowledge of the constraint may cause the student to doubt their answer (Fernandez Cruz, Arango-Muñoz, & Volz, 2016). This feeling of doubt could prompt the student to rerun or modify their procedure.

Our proposal implies that doubt and conceptual knowledge go hand in hand. Both occur in similar situations; doubt during strategy selection may trigger use of conceptual knowledge; and in the presence of an error, conceptual knowledge may generate doubt. These observations suggested a prediction: doubt should be positively associated with using conceptual knowledge when solving routine math problems (Table 1, Prediction 2).

A final implication of our proposal is that in the context of routine problems, doubt is a good indicator of weak procedural knowledge. Thus, doubt should be negatively associated with accuracy in this context (Table 1, Prediction 3). Though this prediction may seem uncontroversial, it is still worth testing, because confidence is not always well calibrated with accuracy in math (Nelson & Fyfe, 2019; Reder & Ritter, 1992). Further, the prediction does not apply for nonroutine problems, for which doubt might have a positive effect by prompting deliberate consideration of alternative strategies. We elaborate on this point in Study 2.

1.3. Arithmetic with Fractions and Decimals

The present studies tested the predictions in Table 1 in the domain of rational number arithmetic. Knowledge of rational numbers is a critical foundation for more advanced math such as algebra (Siegler et al., 2012). Yet, many students experience large and persistent difficulties in this area. Rational number arithmetic is particularly challenging, including fraction arithmetic (Hansen, Jordan, & Rodrigues, 2015; Hecht & Vagi, 2012; Mack, 1995; Siegler et al., 2011) and decimal arithmetic (Hiebert & Wearne, 1985; Hurst & Cordes, 2018b). For example, US sixth graders correctly solved only 46% of fraction arithmetic problems in Siegler and Pyke (2013), and only 57% of decimal arithmetic problems in Tian et al. (2021).

Difficulties with fraction and decimal arithmetic result in part from the large number of procedures that must be learned, the complexity of these procedures, and the ease of confusing the procedures with each other (Braithwaite, Pyke, & Siegler, 2017; Lortie-Forgues, Tian, & Siegler, 2015). Difficulties in this area also reflect lack of conceptual knowledge, without which students have little basis for choosing correct procedures over incorrect ones or for distinguishing between correct and incorrect answers (Siegler, Im, Schiller, Tian, & Braithwaite, 2020). At least four types of conceptual knowledge are relevant in this domain. Because these types provided a framework for our analyses, we describe them in detail here.

First, *magnitude knowledge* refers to understanding that fractions and decimals have magnitudes that can be compared, ordered, and placed on a number line. Students could use magnitude knowledge to detect arithmetic errors that result in implausibly large or small answers (Siegler et al., 2011), as in the earlier example of rejecting 3/5+1/5 = 4/10 because 4/10 < 3/5. Consistent with this possibility, interventions that focus on fraction magnitude understanding have led to improvements in fraction arithmetic skill (Fuchs et al., 2013, 2021).

Second, *fraction interpretation* refers to semantic interpretations of fractions. Five such interpretations were proposed by Kieren (1980; see also Behr, Lesh, Post, & Silver, 1983): measurement, part-whole, quotient, ratio, and operator. Subsequent research explored learning activities conducive to understanding these interpretations, such as splitting wholes into parts and recombining these parts (Braithwaite & Siegler, 2021; Martin et al., 2015; Steffe, 2004; Tzur, 1999; Tzur & Hunt, 2015), and trajectories from less to more advanced interpretations (e.g., from part-whole to measurement, Wilkins & Norton, 2018). Meaningful interpretations of fractions could help students make sense of fraction arithmetic procedures. For example, 3/5+1/5 = 4/5 makes sense when understood as splitting a whole into five equal parts, combining three of those parts, then adding another of the parts, yielding four of the parts. Consistent with this possibility, understanding of meaningful fraction interpretations, especially the part-whole and measurement interpretations, predicts fraction arithmetic skill (Gabriel et al., 2013; Hecht & Vagi, 2010).

Third, *place value knowledge* refers to understanding how the value of each digit in a decimal depends on its position relative to the decimal point and how the position of the decimal point affects the decimal's magnitude. Place value concepts could help individuals understand decimal arithmetic procedures. For example, 4+.3 = 4.3, not 7 or .7, because digits representing different place values should not be added. Consistent with place value knowledge contributing to decimal arithmetic skill, an intervention that emphasized place value led to improvements in decimal arithmetic skill (Wearne & Hiebert, 1988), and interleaving lessons on decimal place value and decimal addition procedures led to greater improvement in decimal addition skill than teaching place value and addition in separate blocks (Rittle-Johnson & Koedinger, 2009).

Finally, *cross-notation knowledge* refers to knowledge of relations between fractions and decimals, such as knowing how to translate fractions into equivalent decimals and vice versa. Cross-notation knowledge could enable students to use knowledge of one notation to help solve problems in a different notation, such as solving 3+1/5 by reasoning that 1/5 = .2 so 3+1/5 = 3+.2 = 3.2. Spontaneous cross-notation translation has previously been reported in the context of number line estimation tasks (Siegler & Thompson, 2014; Siegler et al., 2011). Cross-notation knowledge was emphasized in Moss and Cases's (1999) successful rational number intervention, and cross-notation knowledge predicts fraction and decimal affrithmetic accuracy when controlling for fraction and decimal magnitude knowledge (Braithwaite, McMullen, et al., in press).

1.4. The Present Studies

In contrast to previous studies that have often assessed conceptual and procedural knowledge using separate tasks, the present studies investigated uses of conceptual knowledge in the context of a procedural task—fraction and decimal arithmetic calculation. Two studies were conducted to test the predictions in Table 1. Our approach to testing these predictions was identical in both studies and is described below. Distinctive aspects of each study are described in the Introductions thereof.

Both studies included a *calculation task* in which participants were asked to think aloud while solving fraction and decimal arithmetic problems. Displays of doubt were identified by analysis of think-aloud protocols and written work, as in some investigations of math problem solving (e.g., Schoenfeld, 1992; Stillman & Galbraith, 1998). We adopted this approach rather than eliciting retrospective confidence ratings (e.g., Fitzsimmons et al., 2020; Nelson & Fyfe, 2019) because think-aloud protocols could reveal feelings of doubt that occurred throughout each trial, whereas retrospective confidence ratings might not detect doubts that were resolved during trials. For example, a participant who committed an error, then noticed and fixed it, might report high confidence in their solution, thus concealing the feeling of error experienced earlier.

We also relied on think-aloud protocols to assess use of conceptual knowledge on the calculation task. According to a meta-analysis by Fox, Ericsson, and Best (2011), thinking aloud does not affect accuracy on cognitive tasks, whereas other verbal report procedures, such as explaining, may have positive effects on accuracy. Further, explanations might omit details due to forgetting or a desire to simplify. For example, a participant who used conceptual reasoning to choose between two strategies, if asked to explain their solution, might report only the chosen strategy, whereas think-aloud protocols would also reveal the reasoning that led to the choice.

Participants in both studies were university students. Rational numbers are not a focus of math education in the US after sixth grade (CCSSI, 2010). However, many adults continue to struggle with rational numbers (Fazio, DeWolf, & Siegler, 2016; Opfer & Devries, 2008; Sidney, Thalluri, Buerke, & Thompson, 2019; Stigler, Givvin, & Thompson, 2010), including with arithmetic (Hurst & Cordes, 2016, 2018a; Newton, 2008; Siegler & Lortie-Forgues, 2015). These phenomena are concerning because knowledge of fractions and decimals is expected in many university courses. Further, 68% of adults in the US use rational numbers in their jobs (Handel, 2016). These considerations informed our decision to focus on adults. We report a similar study of children elsewhere (Braithwaite, Sprague, & Siegler, in press).

2. Study 1

In Study 1, after completing the calculation task, participants performed an *explanation task* in which they generated explanations regarding fraction and decimal arithmetic procedures. We expected that many participants who did not use conceptual knowledge when calculating would display such knowledge when explaining. Such a result would permit us to exclude the possibility that failure to use conceptual knowledge during calculation merely reflected participants lacking relevant conceptual knowledge, being unable to verbalize such knowledge, or being unable to use such knowledge to reason about rational number arithmetic.

We further predicted that use of conceptual knowledge on the explanation task, unlike use of conceptual knowledge on the calculation task, would be positively associated with accuracy on the calculation task. Confirmation of this prediction would indicate that relations between conceptual and procedural knowledge may differ depending on whether conceptual knowledge is assessed via tasks designed to elicit such knowledge or in the context of tasks that could be completed using only procedures.

2.1. Method

2.1.1. Participants

Participants were 64 undergraduate students (38 women, 26 men; 32 first year, 13 second year, 18 third year or higher, 1 year not reported) from a mix of majors, the most common being psychology (n = 26). Participants were recruited from the Psychology Department participant pool at Florida State University (FSU) and received course credit for participation (participants were offered alternative options for course credit). In the years Studies 1 and 2 were conducted, students admitted to FSU had average SAT scores ranging from 1312 to 1325 and average ACT scores ranging from 29 to 30. During these years, the participant pool was 79% White or Caucasian, 10% Black or African-American, 4-6% Asian, and 5-7% other; across races, 22% identified as Hispanic or Latino. Interviews were conducted by the two authors, a female graduate student, and a female undergraduate researcher.

2.1.2. Tasks and Materials

Stimuli for the calculation task were 12 arithmetic problems including three fraction addition problems (3/5+1/5, 3/5+1/4, 3+1/5), three fraction multiplication problems (3/5×1/5, 3/5×1/4, and 3×1/5), three decimal addition problems (12.3+5.6, 2.46+4.1, 5.61+23), and three decimal multiplication problems (2.4×1.2, 2.3×0.13, 3.2×31).

Stimuli for the explanation task were 12 other arithmetic problems, three for each of the four problem types in the calculation task: fraction addition (2/3+1/4, 1/2+3/7, 2/5+1/6), fraction multiplication (3/4×1/4, 5/6×1/6, 2/3×1/3), decimal addition (5.73+1.2, 6.15+2.1, 4.32+3.4), and decimal multiplication (7.1×2.1, 5.1×3.1, 3.1×4.1).

2.1.3. Procedure

Participants first completed a training on thinking aloud, which followed a script adapted from Fox et al. (2011). The script emphasized that participants were not to explain their thoughts but only to report them. Participants practiced thinking aloud with several problems that did not involve fractions or decimals (e.g., “How many months start with the letter ‘J’?”).

Participants then completed the calculation task. They were told to think aloud while solving each problem and to do the problems as they would if they were not thinking aloud. Fraction and decimal problems were presented in separate blocks, with the sequence of these blocks counterbalanced. Within each block, problems were presented in one of four semi-random orders, with addition and multiplication problems intermixed.

Finally, participants completed the explanation task. Trials involving the four problem types (fraction addition, fraction multiplication, decimal addition, decimal multiplication) were presented in separate blocks, with order of notations (fractions or decimals first) and operations (addition or multiplication first) counterbalanced between participants. Within each of these blocks, participants completed one Explain-Solution trial, one Explain-Error trial, and one Explain-Algorithm trial, in that order. In Explain-Solution trials, participants were told to solve a problem, then explain how they solved it and why their solution made sense. In Explain-Error trials, participants were shown two solutions to a problem and were told that one was correct and the other incorrect; they were then asked to say which solution was incorrect and why. In

Explain-Algorithm trials, participants were shown a standard algorithm for solving the given type of problem and a worked example illustrating the algorithm; they were then asked to explain why a key step of the algorithm made sense.

All tasks were completed in paper and pencil format. Sessions were audio recorded. All materials shown to participants and the scripts followed by experimenters for this study and Study 2 are provided in the Supplementary Materials.

2.1.4. Coding

Each trial of the calculation task was coded for whether doubt was displayed. Doubt was coded if the participant expressed uncertainty about how to do a problem, explicitly considered multiple strategies, stated that their solution was likely to be incorrect, or crossed out, erased, or modified their written work. These criteria were intended to include feelings of both uncertainty and error, which were not coded separately due to the difficulty of distinguishing them.

Also, each trial of the calculation and explanation tasks was coded for whether conceptual knowledge was overtly used. Use of conceptual knowledge was coded if the participant displayed any of the types of conceptual knowledge detailed in the Introduction—magnitude knowledge, fraction interpretation, place value knowledge, and cross-notation knowledge—or displayed other types of knowledge that coders considered to be conceptual (this occurred on 1% of calculation trials and 1% of explanation trials). Uses of conceptual knowledge were not classified as correct or incorrect because they often were not clearly one or the other. Explanations were not scored for quality, but only as displaying conceptual knowledge or not.

The guidelines used for coding doubt and use of conceptual knowledge, and examples of trials that displayed each of the four types of conceptual knowledge listed above, are provided in the Supplementary Materials. Each trial was coded separately by two coders, who coded the trials in several batches and met to discuss each batch before coding the next batch. On the calculation task, the coders agreed whether doubt was displayed on 92% of trials, whether any conceptual knowledge was used on 94% of trials, and which types of conceptual knowledge were used (if any) on 94% of trials. On the explanation task, the coders agreed whether any conceptual knowledge was used on 88% of trials and which types of conceptual knowledge were used (if any) on 78% of trials. Disagreements were resolved through discussion.

2.2. Results

Table 2 displays descriptives and correlations for our main outcome measures.

Table 2. Descriptives and Correlations for Measures in Study 1.

		Correlations		
	Mean (SD)	1B	1C	2
Calculation task				
1A. Accuracy	.80 (.17)	-.35 **	-.07	.36 **
1B. Doubt	.21 (.12)		.26 *	-.03
1C. Conceptual knowledge use	.12 (.16)			.48 **
Explanation task				
2. Conceptual knowledge use	.60 (.22)			

Note. “Accuracy” denotes proportion of trials correctly answered, “doubt” denotes proportion of trials on which doubt was displayed, and “conceptual knowledge use” denotes proportion of trials on which any conceptual knowledge was used; * indicates $p < .05$ and ** indicates $p < .01$.

Below, we report analyses testing the predictions in Table 1. These predictions were tested using correlation analysis and mixed logistic regression. The correlation analyses tested for between-subjects effects—for example, did participants who displayed doubt more often also use conceptual knowledge more often? The regressions tested for within-subjects effects—for example, were participants more likely to use conceptual knowledge on calculation trials in which they displayed doubt than on trials in which they did not, after accounting for between-participant variation in use of conceptual knowledge? After reporting tests of the predictions in Table 1, we report analyses regarding the explanation task and the types of conceptual knowledge that were used in each task.

Here and in Study 2, analyses were conducted in *R*, using *lme4* (Bates, Maechler, Bolker, & Walker, 2013) and *lmerTest* (Kuznetsova, Brockhoff, & Christensen, 2016) for regressions. In the regressions, participant was a random effect; notation, arithmetic operation, and the notation * operation interaction were included as fixed effects; and notation and operation were centered (i.e., for notation, "fraction" and "decimal" were coded as -0.5 and 0.5, and for operation, "addition" and "multiplication" were coded as -0.5 and 0.5). All significant effects ($p < .05$) are reported.

2.2.1. Prediction 1: Use of conceptual knowledge will not be positively associated with accuracy when solving routine math problems

No evidence was found for a positive correlation between use of conceptual knowledge and accuracy on the calculation task, $r = -.07$, $t(62) = -0.6$, $p = .56$ (Table 2). To seek evidence *against* such a positive correlation, we examined the one-tailed left 95% CI of the correlation coefficient (Lakens, Scheel, & Isager, 2018), which was [-1, .14]. This result does not exclude the correlation being positive, but indicates that it is likely either not positive or small.

The prediction was further tested using mixed logistic regression, with accuracy on each trial of the calculation task as the dependent variable and use of conceptual knowledge on the same trial as a predictor. Like the correlation analysis, the mixed logistic regression did not yield evidence for a positive effect of conceptual knowledge use on accuracy, $B = -0.5$, $z = -1.4$, $p = .15$. Correct answers were advanced on 73% of trials in which conceptual knowledge was used and on 81% of trials in which conceptual knowledge was not used.

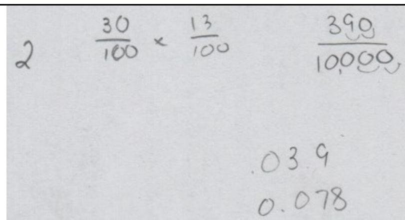
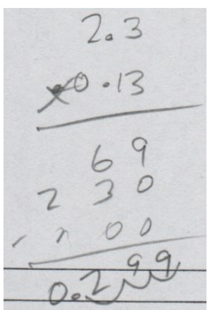
The regression analysis also found an effect of operation, $B = -2.0$, $z = -8.1$, $p < .001$, and a notation * operation interaction, $B = -1.0$, $z = -2.1$, $p < .001$. Participants were more accurate on addition than multiplication problems (92% vs. 68%). This effect was larger for decimal problems (93% vs. 60%) than for fraction problems (92% vs. 76%).

2.2.2. Prediction 2: Doubt will be positively associated with using conceptual knowledge when solving routine math problems

As predicted, the frequency with which participants displayed doubt on the calculation task was positively correlated with how often they displayed conceptual knowledge, $r = .26$, $t(62) = 2.1$, $p = .038$ (Table 2). Similarly, mixed logistic regression found that the likelihood of using conceptual knowledge was higher on trials in which doubt was displayed than when doubt was not displayed (21% vs. 9%), $B = 1.0$, $z = 3.1$, $p = .002$.

Table 3 shows two examples of trials that illustrate relations between doubt and use of conceptual knowledge. Both trials involved the problem 2.3×0.13. The participant in trial A felt doubt about how to solve the problem and used conceptual knowledge to generate a strategy, leading to an incorrect answer, which also elicited doubt. The participant in trial B felt doubt about their answer and used conceptual knowledge to evaluate it, leading to a correct answer.

Table 3. Two Participants' Solutions for 2.3×0.13 in the Calculation Task in Study 1.

Trial	Written Work	Think-Aloud Protocol	Code
A		Wow, there's definitely a way to do this that I don't remember, so	Doubt
		So we get 2 here, let me do, let's try 30 out of 100 times 1, huh, times 13 out of 100	Conceptual (cross-notation)
		13 times 30, that's 10 times 3 which is 300, plus 30 three more times, so 390	
		Over 100 times 100, add 0s, 10,000 ...	
		0, 0 point 3 9 ...	
		I'm gonna do this, times that by 2	
		This doesn't make any sense, but we're gonna do 0 point 0 7 8 ...	Doubt
B		3 times 3 is 9, 3 times 2 is 6	
		Uh, 0, 1 times 3 is 3, 1 times 2 is 2	
		Zeros, and, uh, 0 0 0, 9, 9, uh, 2	
		And then move the decimal, two times? Oh wait, no, three times	Doubt
		So 0 point 2 9 9	
		Yeah, cause it's times, it's smaller.	Conceptual (magnitude)

Note. Protocols have been edited for brevity; omissions are marked with ellipsis (...). "Code" indicates codes that were assigned based on corresponding lines of the protocols.

The mixed logistic regression also found an effect of notation, $B = 1.0$, $z = 3.9$, $p < .001$, indicating that participants displayed conceptual knowledge more often on decimal problems than fraction problems (16% vs. 7%).

2.2.3. Prediction 3: Doubt will be negatively associated with accuracy when solving routine math problems

Consistent with the prediction, how often participants displayed doubt on the calculation task negatively predicted accuracy on that task, $r = -.35$, $t(62) = -3.0$, $p = .004$ (Table 2).

Similarly, mixed logistic regression found that participants answered correctly less often when they displayed doubt than when they did not (57% vs. 86%), $B = -1.3$, $z = -5.2$, $p < .001$. The effect of arithmetic operation reported under Prediction 1 also appeared here, whereas the notation * operation interaction reported there did not reach significance in this analysis.

2.2.4. Relations between the explanation and calculation tasks

Participants displayed conceptual knowledge much more often on the explanation task than on the calculation task (60% vs. 12% of trials, Table 1). Of the 64 participants, 63 displayed conceptual knowledge at least once on the explanation task, whereas only 37 did so at least once on the calculation task. Thus, over half of participants had conceptual knowledge that they revealed during explanation but did not overtly use during calculation.

As predicted, calculation accuracy was positively correlated with the proportion of explanation trials on which participants used conceptual knowledge, $r = .36$, $t(62) = 3.0$, $p = .004$ (Table 2). A possible interpretation of this correlation is that individuals with strong conceptual knowledge used that knowledge to improve their performance on the calculation task, and subsequently revealed that knowledge in their explanations. We reasoned that in this case, the correlation should not appear among participants who never used conceptual knowledge on the calculation task. However, within this group ($n = 27$), calculation accuracy was even more strongly correlated with use of conceptual knowledge use on the explanation task, $r = .47$, $t(25) = 2.7$, $p = .014$. In contrast, among participants who overtly used conceptual knowledge at least once during calculation ($n = 37$), the correlation between calculation accuracy and use of conceptual knowledge on the explanation task was not significant, $r = .24$, $t(35) = 1.5$, $p = .15$.

Participants varied in mean number of words spoken per explanation trial ($M = 110.8$, $SD = 35.2$). This number correlated with proportion of explanation trials coded as displaying conceptual knowledge, $r = .54$, $t(62) = 5.0$, $p < .001$. Thus, apparent effects of conceptual knowledge could be effects of verbosity. However, when controlling for number of words spoken per explanation trial, the partial correlation between calculation accuracy and conceptual knowledge use on the explanation task remained significant, $r = .34$, $t(62) = 2.9$, $p = .006$.

2.2.5. Types of conceptual knowledge used in each task

Table 4 shows the proportions of trials in each task in which participants displayed each of the four types of conceptual knowledge on which our coding scheme was based. The most used types of conceptual knowledge on the calculation task were cross-notation knowledge and decimal place value knowledge. On the explanation task, magnitude knowledge and decimal place value knowledge were displayed most often.

Table 4. Mean (SD) Proportions of Trials in Which Participants Displayed Each Type of Conceptual Knowledge on Each Task in Study 1.

	Type of Conceptual Knowledge			
Task	Magnitude knowledge	Fraction interpretation	Decimal place value	Cross-notation knowledge
Calculation task	.01 (.04)	.00 (.01)	.05 (.09)	.05 (.11)
Explanation task	.32 (.16)	.09 (.11)	.23 (.09)	.05 (.10)
Low bin	.22 (.12)	.04 (.07)	.21 (.09)	.01 (.03)
High bin	.43 (.13)	.15 (.11)	.26 (.09)	.09 (.13)

Note. “Low bin” and “High bin” denote participants whose use of conceptual knowledge on the explanation task was ≤ or > the median (58%), respectively.

To better understand relations between calculation accuracy and performance on the explanation task, we correlated calculation accuracy separately with use of each type of conceptual knowledge on the explanation task. Calculation accuracy was correlated with use of magnitude knowledge ($r = .26$, $t(62) = 2.1$, $p = .04$) and fraction interpretation knowledge ($r = .35$, $t(62) = 2.9$, $p = .005$), but not with use of decimal place value knowledge ($p = .19$) or cross-notation knowledge ($p = .98$), on the explanation task. Related, individual differences in participants' use of conceptual knowledge on the explanation task were largest with respect to magnitude knowledge and fraction interpretations. This effect is shown in the bottom two rows of Table 4: When participants were binned using a median split on proportion of explanation trials in which they used any type of conceptual knowledge, the bins differed most with respect to use of magnitude knowledge and fraction interpretations.

2.3. Discussion

The results lend support to our central hypotheses that when solving routine problems, students are more likely to rely on conceptual knowledge when their relevant procedural knowledge is weak than when it is strong, and that doubt mediates interactions between conceptual and procedural knowledge in such situations. In Study 2, we attempted to replicate these findings with a larger sample and preregistered analyses.

Consistent with previous research on relations between procedural and conceptual knowledge, calculation accuracy was positively correlated with use of conceptual knowledge on the explanation task. This correlation appeared only among participants who did not overtly use conceptual knowledge during calculation, suggesting that it was not entirely driven by use of conceptual knowledge during calculation. We elaborate on this point in the General Discussion.

Calculation accuracy was positively correlated with use of magnitude knowledge and fraction interpretation knowledge on the explanation task, but these two types of conceptual knowledge were rarely spontaneously used on the calculation task (Table 4). In contrast, calculation accuracy was uncorrelated with use of decimal place value knowledge or cross-notation knowledge on the explanation task, but spontaneous uses of these two types of conceptual knowledge on the calculation task were relatively common (Table 4). These results suggest that the types of conceptual knowledge that students use most often when solving routine problems may differ from the types of conceptual knowledge that are most helpful for understanding calculation procedures. Future research should explore this possibility.

3. Study 2

Study 2 was conducted to replicate the main findings from the calculation task of Study 1 with a larger sample and preregistered design. As in Study 1, participants completed the calculation task while thinking aloud, and trials were coded for displays of doubt and use of conceptual knowledge. It was predicted that use of conceptual knowledge would not be positively associated with accuracy, doubt would be positively associated with use of conceptual knowledge, and accuracy would be negatively associated with doubt (Table 1).

Another goal of Study 2 was to investigate boundary conditions for the negative associations between doubt and accuracy found in Study 1. To do so, we distinguished between routine problems, like those in the calculation task, and nonroutine problems, meaning problems for which one has not been taught appropriate solution procedures. We reasoned that when solving nonroutine problems, a student who reflexively uses the first strategy that comes to mind, and unreflectively accepts the output of that strategy as their final answer, might perform poorly. In contrast, a student who deliberately considers alternative strategies, and evaluates their solutions rather than accepting them without question, might perform well (Garofalo & Lester, 1985; Schoenfeld, 1992). Thus, for nonroutine problems, doubt about how to solve a problem or about whether an initial solution is correct could be adaptive, rather than serving as a signal of weak procedural knowledge as in the case of routine problems.

To test this hypothesis, in Study 2, participants completed an extended version of the Cognitive Reflection Task (CRT-Long; Frederick, 2005; Primi, Morsanyi, Chiesi, Donati, & Hamilton, 2016). This task consists of math story problems[2] for which an apparently obvious solution is incorrect. For example, "A bat and a ball cost $1.10 in total. The bat costs a dollar more than the ball. How much does the ball cost?" has an apparently obvious solution of $0.10, but the correct answer is $0.05. Such problems are nonroutine because most university students have not been taught procedures for solving them. Doubting the correctness of the apparently obvious solutions seems especially critical for success on such problems. We therefore predicted that on the CRT-Long, displays of doubt would be positively associated with accuracy.

Study 2 was preregistered at https://osf.io/e9d63.

3.1. Method

3.1.1. Participants

Participants were 138 undergraduate students (99 women, 39 men; 56 first year, 33 second year, 49 third year or higher) whose most common major was psychology (*n* = 37). Participants were recruited from the FSU Psychology participant pool, as in Study 1. They received course credit or cash for participation. Interviews were conducted by five female undergraduate students and two lab managers, one male and one female.

Power analysis conducted in G*Power 3.1 (Faul, Erdfelder, Lang, & Buchner, 2007) prior to data collection indicated that 126 was the minimum sample needed for 85% power to detect correlations of doubt with conceptual knowledge use and accuracy equal to those found in Study 1 (*r* = .26 and -.33). We recruited 140 participants to ensure 126 if up to 10% of them were

[2] The CRT-long was devised as a measure of cognitive reflection rather than math ability. However, the problems in the CRT-long all involve calculation with numbers, and performance on the CRT-long correlates with other measures of numeracy (Primi et al., 2016). Thus, we refer to these problems as math problems.

excluded. In fact, only two were excluded, leaving 138. However, 33 of these were excluded from analyses involving the CRT-Long because they reported having seen three or more of the problems in it before (n = 32) or did not complete that task (n = 1). Other data exclusions are detailed in the Supplementary Materials. All exclusions followed our preregistered criteria.

3.1.2. Tasks and Materials

Stimuli for the calculation task were the 12 problems used for that task in Study 1. Stimuli for the cognitive reflection task were the six problems in the CRT-Long (Primi et al., 2016), which includes the three problems of the original CRT and three other problems with similar structure to the original ones (the problems are provided in the Supplementary Materials).

3.1.3. Procedure

Due to the COVID-19 pandemic, sessions were conducted via video chat using Zoom, and materials were presented using Qualtrics. Participants first were trained to think aloud using a modified version of the script from Study 1. The main modifications were that participants were instructed to say aloud anything they wrote, crossed out, or erased, and that participants completed three additional practice trials that required writing. The instruction to say aloud anything written, crossed out, or erased was repeated at the beginning of each task.

Next, participants completed the calculation task, followed by the CRT-Long. In the calculation task, fraction and decimal problems were presented in blocks, with the sequence of these blocks counterbalanced. Within each block, problems were presented in random order. In the CRT-Long, problems were presented in a single block, within which their order was randomized for each participant.

In both the calculation task and the CRT-Long, one problem appeared at a time. Participants were told to read each problem aloud, then solve it while thinking aloud. After solving the problem, participants were asked to type their answer into a text box on screen and present their written work to the webcam so the interviewer could take a screenshot. Then, participants were asked "How confident are you that your solution is correct?" Responses were on a 5-point scale with endpoints labeled "Not confident at all" (1) and "Extremely confident" (5). These confidence ratings were included so that we could compare our observation-based measure of doubt to a metacognitive measure based on self-report.

After completing the CRT-Long, participants were asked which of the CRT-Long problems they had seen before. Problems that participants reported having seen before (7% of trials for the 105 participants included in analyses of this task) were excluded from analyses.

3.1.4. Coding

Calculation trials were coded for whether doubt and conceptual knowledge were displayed according to the same criteria as in Study 1. Minor revisions were made to the coding guidelines to address questions that arose while coding Study 1. The revised guidelines are provided in the Supplementary Materials.

CRT-Long trials were also coded for whether doubt was displayed, using the same criteria as on the calculation task—that is, doubt was coded for verbal expression of uncertainty; consideration of multiple strategies; saying that a solution was likely wrong; or crossing out, erasing, or modifying written work. CRT-Long trials were not coded for use of conceptual knowledge due to the lack of a clear distinction between conceptual and procedural knowledge on this task. For example, selecting appropriate procedures for routine calculation problems might be possible using procedural knowledge alone, but selecting appropriate procedures for the CRT-Long problems presumably requires both conceptual and procedural knowledge.

Each trial was coded by two coders, and disagreements between them were resolved, following the same procedure as in Study 1. On the calculation task, the coders agreed whether doubt was displayed on 94% of trials, whether any conceptual knowledge was used on 94% of trials, and which types of conceptual knowledge were used (if any) on 92% of trials. On the CRT-Long, the coders agreed whether doubt was displayed on 88% of trials.

3.2. Results

Table 5 displays descriptives and correlations for our main outcome measures. As in Study 1, the types of conceptual knowledge used most often on the calculation task were decimal place value and cross-notation knowledge (8% and 4% of trials), followed by magnitude knowledge and fraction interpretations (2% and 1% of trials). Doubt and confidence on the calculation task were negatively correlated, providing some support for our measure of doubt. On the CRT-Long, the correlation between doubt and confidence was also negative, but did not reach significance.

Table 5. Descriptives and Correlations for Measures in Study 2.

		Correlations					
	Mean (SD)	1B	1C	1D	2A	2B	2C
Calculation task							
1A. Accuracy	0.82 (0.2)	.65 **	-.34 **	.05	.37 **	.25 *	.13
1B. Confidence	4.22 (0.67)		-.49 **	.09	.31 **	.53 **	-.03
1C. Doubt	0.23 (0.18)			.34 **	.09	-.09	.37 **
1D. Conceptual knowledge use	0.14 (0.17)				.33 **	.20 *	.12
CRT-Long							
2A. Accuracy	0.31 (0.29)					.39 **	.32 **
2B. Confidence	3.55 (0.83)						-.13
2C. Doubt	0.32 (0.23)						

Note. "Accuracy" denotes proportion of trials correctly answered, "doubt" denotes proportion of trials on which doubt was displayed, and "conceptual knowledge use" denotes proportion of trials on which any conceptual knowledge was used; * indicates $p < .05$ and ** indicates $p < .01$.

Below, we report analyses of the calculation task, then analyses of the CRT-Long. Accuracy, doubt, and conceptual knowledge use on the calculation task were analyzed as in Study 1. Accuracy and doubt on the CRT-Long were analyzed as for the calculation task, except that regression analyses of the CRT-Long did not include arithmetic operation or notation as predictors. All reported analyses were preregistered unless marked as exploratory, and all preregistered analyses are reported.

3.2.1. Prediction 1: Use of conceptual knowledge will not be positively associated with accuracy when solving routine math problems

The correlation between use of conceptual knowledge and accuracy on the calculation task was not statistically significant, $r = .05$, $t(135) = 0.6$, $p = .57$ (Table 5). As in Study 1, we calculated the one-tailed left 95% CI of the correlation: [-1, .19] (this analysis was exploratory). This result indicates that conceptual knowledge use and accuracy likely either are not positively correlated or weakly positively correlated.

Similarly, in mixed logistic regression, the effect of conceptual knowledge use on accuracy was not significant, $B = 0.2$, $z = 0.7$, $p = .48$. Correct answers were about equally likely on trials during which participants used or did not use conceptual knowledge (82% vs. 81%).

The regression also found an effect of arithmetic operation, $B = -2.4$, $z = -11.7$, $p < .001$, and a notation * operation interaction, $B = -2.1$, $z = -5.5$, $p < .001$. As in Study 1, participants were more accurate on addition than multiplication problems (93% vs. 70%), and this effect was larger for decimal problems (95% vs. 64%) than for fraction problems (90% vs. 77%).

3.2.2. Prediction 2: Doubt will be positively associated with using conceptual knowledge when solving routine math problems

As in Study 1, the frequency with which participants displayed doubt on the calculation task was positively correlated with how often they used conceptual knowledge, $r = .34$, $t(135) = 4.2$, $p < .001$ (Table 5). Similarly, mixed logistic regression found an effect of doubt on use of conceptual knowledge, $B = 1.5$, $z = 6.9$, $p < .001$, indicating that participants used conceptual knowledge more often on trials during which they displayed doubt than when they did not (25% vs. 10%). The analysis also found an effect of notation, $B = 1.7$, $z = 8.4$, $p < .001$, and a notation * operation interaction, $B = -0.8$, $z = -2.1$, $p = .036$. Participants used conceptual knowledge more often on decimal problems than fraction problems (19% vs. 7%), and this difference was slightly larger for addition (20% vs. 6%) than for multiplication (20% vs. 8%).

3.2.3. Prediction 3: Doubt will be negatively associated with accuracy when solving routine math problems

Consistent with the prediction, how often participants displayed doubt on the calculation task negatively predicted accuracy on that task, $r = -.34$, $t(135) = -4.3$, $p < .001$ (Table 5). Mixed logistic regression with the same model structure as in Study 1 similarly found a negative effect of doubt on accuracy, $B = -1.2$, $z = -6.0$, $p < .001$, indicating that participants answered correctly less often when they displayed doubt than when they did not (62% vs. 88%). The effect of arithmetic operation and the notation * operation interaction reported under Prediction 1 also appeared in this analysis.

3.2.4. Doubt and accuracy on the CRT-Long

In contrast to the calculation task, on the CRT-Long, the frequency with which participants displayed doubt was positively correlated with accuracy, $r = .32$, $t(102) = 3.4$, $p = .001$ (Table 5). Similarly, mixed logistic regression with accuracy on each trial as the dependent variable, whether doubt was displayed as a fixed effect, and participant as a random effect found an effect of doubt, $B = 0.5$, $z = 2.2$, $p = .028$, indicating that participants answered correctly more often on trials during which they displayed doubt than when they did not (39% vs. 26%).

3.3. Discussion

Results of Study 2 were consistent with the main findings of Study 1. Replication of these findings with an adequately powered sample increases confidence in their robustness. Further, the positive relation between doubt and accuracy on the CRT-Long highlights boundary conditions for the negative relations found in both studies between doubt and accuracy on the calculation task. Discussion of these central findings is deferred to the General Discussion.

The findings of Study 2 are informative regarding the validity of our measure of doubt. Rather than assessing (as intended) operation of metacognitive processes relating to feelings of uncertainty and error, this measure could have functioned as a proxy for other variables that are negatively associated with calculation accuracy, such as weak procedural knowledge, high math anxiety, or weak math self-efficacy. Such interpretations, however, would have difficulty accounting for the positive associations found in Study 2 between doubt and accuracy on the CRT-Long. In contrast, those associations are consistent with our interpretation of our measure of doubt as an indicator of metacognitive processes.

Finally, participants' poor performance on the CRT-Long is consistent with previous studies of cognitive reflection by adults (Frederick, 2005; Primi et al., 2016). Interestingly, although accuracy was much lower on the CRT-Long than on the calculation task (31% vs. 82%), expressions of doubt were only slightly more common (32% vs. 23%). Similarly, correlations between confidence ratings and accuracy were much weaker on the CRT-Long than on the calculation task (r = .39 vs. .65). These results dovetail with previous studies that have found relatively weak calibration between confidence and accuracy on nonroutine or unfamiliar problems (Nelson & Fyfe, 2019; Reder & Ritter, 1992).

4. General Discussion

The present studies investigated university students' accuracy, use of conceptual knowledge, and displays of doubt while solving routine calculation problems involving fraction and decimal arithmetic; use of conceptual knowledge while explaining solutions to such problems; and accuracy and displays of doubt while solving nonroutine problems for which apparently obvious answers were incorrect. Below, we discuss implications of the findings regarding relations between procedural knowledge and conceptual knowledge, metacognitive processes as mediators between conceptual and procedural knowledge, and differences between routine and nonroutine problem solving.

4.1. Relations between procedural knowledge and conceptual knowledge

Many theories of routine problem solving in math assume that such problems are solved using procedural knowledge alone (Anderson, 2005; Braithwaite et al., 2017; Brown & VanLehn, 1980; Chen & Campbell, 2018; Rickard, 2005; for exceptions, see Ohlsson & Rees, 1991; Shrager & Siegler, 1998). For example, describing a cognitive model of children's decimal arithmetic calculation, Hiebert & Wearne (1985) stated, "Predictions of performance were made without considering conceptual knowledge, and most of the predictions were verified. It appears that students compute without calling on conceptual understandings" (p. 200).

However, on the calculation task of the present studies, participants used conceptual knowledge to solve problems that they could, in principle, have solved using only procedures. They did so on a small but not trivial proportion of trials (12% in Study 1 and 14% in Study 2). Most participants (58% in Study 1 and 60% in Study 2) did so at least once. A complete theory of adults' fraction and decimal arithmetic, even one restricted to routine calculation, should account for such uses of conceptual knowledge.

Use of conceptual knowledge on the calculation task was either not positively correlated or weakly positively correlated with accuracy. This finding is consistent with our hypothesis that in the context of routine calculation, students recruit conceptual knowledge mainly when their procedural knowledge is weak. Such uses of conceptual knowledge might partially compensate for weak procedural knowledge, but might not lead to higher accuracy than that of students who rely primarily on their knowledge of appropriate procedures for solving the problems at hand.

The above proposal assumes that when solving routine problems, students use procedures when they can and concepts only when they must. Indeed, evidence suggests that students favor procedures over concepts in many contexts. For example, in Perry (1991), fourth and fifth graders displayed higher accuracy on transfer problems involving math equivalence following an intervention that emphasized principles than after one that emphasized both principles and procedures. Perry (1991) argued that students "may ignore the conceptually rich information inherent in the principle when procedures are also provided" (p. 449). Similarly, McNeil (2007) found that children's generation of conceptually correct solutions to math equivalence problems declined following formal instruction and practice in calculation procedures, suggesting that procedural knowledge obtained from instruction "crowded out" conceptual understanding. Finally, some models of simple whole number arithmetic assume that when children are presented an arithmetic problem, they first attempt to retrieve the answer from memory (analogous to using procedural knowledge) and only rely on counting-based backup strategies (analogous to using conceptual knowledge) when retrieval fails (e.g., Siegler, 1988).

Our hypothesis that students are unlikely to recruit conceptual knowledge when their procedural knowledge is strong applies only to routine problems. For nonroutine problems, conceptual knowledge enables adaptation of learned procedures or generation of new ones (Baroody, Feil, & Johnson, 2007; Perry, 1991; Rittle-Johnson & Alibali, 1999). For example, in Rittle-Johnson and Alibali (1999), fourth and fifth graders who received a conceptual intervention regarding math equivalence subsequently invented their own procedures for solving equivalence problems of a type they had not seen before. Such uses of conceptual knowledge for nonroutine problems are likely to occur even among individuals with strong procedural knowledge. We return to the distinction between routine and nonroutine problems in Section 4.3.

Conceptual knowledge use during explanation was positively correlated with calculation accuracy (Study 1), consistent with relations that have previously been found between conceptual knowledge of rational numbers and rational number arithmetic accuracy (Bailey et al., 2017; Gabriel et al., 2013; Hecht & Vagi, 2010; Siegler & Pyke, 2013; Siegler et al., 2011). The finding is also consistent with theoretical proposals that conceptual knowledge supports procedural knowledge (e.g., Rittle-Johnson, 2017; Rittle-Johnson et al., 2015).

Conceptual knowledge could contribute to performance on procedural tasks through either online or offline mechanisms. In an online mechanism, individuals use conceptual knowledge while performing procedural tasks, for example to aid in strategy selection or error detection, resulting in higher accuracy on the procedural tasks. In an offline mechanism, conceptual knowledge facilitates acquisition of procedural knowledge, such as learning from instruction or practice. In this case, conceptual knowledge could lead to improved accuracy on procedural tasks without conceptual knowledge being used while performing those tasks.

If an online mechanism generated the correlation between calculation accuracy and use of conceptual knowledge during explanation in Study 1, then that correlation should have appeared primarily among individuals who used conceptual knowledge on the calculation task. However, the opposite was true: the correlation appeared only among individuals who did *not* use conceptual knowledge during calculation. This fact does not imply that online uses of conceptual knowledge during calculation did not occur or that they were not helpful, but does suggest that such uses of conceptual knowledge did not cause the correlation between conceptual knowledge and calculation accuracy. This correlation appears more consistent with an offline mechanism.

Online and offline effects are not mutually exclusive in principle. However, if we are correct in proposing that students generally prefer procedures over concepts for solving routine problems, then factors that facilitate procedure learning should also reduce reliance on concepts in this context. Thus, strong conceptual knowledge could contribute to procedural proficiency and thereby obviate the need to use conceptual knowledge during routine problem solving. This possibility is analogous to how, in some theories of skill acquisition, declarative knowledge is required for initial acquisition of procedures, but reliance on declarative knowledge during task performance decreases as skills are increasingly automatized (Anderson, 2013).

4.2. Metacognitive processes as mediators between conceptual and procedural knowledge

Metacognitive processes are considered particularly important for solving complex, nonroutine problems (Carr et al., 1994; Garofalo & Lester, 1985; W. Schneider & Artelt, 2010; Schoenfeld, 1992). Although the calculation task in the present studies involved relatively simple routine problems, participants displayed doubt on a small but not trivial proportion of trials of that task (21% in Study 1, 23% in Study 2). Most participants (92% in Study 1, 88% in Study 2) did so at least once. These displays of doubt demonstrate involvement of metacognitive processes, including feelings of uncertainty and feelings of error, during routine problem solving.

Doubt was negatively associated with accuracy, and positively associated with overt use of conceptual knowledge, on the calculation task. These results were predicted based on our hypothesis that doubt mediates interactions between conceptual and procedural knowledge. As proposed in the Introduction, such interactions may occur in at least two ways. First, failure to retrieve an appropriate procedure or retrieval of multiple procedures may generate feelings of uncertainty that trigger recruitment of conceptual knowledge. Second, conceptual knowledge may enable detection of errors, generating feelings of error that may lead students to rerun or modify their solution procedures.

Correlations between doubt and use of conceptual knowledge were moderate (r = .26 and .34 in Studies 1 and 2), indicating that doubt and conceptual knowledge do not always go together. Students might feel uncertain how to solve a problem but forge ahead with their best guess without using conceptual knowledge because they lack relevant conceptual knowledge or do not wish to think deeply. Feelings of error might occur for non-conceptual reasons, such as the configuration of symbols on the page appearing unfamiliar. Thus, the proposed mechanisms involving doubt and conceptual knowledge are only part of the whole picture.

Our proposal is similar to that of Crowley, Shrager, and Siegler (1997), who argued that children's problem solving involves a competition between fast associative processes, such as fact retrieval and automatized procedures, and slower explicit reasoning based on conceptual knowledge, which Crowley et al. (1997) called "metacognitive processes." Associative processes typically generate an initial answer, which is accepted if accompanied by high confidence; otherwise, metacognitive processes may generate a solution. Thus, feelings of confidence mediate between different processes in Crowley et al. (1997), as in the present proposal. However, our proposal separates what Crowley et al. (1997) called "metacognitive processes" into two parts, conceptual knowledge and metacognitive processes proper. These two parts appear to be distinct, though related, as evidenced by the moderate positive correlations found in the present studies between doubt and conceptual knowledge use on the calculation task.

Our proposal is also analogous to recent proposals in dual-process theories of reasoning (Evans, 2019; Thompson, Prowse Turner, & Pennycook, 2011). For example, in the default-interventionist model of Evans (2019), fast, intuitive Type 1 processes generate default responses that are accepted or revised by slow, reflective Type 2 processes. Initial intuitive responses are accompanied by *Feelings of Rightness* (FoR); weaker FoRs lead to greater involvement by Type 2 processes. This modulation of Type 2 involvement is carried out by separate monitoring and control processes, which Evans (2019) called "Type 3." We propose that procedures, like Type 1 processes, are the default approach for solving routine problems; that conceptual knowledge, like Type 2 processes, is most involved when one feels doubt (analogous to low FoR) about the default response; and that metacognitive processes modulate the involvement of conceptual knowledge as Type 3 processes modulate involvement of Type 2 processes.

Despite these similarities, our proposal diverges from that of Evans (2019) in several ways. First, Type 1 processes are fast and do not use working memory, whereas execution of mathematical procedures may be slow and require working memory. Second, Evans (2019) describes Type 3 processes as unconscious, whereas in our proposal, metacognitive processes may be conscious. Third, in Evans' (2019) model, the flow of information among processes is unidirectional: Type 1 processes generate FoRs, which feed into Type 3 processes, which govern the effort devoted to Type 2 processing. In our proposal, information flows bidirectionally: procedural knowledge can generate doubt leading to recruitment of conceptual knowledge, but conceptual knowledge also can generate doubt leading to rerunning or modifying a procedure. Future research should investigate both similarities and differences between interactions of Type 1 and 2 processes in reasoning and interactions of procedural and conceptual knowledge in math.

4.3. Routine and nonroutine problem solving

Study 2 found a positive correlation between participants' accuracies on the rational number arithmetic calculation task and the CRT-Long. This result points to commonalities between routine and nonroutine problem solving in math. Several other studies have also found positive associations between performance on math assessments and on various versions of the CRT (Primi et al., 2016; Sobkow, Olszewska, & Traczyk, 2020; Young & Shtulman, 2020).

However, the present findings suggest that metacognitive processes, specifically doubt, serve different functions when solving routine and nonroutine problems. Doubt was negatively associated with accuracy on the calculation task, but positively associated with accuracy on the CRT-Long. To explain the former finding, we proposed that when solving routine problems such as those in the calculation task, doubt often reflects weak procedural knowledge. However, for nonroutine problems, such as those in the CRT-Long, procedural knowledge is inadequate by definition (i.e., nonroutine problems by definition cannot be solved correctly merely by straightforward application of learned procedures). In this context, overt displays of doubt may signal a student recognizing that the solution is not obvious and consequently deciding consciously to regulate and/or monitor the problem solving process. This approach might be unnecessary when one is solving routine problems and has strong procedural knowledge, but is likely adaptive when solving nonroutine problems.

The above proposal implies that metacognitive skills play a more causal role in nonroutine than routine problem solving. In routine problem solving, a student with strong procedural knowledge might perform well regardless of metacognitive skill; the direct cause of error on such problems is likely to be weak procedural knowledge (though metacognitive skills and conceptual knowledge might help students to cope with weak procedural knowledge). In nonroutine problem solving, performance likely depends much more on whether and how well students regulate and monitor their thinking (Carr et al., 1994; Garofalo & Lester, 1985; W. Schneider & Artelt, 2010; Schoenfeld, 1992); the cause of error on such problems might often be poor metacognitive skill. This hypothesis implies that individual differences in metacognitive skill should be more strongly correlated with differences in performance on nonroutine than routine problems, and interventions that emphasize metacognitive skills (e.g., Desoete et al., 2003; Hacker et al., 2019) should affect performance on nonroutine problems more than on routine ones, a prediction worth testing in the future.

Finally, even among nonroutine problems, benefits of metacognitive skill are likely to vary as a function of metacognitive and procedural demands. The problems on the CRT-Long are metacognitively demanding, because they elicit strong but incorrect intuitions, but not very procedurally demanding (for university students), because they require only simple calculations such as $1.05 + $0.05 = $1.10. Metacognitive processes, such as feelings of doubt, might have less effect on accuracy for nonroutine problems that are less metacognitively demanding and/or more procedurally demanding than those on the CRT-Long.

4.4. Limitations

Unobserved variables could account for some or all of the correlations found in the present studies. This limitation is partially addressed by the converging results obtained through mixed logistic regressions, which tested for between-trial effects within participants. Nevertheless, future research should test whether the relations found among accuracy, doubt, and use of conceptual knowledge still appear when controlling for other aspects of individual differences, such as executive function, fluid intelligence, verbal ability, and motivation.

The present studies relied on think-aloud protocols and written work to identify displays of doubt and uses of conceptual knowledge. This approach has advantages (as described in the Introduction), but could not detect feelings of doubt or uses of conceptual knowledge that did not result in overt displays, for example because a participant was not verbally expressive or used conceptual knowledge unconsciously. Future research should test the robustness of the present findings using a wider range of measures of doubt and conceptual knowledge.

4.5. Conclusion

Students used conceptual knowledge and displayed evidence of metacognitive processes while performing routine calculations involving fractions and decimals. Thus, procedural knowledge does not operate in a vaccuum; conceptual knowledge and metacognition are sometimes involved even in apparently procedural tasks. Further, conceptual knowledge and metacognition in math appear to be connected, like Type 2 and Type 3 processes in reasoning. It is hoped that these findings will motivate the development of formal theories describing mechanisms of interaction between conceptual and procedural knowledge and metacognition in fraction and decimal arithmetic, and in other mathematical domains as well.

5. References

Alibali, M. W., Brown, S. A., & Menendez, D. (2019). *Understanding Strategy Change: Contextual, Individual, and Metacognitive Factors*. *Advances in Child Development and Behavior* (1st ed., Vol. 56). Elsevier Inc. https://doi.org/10.1016/bs.acdb.2018.11.004

Anderson, J. R. (2005). Human symbol manipulation within an integrated cognitive architecture. *Cognitive Science*, *29*(3), 313–341. https://doi.org/10.1207/s15516709cog0000_22

Anderson, J. R. (2013). *The Architecture of Cognition*. Psychology Press.

Bailey, D. H., Hansen, N., & Jordan, N. C. (2017). The codevelopment of children's fraction arithmetic skill and fraction magnitude understanding. *Journal of Educational Psychology*, *109*(4), 509–519. https://doi.org/10.1037/edu0000152

Baroody, A. J. (2003). The development of adaptive expertise and flexibility: The integration of conceptual and procedural knowledge. In A. J. Baroody & A. Dowker (Eds.), *The*

Development of Arithmetic Concepts and Skills: Constructing Adaptive Expertise (pp. 1–34). Mahwah, NJ: Erlbaum.

Baroody, A. J., Feil, Y., & Johnson, A. R. (2007). An alternative reconceptualization of procedural and conceptual knowledge. *Journal for Research in Mathematics Education*, *38*(2), 115–131. https://doi.org/10.2307/30034952

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2013). lme4: Linear mixed-effects models using Eigen and S4. Retrieved from http://cran.r-project.org/web/packages/lme4/index.html

Behr, M. J., Lesh, R., Post, T. R., & Silver, E. A. (1983). Rational Number Concepts. In *Acquisition of Mathematics Concepts and Processes* (pp. 1–41).

Braithwaite, D. W., McMullen, J., & Hurst, M. A. (in press). Cross-notation knowledge of rational numbers. *Journal of Experimental Child Psychology*.

Braithwaite, D. W., Pyke, A. A., & Siegler, R. S. (2017). A computational model of fraction arithmetic. *Psychological Review*, *124*(5), 603–625. https://doi.org/10.1037/rev0000072

Braithwaite, D. W., & Siegler, R. S. (2021). Putting fractions together. *Journal of Educational Psychology*, *113*(3), 556–571. https://doi.org/10.1037/edu0000477

Braithwaite, D. W., Sprague, L., & Siegler, R. S. (in press). Toward a unified theory of rational number arithmetic. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Brown, J. S., & VanLehn, K. (1980). Repair theory: A generative theory of bugs in procedural skills. *Cognitive Science*, *4*(4), 379–426. https://doi.org/10.1207/s15516709cog0404_3

Canobi, K. H. (2009). Concept–procedure interactions in children's addition and subtraction. *Journal of Experimental Child Psychology*, *102*(2), 131–149. https://doi.org/10.1016/j.jecp.2008.07.008

Carr, M., Alexander, J., & Folds-Bennett, T. (1994). Metacognition and mathematics strategy use. *Applied Cognitive Psychology*, *8*(6), 583–595. https://doi.org/10.1002/acp.2350080605

Chen, Y., & Campbell, J. I. D. (2018). “Compacted” procedures for adults’ simple addition: A review and critique of the evidence. *Psychonomic Bulletin & Review*, *25*(2), 739–753. https://doi.org/10.3758/s13423-017-1328-2

Common Core State Standards Initiative. (2010). *Common core state standards for mathematics*. Washington, D.C.: National Governors Association Center for Best Practices and the Council of Chief State School Officers. Retrieved from http://www.corestandards.org/math

Crowley, K., Shrager, J., & Siegler, R. S. (1997). Strategy discovery as a competitive negotiation between metacognitive and associative mechanisms. *Developmental Review*, *17*, 462–489. https://doi.org/10.1006/drev.1997.0442

Desoete, A., Roeyers, H., & De Clercq, A. (2003). Can offline metacognition enhance mathematical problem solving? *Journal of Educational Psychology*, *95*(1), 188–200. https://doi.org/10.1037/0022-0663.95.1.188

Evans, J. S. B. T. (2019). Reflections on reflection: the nature and function of type 2 processes in dual-process theories of reasoning. *Thinking & Reasoning*, *25*(4), 383–415. https://doi.org/10.1080/13546783.2019.1623071

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Fazio, L. K., DeWolf, M., & Siegler, R. S. (2016). Strategy use and strategy choice in fraction magnitude comparison. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(1), 1–16. https://doi.org/10.1037/xlm0000153

Fernandez Cruz, A. L., Arango-Muñoz, S., & Volz, K. G. (2016). Oops, scratch that! Monitoring one's own errors during mental calculation. *Cognition*, *146*, 110–120. https://doi.org/10.1016/j.cognition.2015.09.005

Fitzsimmons, C. J., Thompson, C. A., & Sidney, P. G. (2020). Confident or familiar? The role of familiarity ratings in adults' confidence judgments when estimating fraction magnitudes. *Metacognition and Learning*, *15*(2), 215–231. https://doi.org/10.1007/s11409-020-09225-9

Fox, M. C., Ericsson, K. A., & Best, R. (2011). Do procedures for verbal reporting of thinking have to be reactive? A meta-analysis and recommendations for best reporting methods. *Psychological Bulletin*, *137*(2), 316–344. https://doi.org/10.1037/a0021663

Frederick, S. (2005). Cognitive Reflection and Decision Making. *Journal of Economic Perspectives*, *19*(4), 25–42. https://doi.org/10.1257/089533005775196732

Fuchs, L. S., Geary, D. C., Compton, D. L., Fuchs, D., Hamlett, C. L., & Bryant, J. D. (2010). The contributions of numerosity and domain-general abilities to school readiness. *Child Development*, *81*(5), 1520–1533. https://doi.org/10.1111/j.1467-8624.2010.01489.x

Fuchs, L. S., Schumacher, R. F., Long, J., Namkung, J., Hamlett, C. L., Cirino, P. T., … Changas, P. (2013). Improving at-risk learners' understanding of fractions. *Journal of Educational Psychology*, *105*(3), 683–700. https://doi.org/10.1037/a0032446

Fuchs, L. S., Wang, A. Y., Preacher, K. J., Malone, A. S., Fuchs, D., & Pachmayr, R. (2021). Addressing challenging mathematics standards with at-risk learners: A randomized controlled trial on the effects of fractions intervention at third grade. *Exceptional Children*, *87*(2), 163–182. https://doi.org/10.1177/0014402920924846

Fyfe, E. R., DeCaro, M. S., & Rittle-Johnson, B. (2014). An alternative time for telling: When conceptual instruction prior to problem solving improves mathematical knowledge. *British*

Journal of Educational Psychology, *84*(3), 502–519. https://doi.org/10.1111/bjep.12035

Gabriel, F. C., Coché, F., Szucs, D., Carette, V., Rey, B., & Content, A. (2013). A componential view of children's difficulties in learning fractions. *Frontiers in Psychology*, *4*(715), 1–12. https://doi.org/10.3389/fpsyg.2013.00715

Garofalo, J., & Lester, F. K. J. (1985). Metacognition, cognitive monitoring, and mathematical performance. *Journal for Research in Mathematics Education*, *16*(3), 163–176. https://doi.org/10.5951/jresematheduc.16.3.0163

Hacker, D. J., Kiuhara, S. A., & Levin, J. R. (2019). A metacognitive intervention for teaching fractions to students with or at-risk for learning disabilities in mathematics. *ZDM*, *51*(4), 601–612. https://doi.org/10.1007/s11858-019-01040-0

Handel, M. J. (2016). What do people do at work? *Journal for Labour Market Research*, *49*(2), 177–197. https://doi.org/10.1007/s12651-016-0213-1

Hansen, N., Jordan, N. C., & Rodrigues, J. (2015). Identifying learning difficulties with fractions: A longitudinal study of student growth from third through sixth grade. *Contemporary Educational Psychology*, 1–15. https://doi.org/10.1016/j.cedpsych.2015.11.002

Hecht, S. A., & Vagi, K. J. (2010). Sources of group and individual differences in emerging fraction skills. *Journal of Educational Psychology*, *102*(4), 843–859. https://doi.org/10.1037/a0019824

Hecht, S. A., & Vagi, K. J. (2012). Patterns of strengths and weaknesses in children's knowledge about fractions. *Journal of Experimental Child Psychology*, *111*(2), 212–229. https://doi.org/10.1016/j.jecp.2011.08.012

Hiebert, J., & Lefevre, P. (1986). Conceptual and procedural knowledge in mathematics: An introductory analysis. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: The case of mathematics* (pp. 1–27). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc. Retrieved from https://www.taylorfrancis.com/books/e/9781136559761/chapters/10.4324%2F9780203063538-6

Hiebert, J., & Wearne, D. (1985). A model of students' decimal computation procedures. *Cognition and Instruction*, *2*(3), 175–205. https://doi.org/10.1080/07370008.1985.9648916

Hurst, M. A., & Cordes, S. (2016). Rational-number comparison across notation: Fractions, decimals, and whole numbers. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(2), 281–293. https://doi.org/10.1037/xhp0000140

Hurst, M. A., & Cordes, S. (2018a). A systematic investigation of the link between rational number processing and algebra ability. *British Journal of Psychology*, *109*(1), 99–117. https://doi.org/10.1111/bjop.12244

Hurst, M. A., & Cordes, S. (2018b). Children's understanding of fraction and decimal symbols and the notation-specific relation to pre-algebra ability. *Journal of Experimental Child Psychology*, *168*, 32–48. https://doi.org/10.1016/j.jecp.2017.12.003

Kieren, T. E. (1980). The rational number construct--Its elements and mechanisms. In T. E. Kieren (Ed.), *Recent research on number learning* (pp. 125–150). Columbus, OH: Information Reference Center (ERIC/IRC), The Ohio State University. Retrieved from https://eric.ed.gov/?id=ED212463

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2016). lmerTest: Tests in Linear Mixed Effects Models.

Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence Testing for Psychological

Research: A Tutorial. *Advances in Methods and Practices in Psychological Science*, *1*(2), 259–269. https://doi.org/10.1177/2515245918770963

Lortie-Forgues, H., Tian, J., & Siegler, R. S. (2015). Why is learning fraction and decimal arithmetic so difficult? *Developmental Review*, *38*, 201–221. https://doi.org/10.1016/j.dr.2015.07.008

Mack, N. K. (1995). Confounding whole-number and fraction concepts when building on informal knowledge. *Journal for Research in Mathematics Education*, *26*(5), 422–441. https://doi.org/10.2307/749431

Martin, T., Petrick Smith, C., Forsgren, N., Aghababyan, A., Janisiewicz, P., & Baker, S. (2015). Learning fractions by splitting: Using learning analytics to illuminate the development of mathematical understanding. *Journal of the Learning Sciences*, *24*(4), 593–637. https://doi.org/10.1080/10508406.2015.1078244

McNeil, N. M. (2007). U-shaped development in math: 7-year-olds outperform 9-year-olds on equivalence problems. *Developmental Psychology*, *43*(3), 687–695. https://doi.org/10.1037/0012-1649.43.3.687

Moss, J., & Case, R. (1999). Developing children's understanding of the rational numbers: A new model and an experimental curriculum. *Journal for Research in Mathematics Education*, *30*(2), 122. https://doi.org/10.2307/749607

Nelson, L. J., & Fyfe, E. R. (2019). Metacognitive monitoring and help-seeking decisions on mathematical equivalence problems. *Metacognition and Learning*, *14*(2), 167–187. https://doi.org/10.1007/s11409-019-09203-w

Newton, K. J. (2008). An extensive analysis of preservice elementary teachers' knowledge of fractions. *American Educational Research Journal*, *45*(4), 1080–1110. https://doi.org/DOL

10.3102/0002831208320851

Ohlsson, S. (1996). Learning from performance errors. *Psychological Review*, *103*(2), 241–262. https://doi.org/10.1037/0033-295X.103.2.241

Ohlsson, S., & Rees, E. (1991). The function of conceptual understanding in the learning of arithmetic procedures. *Cognition and Instruction*, *8*(2), 103–179. https://doi.org/10.1207/s1532690xci0802_1

Opfer, J. E., & Devries, J. M. (2008). Representational change and magnitude estimation: Why young children can make more accurate salary comparisons than adults. *Cognition*, *108*(3), 843–849. https://doi.org/10.1016/j.cognition.2008.05.003

Perry, M. (1991). Learning and transfer: Instructional conditions and conceptual change. *Cognitive Development*, *6*(4), 449–468. https://doi.org/10.1016/0885-2014(91)90049-J

Primi, C., Morsanyi, K., Chiesi, F., Donati, M. A., & Hamilton, J. (2016). The development and testing of a new version of the Cognitive Reflection Test applying Item Response Theory (IRT). *Journal of Behavioral Decision Making*, *29*(5), 453–469. https://doi.org/10.1002/bdm.1883

Reder, L. M., & Ritter, F. E. (1992). What determines initial feeling of knowing? Familiarity with question terms, not with the answer. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(3), 435–451. https://doi.org/10.1037/0278-7393.18.3.435

Rickard, T. C. (2005). A revised identical elements model of arithmetic fact representation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(2), 250–257. https://doi.org/10.1037/0278-7393.31.2.250

Rittle-Johnson, B. (2017). Developing mathematics knowledge. *Child Development Perspectives*, *11*(3), 184–190. https://doi.org/10.1111/cdep.12229

Rittle-Johnson, B., & Alibali, M. W. (1999). Conceptual and procedural knowledge of mathematics: Does one lead to the other? *Journal of Educational Psychology*, *91*(1), 175–189. https://doi.org/10.1037/0022-0663.91.1.175

Rittle-Johnson, B., & Koedinger, K. R. (2009). Iterating between lessons on concepts and procedures can improve mathematics knowledge. *The British Journal of Educational Psychology*, *79*(3), 483–500. https://doi.org/10.1348/000709908X398106

Rittle-Johnson, B., Schneider, M., & Star, J. R. (2015). Not a one-way street: Bidirectional relations between procedural and conceptual knowledge of mathematics. *Educational Psychology Review*, *27*(4), 587–597. https://doi.org/10.1007/s10648-015-9302-x

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. https://doi.org/10.1037/0022-0663.93.2.346

Schneider, M., Rittle-Johnson, B., & Star, J. R. (2011). Relations among conceptual knowledge, procedural knowledge, and procedural flexibility in two samples differing in prior knowledge. *Developmental Psychology*, *47*(6), 1525–1538. https://doi.org/10.1037/a0024997

Schneider, W., & Artelt, C. (2010). Metacognition and mathematics education. *ZDM*, *42*(2), 149–161. https://doi.org/10.1007/s11858-010-0240-2

Schoenfeld, A. H. (1992). Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 334–370). New York, NY: MacMillan.

Shrager, J., & Siegler, R. S. (1998). SCADS: A model of children's strategy choices and strategy discoveries. *Psychological Science*, *9*(5), 405–410. https://doi.org/10.1111/1467-

9280.00076

Sidney, P. G. G., Thalluri, R., Buerke, M. L. L., & Thompson, C. A. A. (2019). Who uses more strategies? Linking mathematics anxiety to adults' strategy variability and performance on fraction magnitude tasks. *Thinking & Reasoning*, *25*(1), 94–131. https://doi.org/10.1080/13546783.2018.1475303

Siegler, R. S. (1988). Strategy choice procedures and the development of multiplication skill. *Journal of Experimental Psychology: General*, *117*(3), 258–275. https://doi.org/10.1037/0096-3445.117.3.258

Siegler, R. S., Duncan, G. J., Davis-Kean, P. E., Duckworth, K., Claessens, A., Engel, M., … Chen, M. (2012). Early predictors of high school mathematics achievement. *Psychological Science*, *23*(7), 691–697. https://doi.org/10.1177/0956797612440101

Siegler, R. S., Im, S., Schiller, L. K., Tian, J., & Braithwaite, D. W. (2020). The sleep of reason produces monsters: How and when biased input shaped mathematics learning. *Annual Review of Developmental Psychology*, *2*, 413–435. https://doi.org/10.1146/annurev-devpsych-041620-031544

Siegler, R. S., & Lortie-Forgues, H. (2015). Conceptual knowledge of fraction arithmetic. *Journal of Educational Psychology*, *107*(3), 909–918. https://doi.org/10.1037/edu0000025

Siegler, R. S., & Pyke, A. A. (2013). Developmental and individual differences in understanding of fractions. *Developmental Psychology*, *49*(10), 1994–2004. https://doi.org/10.1037/a0031200

Siegler, R. S., & Ramani, G. B. (2009). Playing linear number board games—but not circular ones—improves low-income preschoolers' numerical understanding. *Journal of Educational Psychology*, *101*(3), 545. https://doi.org/10.1037/a0014239

Siegler, R. S., & Thompson, C. A. (2014). Numerical landmarks are useful--except when they're not. *Journal of Experimental Child Psychology*, *120*, 39–58. https://doi.org/10.1016/j.jecp.2013.11.014

Siegler, R. S., Thompson, C. A., & Schneider, M. (2011). An integrated theory of whole number and fractions development. *Cognitive Psychology*, *62*(4), 273–296. https://doi.org/10.1016/j.cogpsych.2011.03.001

Sobkow, A., Olszewska, A., & Traczyk, J. (2020). Multiple numeric competencies predict decision outcomes beyond fluid intelligence and cognitive reflection. *Intelligence*, *80*, 101452. https://doi.org/10.1016/J.INTELL.2020.101452

Steffe, L. P. (2004). On the construction of learning trajectories of children: The case of commensurate fractions. *Mathematical Thinking and Learning*, *6*(2), 129–162. https://doi.org/10.1207/s15327833mtl0602_4

Stigler, J., Givvin, K., & Thompson, A. (2010). What community college developmental mathematics students understand about mathematics. *MathAMATYC Educator*, *1*(3), 4–16. https://doi.org/10.1016/j.learninstruc.2014.03.002

Stillman, G. A., & Galbraith, P. L. (1998). Applying mathematics with real world connections: metacognitive characteristics of secondary students. *Educational Studies in Mathematics*, *36*(2), 157–194. https://doi.org/10.1023/A:1003246329257

Thompson, V. A., Prowse Turner, J. A., & Pennycook, G. (2011). Intuition, reason, and metacognition. *Cognitive Psychology*, *63*(3), 107–140. https://doi.org/10.1016/j.cogpsych.2011.06.001

Tian, J., Braithwaite, D. W., & Siegler, R. S. (2021). Distributions of textbook problems predict student learning: Data from decimal arithmetic. *Journal of Educational Psychology*, *113*(3), 516–529. https://doi.org/10.1037/edu0000618

Tzur, R. (1999). An integrated study of children's contruction of improper fractions and the teacher's role in promoting that learning. *Journal for Research in Mathematics Education*, *30*(4), 390–416. https://doi.org/10.2307/749707

Tzur, R., & Hunt, J. (2015). Iteration: Unit Fraction Knowledge and the French Fry Tasks. *Teaching Children Mathematics*, *22*(3), 148. https://doi.org/10.5951/teacchilmath.22.3.0148

Wearne, D., & Hiebert, J. (1988). A cognitive approach to meaningful mathematics instruction: Testing a local theory using decimal numbers. *Journal for Research in Mathematics Education*, *19*(5), 371–384. https://doi.org/10.2307/749172

Wilkins, J. L. M., & Norton, A. (2018). Learning progression toward a measurement concept of fractions. *International Journal of STEM Education*, *5*(1), 27. https://doi.org/10.1186/s40594-018-0119-2

Wong, H., & Odic, D. (2021). The intuitive number sense contributes to symbolic equation error detection abilities. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *47*(1), 1–10. https://doi.org/10.1037/xlm0000803

Young, A. G., & Shtulman, A. (2020). Children's cognitive reflection predicts conceptual understanding in science and mathematics. *Psychological Science*, *31*(11), 1396–1408. https://doi.org/10.1177/0956797620954449