

Journal of Educational Psychology

Feedback Both Helps and Hinders Learning: The Causal Role of Prior Knowledge

Emily R. Fyfe and Bethany Rittle-Johnson

Online First Publication, June 8, 2015. <http://dx.doi.org/10.1037/edu0000053>

CITATION

Fyfe, E. R., & Rittle-Johnson, B. (2015, June 8). Feedback Both Helps and Hinders Learning: The Causal Role of Prior Knowledge. *Journal of Educational Psychology*. Advance online publication. <http://dx.doi.org/10.1037/edu0000053>

Feedback Both Helps and Hinders Learning: The Causal Role of Prior Knowledge

Emily R. Fyfe and Bethany Rittle-Johnson
Vanderbilt University

Feedback can be a powerful learning tool, but its effects vary widely. Research has suggested that learners' prior knowledge may moderate the effects of feedback; however, no causal link has been established. In Experiment 1, we randomly assigned elementary schoolchildren ($N = 108$) to a condition based on a crossing of 2 factors: induced strategy knowledge (yes vs. no) and immediate, verification feedback (present vs. absent). Feedback had positive effects for children who were not taught a correct strategy, but negative effects for children with induced knowledge of a correct strategy. In Experiment 2, we induced strategy knowledge in all children ($N = 101$) and randomly assigned them to 1 of 3 conditions: no feedback, immediate correct-answer feedback, or summative correct-answer feedback. Again, feedback had negative effects relative to no feedback. Results provide evidence for a causal role of prior knowledge and indicate that minimal feedback can both help and hinder learning.

Keywords: feedback, problem solving, prior knowledge, mathematics learning

Feedback is a ubiquitous learning tool that has been studied by cognitive scientists, learning theorists, and educational psychologists alike (e.g., Hattie & Timperley, 2007; Kluger & DeNisi, 1996). It is broadly defined as any information about performance or understanding that the learner can use to confirm, reject, or modify prior knowledge (Mory, 2004). In this study, we focus on two common types of feedback: right-wrong verification and verification plus the correct answer. However, the amount of information can vary on a continuum from simple right-wrong verification to more elaborate explanations, such as a conceptual rationale of the correct answer or a hint at a correct problem-solving procedure. Feedback is theorized to benefit learning by reinforcing correct responses (Smith & Kimball, 2010), reducing perseveration on incorrect responses (Kulhavy, 1977), and facilitating the generation of correct alternatives (Butler & Winne, 1995). Further, meta-analyses confirm its powerful influence on learning (e.g., Kluger & DeNisi, 1996). For example, a recent analysis reported an average positive effect size of .46 for feedback relative to no feedback conditions (Alfieri, Brooks, Aldrich, & Tenenbaum, 2011).

In addition to theoretical and empirical support, there also are practical reasons for the popularity of feedback. Feedback can be applied by parents and teachers in nearly any learning situation. Indeed, parents frequently provide feedback to their children on learning tasks at home (e.g., Evans, Barraball, & Eberle, 1998; Hoover-Dempsey et al., 2001), and teachers also provide feedback on student performance in the classroom (e.g., Pianta, Belsky, Houts, & Morrison, 2007). In general, feedback is often assumed to be helpful and many agree that "the importance of feedback in promoting learning is inarguable" (Moreno, 2004, p. 100).

Despite broad endorsement of feedback, research has indicated that the effects of feedback vary considerably and are not universally beneficial (see Mory, 2004). For example, in two meta-analyses, feedback had mostly positive effects, but neutral or negative effects in a third of the cases (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Kluger & DeNisi, 1996). Negative effects occur when feedback leads to lower learning outcomes compared to no feedback. Feedback is theorized to have negative effects when it reduces mindfulness (e.g., overrelying on the feedback; Butler & Winne, 1995), draws attention to the self (e.g., evaluating one's abilities; Kluger & DeNisi, 1996), or produces cognitive interference (e.g., confusing one's response with the correct one; Kulhavy, 1977). However, the majority of feedback research is with adults in lab contexts recalling test-like material (e.g., multiple choice, list learning), and feedback may function differently for children generating problem solutions. Further, learner characteristics that interact with feedback (i.e., moderators) have rarely been experimentally tested. The goal of the current research was to experimentally test one potential moderator in the context of children's problem solving to better predict when feedback will help versus harm learning.

There are several reasons to suggest that prior knowledge is a key moderator to consider in the context of feedback. First, nearly all theoretical models of feedback give prior knowledge a primary role (e.g., Mory, 2004; Narciss & Huth, 2004). In particular, learning from feedback is viewed as an interaction between infor-

Emily R. Fyfe and Bethany Rittle-Johnson, Department of Psychology and Human Development, Vanderbilt University.

Emily R. Fyfe was supported by a Graduate Research Fellowship from the National Science Foundation. The research described in this paper was supported in part by National Science Foundation Grant DRL-9746565 to Bethany Rittle-Johnson, Institute of Education Sciences, U.S. Department of Education, Training Grant R305B080025, and a dissertation research award from the American Psychological Association to Emily R. Fyfe. We thank Emilie Hall, Emily Litzow, and Hannah Wolfe for their help with data collection and coding as well as the teachers and children at Westmeade Elementary, Carter Lawrence Elementary, Lockeland Elementary, and Saint Ann School for their participation.

Correspondence concerning this article should be addressed to Emily R. Fyfe, 230 Appleton Place #552, Vanderbilt University, Nashville TN 37203. E-mail: emily.r.fyfe@vanderbilt.edu

mation in long-term memory (i.e., prior knowledge) and the new information provided in the feedback message. Second, prior knowledge determines the effectiveness of various instructional techniques, such that a technique that is effective for low-knowledge learners loses its benefits for high-knowledge learners (see Kalyuga, 2007).

Third, evidence from multiple experimental studies suggests that prior knowledge often predicts learning from feedback (Fyfe, Rittle-Johnson, & DeCaro, 2012; Gielen, Peeters, Dochy, Onghena, & Struyven, 2010; Krause, Stark, & Mandl, 2009; Luwel, Foustana, Papadatos, & Verschaffel, 2011; Nihalani, Mayrath, & Robinson, 2011). For example, undergraduate students with low prior knowledge of statistics exhibited higher learning on a posttest if they received explicit feedback during training than if they did not. However, those with higher prior knowledge performed just as well when feedback was not provided (Krause et al., 2009). In problem-solving domains, prior knowledge of correct strategies seems particularly relevant. For example, for children with low knowledge of target math problems at pretest, feedback facilitates the generation of diverse strategies relative to no feedback (Alibali, 1999; Fyfe et al., 2012). In contrast, adolescents with high knowledge of correct strategies at pretest gain similar strategy knowledge whether feedback is provided or not (Hofer, Nussbaumer, & Schneider, 2011; Nussbaumer, Schneider, & Stern, 2014).

The possibility that feedback has negative effects for some learners is of more concern. Preliminary evidence for this comes from two previous experiments in which second- and third-grade children solved novel math problems prior to receiving instruction (Fyfe et al., 2012). During problem solving, some children received feedback after each problem while others did not. For children with low prior knowledge of correct strategies, feedback facilitated posttest problem solving. But, for children with moderate prior knowledge, feedback hindered accuracy relative to no feedback. This occurred even though most of these “moderate-knowledge” children used correct strategies on less than 40% of problems at pretest. The effects were maintained 2 weeks later, and did not depend on feedback type. For example, feedback that contained the correct answer and feedback that provided right-wrong verification only yielded similar results.

Given the counterintuitive nature of these results, more work is needed to verify and clarify the conclusions. First, no causal link has been established between learners’ prior knowledge and the effects of feedback on learning. Previous studies have relied on preexisting indicators of prior knowledge (e.g., researcher-created pretests; Fyfe et al., 2012; Gielen et al., 2010; Krause et al., 2009). However, these learners may vary on a number of factors (e.g., motivation, intelligence) that influence their response to feedback. In the current study, we used a prefamiliarization technique, in which some learners were exposed to the target material and others were not (Petersen & McNeil, 2013; Rey & Buchwald, 2011). This avoided confounding variables and allowed for random assignment to prior knowledge condition (Tobias, 2010).

A second issue is the need to specify the type and level of prior knowledge at which feedback may have negative effects. Past research has relied on a median split to classify low- and high-knowledge learners (e.g., Fyfe et al., 2012; Krause et al., 2009). Such a sample-specific, post hoc approach does not allow for any prespecified classification criteria. Yet, a priori predictions are necessary for good theory and for translation into practice. One

may intuitively set the criteria as mastery of a correct strategy. Yet, in Fyfe et al. (2012), results indicated that the negative effects of feedback occur for learners with only moderate knowledge of correct strategies (e.g., they use correct and incorrect strategies inconsistently), suggesting the threshold may be some versus no knowledge of a correct strategy. In the current study, we provided some children with knowledge of a correct strategy and other children with no knowledge of a correct strategy.

A final issue is to better understand why feedback may have negative effects. Learners with moderate prior knowledge in the domain may be particularly susceptible to negative effects precisely because they can activate their knowledge during the task. Although generally helpful, knowledge activation may have potential consequences when processing feedback. First, it may increase learners’ expectation of performing well compared to learners with no prior knowledge, and thus heighten their sensitivity to feedback that states otherwise (Kluger & DeNisi, 1996). Second, greater knowledge activation may increase the processing of redundant information, and thus heighten the cognitive demands of the task. For example, for higher knowledge learners, feedback provides information about a problem the learner already knows to a certain degree. This redundancy may cause learners to spend cognitive resources on unnecessary information and reduce learning (Sweller, Ayres, & Kalyuga, 2011). In the current study, we included trial-by-trial microgenetic analyses and subjective student reports to better understand how feedback impacted the learning process. Also, to examine if the negative effects of feedback were robust, we varied features of the feedback provided. In particular, we varied whether the feedback included the correct answer and whether feedback was provided immediately or after a delay.

In two experiments, we worked with elementary schoolchildren learning about math equivalence. Math equivalence is the relation between two quantities that are equal and interchangeable (Kieran, 1981) and it is arguably one of the most important concepts for developing young children’s algebraic thinking (Falkner, Levi, & Carpenter, 1999; Knuth, Stephens, McNeil, & Alibali, 2006). Indeed, the Common Core State Standards recognize the importance of math equivalence and have included it in their standards. Unfortunately, children struggle to understand math equivalence and have difficulty solving math equivalence problems (i.e., problems with operations on both sides of the equal sign, such as $3 + 4 + 5 = 3 + _$; e.g., McNeil, 2008; Rittle-Johnson & Alibali, 1999; Weaver, 1973). Poor performance on these problems often stems from misinterpretations of the equal sign as an operator symbol meaning “get the answer,” as opposed to a symbol relating two equal amounts (Kieran, 1981; McNeil & Alibali, 2005). Thus, math equivalence is both educationally relevant and difficult for children to understand by themselves.

Experiment 1

The goal of Experiment 1 was to test the causal role of prior knowledge on the impact of feedback. We manipulated children’s strategy knowledge prior to problem solving as well as the provision of feedback during problem solving. We provided simple right-wrong verification feedback for several reasons. First, we were interested in studying the potentially powerful effects of seemingly minor input during problem solving. In particular, given concerns that extensive feedback might overwhelm or disrupt

ongoing cognitive processing, we opted to minimize the amount of information in the feedback message. Second, in previous work, we found that the content of the feedback did not matter. For example, whether the feedback message included the correct answer did not impact its effects (Fyfe et al., 2012). We predicted that children with no initial knowledge of a correct strategy would benefit from right–wrong verification feedback, but that, contrary to conventional wisdom, children with induced knowledge of a correct strategy would be hindered by it relative to no feedback.

Method

Participants. Initial participants were 159 children from second- and third-grade classrooms in two public schools and one private school. Of those children, 112 met criteria for participation because they could not solve any math equivalence problems correctly (out of 4) on a screening measure. This ensured that any effects due to strategy knowledge level were a result of the strategy knowledge manipulation and not preexisting differences. Data from four additional children were excluded for failing to complete all activities. The final sample contained 108 children (*M* age = 8.4 years, *min* = 7.2 years, *max* = 9.8 years; 67 girls; 41 boys).

Materials and coding. The materials consisted of a screening measure, intervention problems, a subjective task difficulty measure, and a posttest.

Screening measure. The screening measure was three tasks that tap understanding of math equivalence (from McNeil, Fyfe, Petersen, Dunwiddie, & Brletic-Shipley, 2011). For *equation solving*, children solved four math equivalence problems. Inclusion criterion was based solely on equation solving, as we were interested in children’s knowledge of solution strategies. The two remaining tasks allowed us to test if conditions were matched on different aspects of prior knowledge. For *equation encoding*, children reconstructed four math equivalence problems in writing after viewing each for 5 s to assess how they mentally represented the structure of the problem. They received one point for each accurate reconstruction (up to four points). For *defining the equal sign*, children provided a written definition of the equal sign and received one point if they provided a relational definition (e.g., “the same amount”).

Intervention problems. The 12 intervention problems (from Fyfe et al., 2012) consisted mostly of four- and five-addend math equivalence problems with operations on both sides of the equal sign, with the unknown after the equal sign (e.g., $3 + 7 = _ + 6$) or at the end (e.g., $5 + 3 + 9 = 5 + _$). Three problems had an operation on the right side only (e.g., $9 = 6 + _$).

Task difficulty. We obtained children’s subjective ratings of task difficulty as a component of their experience of cognitive load. We administered a nine-item measure that we adapted from previous measures in the cognitive load literature (Fyfe, DeCaro, & Rittle-Johnson, 2015; Hart & Staveland, 1988; Paas, Tuovinen, Tabbers, & Van Gerven, 2003) to be suitable for young children. The measure included three 3-item subscales: mental effort, mental frustration, and task difficulty. Children responded to each item by circling their answer on a 4-point scale ranging from 1 (*strongly disagree*) to 4 (*strongly agree*). The task difficulty scale was the only reliable and valid scale (see Appendix for these results), so it is the only scale we report. Several additional cognitive measures

were used in the validation process, and they are described in the Appendix.

Posttest. The posttest, adapted from past work (Matthews, Rittle-Johnson, McEldoon, & Taylor, 2012; Rittle-Johnson, Matthews, Taylor, McEldoon, 2011), was a broader measure that included procedural and conceptual knowledge scales. The procedural knowledge scale included eight items (see Table 1) that assessed children’s use of correct strategies to solve math equivalence problems ($\alpha = .90$). Half of the items were similar to those presented during the intervention (i.e., learning items) and half differed on a key problem feature, such as inclusion of subtraction (i.e., transfer items). The conceptual knowledge scale included 10 items (see Table 2 for examples) that assessed two key concepts: the relational meaning of the equal sign and the structure of equations ($\alpha = .73$).

Coding. We coded children’s problem-solving strategies. On the screening measure and posttest, strategies were coded from children’s written numerical answers. For example, for the problem $2 + 7 = 6 + _$, an answer of 15 indicated an incorrect “add all” strategy and an answer of 3 indicated a correct strategy. Responses within plus or minus one of the correct answer were coded as reflecting a correct strategy. On the intervention problems, strategies were based on children’s verbal reports (see Table 3; e.g., strategy reports). A second rater coded 30% of the responses. Interrater agreement on specific strategy use was high ($\kappa = .94$), and even higher for whether the strategy was correct or incorrect ($\kappa = .99$). We also coded the conceptual knowledge items on the screening measure and posttest that required a written explanation (e.g., definition of the equal sign). A second rater coded 30% of the responses and interrater agreement was high ($\kappa_s = .94-.98$).

Design. The study had a 2 (induced strategy knowledge: yes vs. no) \times 2 (feedback: present vs. absent) between-subjects design with children randomly assigned to conditions: strategy knowledge with feedback ($n = 27$), strategy knowledge without feedback ($n = 26$), no knowledge with feedback ($n = 27$), and no knowledge without feedback ($n = 28$). There were no significant differences between conditions in terms of age, gender, or grade ($ps > .5$).

Procedure. Children completed the screening measure in their classrooms in a 10-min session. Those who met the inclusion criteria then completed a one-on-one tutoring intervention in a single session lasting approximately 50 min. This session was conducted in a quiet area at the school with Emily R. Fyfe. The one-on-one session included four components: knowledge manipulation, knowledge check, problem solving, and an immediate posttest.

Knowledge manipulation. Children assigned to the strategy-knowledge condition received instruction on a correct problem-

Table 1
Problems Presented on the Procedural Knowledge Scale at Posttest

Learning items	Transfer items
$8 = 6 + _$	$_ + 2 = 6 + 4$
$3 + 4 = _ + 5$	$8 + _ = 8 + 6 + 4$
$3 + 7 + 6 = _ + 6$	$5 + 6 - 3 = 5 + _$
$7 + 6 + 4 = 7 + _$	$5 - 2 + 4 = _ + 4$

Table 2
Example Problems Presented on the Conceptual Knowledge Scale at Posttest

Equal sign items	Equation structure items
What does the equal sign (=) mean?	Reproduce $4 + 3 + 9 = 4 + \underline{\quad}$ from memory after viewing for 5 s.
Is “two amounts are the same” a good definition of the equal sign?	Decide if $3 = 3$ and $7 = 3 + 4$ are true or false.
What goes in the box to show that 10 cents is the same amount of money as 1 dime?	Decide if $6 + 4 = 5 + 5$ is true or false and explain how you know.

solving strategy with four math equivalence problems presented on a computer one at a time. Two problems had the blank immediately following the equal sign (e.g., $5 + 4 + 3 = \underline{\quad} + 5$) and two had the blank at the end (e.g., $3 + 4 + 2 = 3 + \underline{\quad}$). We used two problem types to increase the generalizability of the strategy (Matthews & Rittle-Johnson, 2009). Children were instructed on the commonly used equalize strategy, which involves adding the numbers on one side of the equal sign and then counting up from the number on the other side to get the same amount. We taught the equalize strategy as it is the strategy children tend to generate first on their own (e.g., Alibali, Phillips, & Fischer, 2009). The experimenter provided instruction and demonstrated the procedure (adapted from Alibali et al., 2009) on all four problems. Children were asked to answer questions for each problem (e.g., “if we add up this side, what do we get?”) to ensure they were attending to instruction.

Children assigned to the no-knowledge condition received instruction on a filler task (adapted from Hattikudur & Alibali, 2010) to control for time on task and practice with addition. Children were directed to look at two boxes on the computer screen. One contained a single digit (e.g., 9) and the other contained a pair of addends (e.g., $3 + 4$). Children were taught to decide which box has the biggest total number. The experimenter demonstrated the procedure on all four problems. Children were asked to add the pair of addends for each problem and answer questions (e.g., “what is the next step?”) to ensure they were attending to the task.

Knowledge check. To ensure the knowledge manipulation worked, all children then solved a math equivalence problem on their own (i.e., $7 + 6 + 2 = 7 + \underline{\quad}$) and reported how they solved the problem. In the no-knowledge condition, children did not receive any feedback and simply moved on to the next task. In the strategy-knowledge condition, children were told whether or not

they solved the problem correctly. If they used an incorrect strategy, instruction on the equalize strategy was repeated and they were asked to solve another problem until they solved one using a correct strategy and received feedback that it was correct. This procedure ensured that children in the strategy-knowledge condition had knowledge of a correct strategy and were aware that they used it correctly. We set the protocol such that after five failed attempts the experiment was discontinued and children received more remedial tutoring.

Problem solving. Children were then asked to solve 12 math problems presented one at a time on a computer (see the Materials section for a description of the problems). After each problem, children reported how they solved the problem and either did or did not receive feedback.

In the no-feedback condition, children did not receive feedback and were told to go to the next problem. In the feedback condition, children received right/wrong verification feedback on their answer to the problem (e.g., “Good job! You got the right answer.”/ “Good try, but you did not get the right answer.”). The feedback was based solely on the correctness of the child’s numerical answer and did not depend on the strategy reported. On a few occasions, children reported using a correct strategy, but obtained an incorrect answer due to an arithmetic mistake and received negative feedback. However, these mismatches were rare (6% of all trials) and exclusion of children who experienced two or more mismatches ($n = 12$) did not impact the results. Feedback was presented verbally by the experimenter and visually on the computer.

Following the intervention, children rated their task difficulty, took a brief break (3–5 min), and then completed the posttest. We intended for all children to complete a 2-week retention test. However, data collection issues related to absences and school

Table 3
Strategies Children Used to Solve the Intervention Problems

Strategy	For the problem $4 + 5 + 3 = 4 + \underline{\quad}$	
	Solution	Example verbal report
Correct strategies		
Equalize	8	“4 plus 5 plus 3 is 12 and 4 plus 8 is 12.”
Add-subtract	8	“I added 4 plus 5 plus 3 and took away 4 from that.”
Grouping	8	“I saw the 4 and the 4 and I just added the 5 and the 3.”
Incorrect strategies		
Add-all	16	“I just added them all up.”
Add-to-equal	12	“I added 4 plus 5 plus 3 and that equals 12.”
Add-two	9	“I added 4 plus 5 and that is 9.”
Carry	5	“I saw 4 plus 5 here so I made it 4 plus 5 over here.”

breaks prevented us from obtaining a reliable sample to analyze. Indeed, only 58% of the sample completed a 2-week retention test (plus or minus 2 days), so these data are not reported.

Data analysis. To examine children's performance on the primary outcome measures, we performed a series of analyses of covariances (ANCOVAs) with strategy knowledge (yes vs. no) and feedback (present vs. absent) as between-subjects variables. We included children's age and their scores on the screening measure as covariates. Preliminary analyses revealed no interactions with age or screening measure scores, so these interaction terms were not retained in the final models. Partial eta squared was adopted as the measure of effect size. According to Cohen (1988), values of .01, .06, and .14 can be interpreted respectively as small, medium, and large effects.

Results

Screening measure. Because of our inclusion criteria, all children in the final sample solved zero problems correctly on the equation-solving task on the screening measure. Some children succeeded on the other two tasks. On average, children in the final sample encoded 1.6 ($SD = 1.3$) math equivalence problems correctly (out of 4), and 11% of children provided a relational definition of the equal sign. Performance on the equation-encoding task and on the equal sign definition task did not differ as a function of condition, $ps > .3$. In line with previous research (Chesney et al., 2014; McNeil et al., 2011), we created a composite measure of children's performance across these two tasks by summing z scores from each task. Composite scores ranged from -1.64 to 4.69 ($M = 0.00$, $SD = 1.39$) and did not differ as a function of condition, $p = .97$. Composite scores served as a covariate in subsequent analyses, though the pattern of findings was the same if it was not included.

Knowledge check. The knowledge manipulation was successful. All children in the strategy-knowledge condition exhibited knowledge of a correct strategy after the manipulation; 88% of them used a correct strategy to solve the first problem, 6% needed a second problem, and 6% needed a third attempt. Nine (16%) children in the no-knowledge group solved the problem correctly, despite receiving no instruction on a correct strategy. Because these children could not be considered to have no knowledge of a correct strategy, they were excluded from analyses, leaving the no-knowledge group with 24 children in the feedback condition and 22 in the no-feedback condition. The pattern of findings was the same if these nine children were included.

Intervention measures. For analyses of children's verbal strategy reports, we focused on the nine math equivalence problems with operations on both sides of the equal sign. These problems elicit more easily identified strategies than problems with operations on only one side of the equal sign. However, the pattern of results was similar if we considered all 12 problems.

Correct strategy use. We examined the percentage of trials on which children reported using a correct strategy. See Table 3 for example strategy reports. There was a significant feedback by knowledge interaction, $F(1, 93) = 11.13$, $p = .001$, $\eta_p^2 = .11$. For the strategy-knowledge group, children who received feedback used a correct strategy less often ($M = 77\%$, $SE = 5\%$) than children who did not ($M = 90\%$, $SE = 5\%$), $F(1, 93) = 3.51$, $p = .06$, $\eta_p^2 = .04$, though this difference was only marginally signif-

icant. In contrast, for the no-knowledge group, children who received feedback used a correct strategy significantly more often ($M = 30\%$, $SE = 5\%$) than children who did not ($M = 9\%$, $SE = 5\%$), $F(1, 93) = 7.87$, $p = .01$, $\eta_p^2 = .08$. There was no overall main effect of feedback, $p = .45$, but there was a main effect of knowledge, $F(1, 93) = 156.24$, $p < .001$, $\eta_p^2 = .63$, with the strategy-knowledge group ($M = 84\%$, $SE = 4\%$) outperforming the no-knowledge group ($M = 20\%$, $SE = 4\%$). A trial-by-trial examination of children's reported strategy use supports these conclusions (see Figure 1).¹

Correct strategy generation. We also examined the types of strategies children reported using (see Table 3). For each child, we calculated the number of different types of correct strategies he or she used. As long as children reported the strategy on at least one problem, they received credit for using it. The number of different correct strategies used during the intervention ranged from zero to three ($M = 1.1$, $SD = 0.8$). There was a significant feedback by knowledge interaction, $F(1, 93) = 6.30$, $p = .01$, $\eta_p^2 = .06$. For the strategy-knowledge group, children generated a similar number of correct strategies whether they received feedback ($M = 1.3$, $SE = 0.1$) or not ($M = 1.4$, $SE = 0.1$), $p = .42$. For the no-knowledge group, children who received feedback generated a greater number of correct strategies ($M = 1.0$, $SE = 0.1$) than children who did not ($M = 0.4$, $SE = 0.1$), $F(1, 93) = 7.13$, $p = .01$, $\eta_p^2 = .07$. There was no overall main effect of feedback, $p = .17$. There was a main effect of knowledge, $F(1, 93) = 23.65$, $p < .001$, $\eta_p^2 = .20$, with the strategy-knowledge group using a greater number of correct strategies ($M = 1.4$, $SE = 0.1$) than the no-knowledge group ($M = 0.7$, $SE = 0.1$).

Incorrect strategy generation. For each child, we also calculated the number of different types of incorrect strategies he or she used. The number of different incorrect strategies ranged from zero to four ($M = 1.0$, $SD = 0.9$; see Table 2). There was no feedback by knowledge interaction, $p = .76$. However, there were main effects of feedback, $F(1, 93) = 4.43$, $p = .04$, $\eta_p^2 = .05$, and knowledge, $F(1, 93) = 64.24$, $p < .001$, $\eta_p^2 = .41$. Children who received feedback generated a greater number of incorrect strategies ($M = 1.2$, $SE = 0.1$) than children who did not ($M = 0.9$, $SE = 0.1$). Children in the no-knowledge group generated a greater number of incorrect strategies ($M = 1.6$, $SE = 0.1$) than children in the strategy-knowledge group ($M = 0.4$, $SE = 0.1$). There also were differences in perseveration—reporting the same incorrect strategy on all nine math equivalence problems. Nine children

¹ The trial-by-trial data revealed an unexpected difference within the strategy-knowledge group between children in the no-feedback and feedback conditions on the very first problem. In particular, within the strategy-knowledge group, more children in the no-feedback condition solved the problem correctly than children in the feedback condition. However, several points suggest this initial difference was not a concern. First, the first intervention problem was an easier problem with an operation on the right side of the equal sign only. We were primarily interested in children's knowledge of math equivalence problems with operations on both sides of the equal sign. Second, this initial difference was not reliable. Focusing just on the strategy-knowledge group, we performed a logistic regression to predict performance on this initial item and included feedback condition as a factor. There was no significant effect of feedback, $\beta = -.68$, $z = 0.80$, $p = .42$. Finally, all subsequent results (e.g., significance, effect size) remained largely unchanged when we controlled for performance on this initial intervention item. Thus, any initial differences on the first intervention problem cannot explain differences on the posttest.

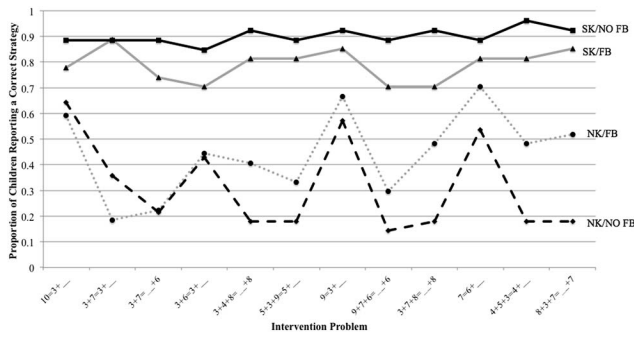


Figure 1. Proportion of children reporting a correct strategy on each problem (Experiment 1). The first, seventh, and 10th problems were easier items with an operation on the right side of the equal sign only. The nine remaining problems were math equivalence problems with operations on both sides of the equal sign. SK = strategy-knowledge condition; NK = no-knowledge condition; FB = feedback condition; NO FB = no-feedback condition.

(41%) in the no-knowledge without feedback condition perseverated, whereas none of the children in the three remaining conditions did.

Positive versus negative feedback. The majority of children in the feedback conditions received a mix of both positive (i.e., that's correct) and negative (i.e., that's incorrect) feedback. For the strategy-knowledge group, a small portion ($n = 6$) received positive feedback on all 12 trials, but no children received negative feedback on all 12 trials. On the other hand, in the no-knowledge group, a small portion ($n = 5$) received negative feedback on all 12 trials, but only one child received positive feedback on all 12 trials.

Task difficulty. We also analyzed children's subjective ratings of task difficulty as an indicator of their cognitive load. There was a significant feedback by knowledge interaction, $F(1, 93) = 4.81$, $p = .03$, $\eta_p^2 = .05$. For the strategy-knowledge group, children reported similar levels of task difficulty whether they received feedback ($M = 2.1$ out of 4, $SE = 0.2$) or not ($M = 2.3$, $SE = 0.2$), $p = .48$. For the no-knowledge group, children who received feedback reported higher levels of task difficulty ($M = 2.7$, $SE = 0.2$) than children who did not ($M = 2.1$, $SE = 0.2$), $F(1, 93) = 5.46$, $p = .02$, $\eta_p^2 = .06$. There were no overall main effects of feedback, $p = .23$, or knowledge group, $p = .35$.

Intervention summary. During the intervention, right-wrong verification feedback had positive effects for low-knowledge children. In particular, for children in the no-knowledge condition, feedback increased the frequency of correct strategy use, prevented perseveration on the same incorrect strategy, facilitated the generation of more diverse strategies (both correct and incorrect), and also led to increased ratings of task difficulty relative to no feedback. In contrast, right/wrong verification feedback had neutral or negative effects for higher knowledge children. In particular, for children in the strategy-knowledge condition, feedback had no significant effect on the frequency of correct strategy use, it facilitated the generation of more incorrect strategies, and it led to similar ratings of task difficulty relative to no feedback.

Posttest. To evaluate children's performance on the posttest, we conducted two separate ANCOVAs for procedural knowledge and conceptual knowledge.

Procedural knowledge. We examined children's percentage correct on the procedural knowledge scale (across learning and transfer items; see Figure 2). As expected, there was a large, significant feedback by knowledge interaction, $F(1, 93) = 16.80$, $p < .001$, $\eta_p^2 = .15$. For the strategy-knowledge group, children who received feedback exhibited significantly lower procedural knowledge ($M = 62\%$, $SE = 6\%$) than children who did not ($M = 81\%$, $SE = 6\%$), $F(1, 93) = 4.96$, $p = .03$, $\eta_p^2 = .05$. For the no-knowledge group, children who received feedback exhibited significantly higher procedural knowledge ($M = 50\%$, $SE = 6\%$) than children who did not ($M = 18\%$, $SE = 7\%$), $F(1, 93) = 12.38$, $p < .001$, $\eta_p^2 = .12$. There was no overall main effect of feedback, $p = .30$. There was a main effect of knowledge, $F(1, 93) = 36.76$, $p < .001$, $\eta_p^2 = .28$, as the strategy-knowledge group as a whole exhibited higher procedural knowledge ($M = 72\%$, $SE = 4\%$) than the no-knowledge group ($M = 34\%$, $SE = 5\%$). However, the effects of feedback were so positive for the no-knowledge group and so negative for the strategy-knowledge group that there were no statistical differences in the feedback condition between no-knowledge children ($M = 50\%$, $SE = 6\%$) and strategy-knowledge children ($M = 62\%$, $SE = 6\%$), $p = .16$.

Exploratory analyses revealed that the results remain unchanged after excluding children who received only positive feedback or children who received only negative feedback. Further, the effects were similar on learning and on transfer problems. Overall, children with no knowledge of a correct strategy benefited from right-wrong verification feedback relative to no feedback, but, for children with induced knowledge of a correct strategy, the reverse was true.

Conceptual knowledge. We examined children's percentage correct on the conceptual knowledge scale (see Figure 3). There was no feedback by knowledge interaction, $F(1, 93) = 0.60$, $p = .44$, $\eta_p^2 = .01$. There also was no significant effects of feedback, $p = .50$, but there was a marginal effect of knowledge group, $F(1, 93) = 2.86$, $p = .09$, $\eta_p^2 = .03$. Children in the strategy-knowledge condition exhibited somewhat higher conceptual knowledge ($M = 43\%$, $SE = 3\%$) than children in the no-knowledge condition ($M = 35\%$, $SE = 3\%$), but not reliably so.

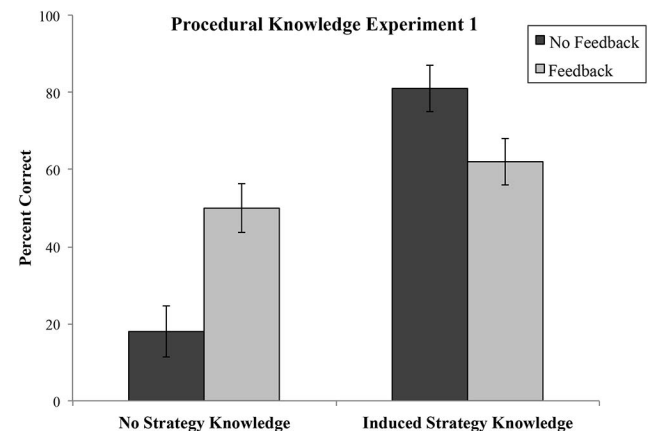


Figure 2. Procedural knowledge at posttest by condition (Experiment 1). Scores are estimated marginal means. Error bars represent standard errors.

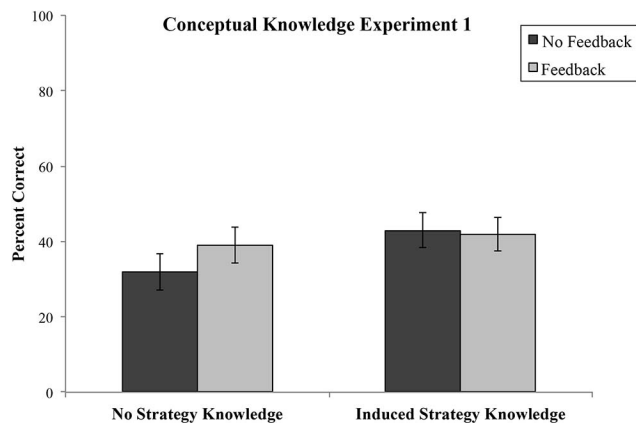


Figure 3. Conceptual knowledge at posttest by condition (Experiment 1). Scores are estimated marginal means. Error bars represent standard errors.

Discussion

Experiment 1 is the first study to provide causal evidence for the moderating role of prior knowledge on the effects of right–wrong verification feedback. Children with no knowledge of a correct problem-solving strategy benefited from verification feedback relative to no feedback. In contrast, children who were taught a correct strategy learned more if they did not receive verification feedback. The reversal occurred on children’s procedural knowledge at posttest. Further, intervention results shed light on these effects. For example, for the no-knowledge group, verification feedback reduced perseveration on incorrect strategies and facilitated the generation of correct strategies. In contrast, for children who already knew a correct strategy, verification feedback merely served to facilitate the generation of incorrect strategies.

An important next step is to consider why feedback has negative effects for learners with prior knowledge. One possibility is that the activation of their prior knowledge causes some interference when processing the feedback. For example, their prior knowledge may increase their expectation of performing well. Given feedback, they may become fixated on whether they were right or wrong (and how that reflects on their abilities), rather than on ways to improve on subsequent problems (Kluger & DeNisi, 1996). Further, higher knowledge learners often rely on their existing knowledge to guide task performance and to generate internal feedback (Butler & Winne, 1995). External feedback may provide redundant information that competes for working memory resources and ultimately impairs knowledge acquisition (Sweller et al., 2011).

In Experiment 1, the immediacy of the feedback may have heightened these potential consequences. Immediate feedback is provided right after a learner has responded to a problem and has the chance to impact ongoing task processing. For example, if feedback is provided right after the first problem, then any resulting affective or cognitive interference will likely hinder performance on the second problem. One potential solution is to delay feedback rather than provide it on a trial-by-trial basis. For example, summative feedback is provided after the learner has responded to all problems in a set. Although summative feedback may still produce cognitive or affective responses, it is not pro-

vided during problem solving when task-relevant processing is ongoing. More important, several studies have found benefits of summative feedback relative to immediate feedback (Butler, Karpicke, Roediger, 2007; Clariana, Wagner, & Roher Murphy, 2000). Further, several researchers suggest that delaying feedback may be particularly beneficial for learners with higher knowledge in the target domain relative to lower knowledge learners (Mason & Bruning, 2001; Shute, 2008), though this has never been experimentally tested.

To address this possibility, a second experiment was conducted with key modifications. First, we provided strategy instruction to all participating children to induce some strategy knowledge in all learners. Second, we manipulated the presence and timing of feedback by including no-feedback, immediate-feedback, and summative-feedback conditions. Third, we employed correct-answer feedback rather than right–wrong verification feedback to ensure the negative effects of feedback were not specific to verification feedback. Indeed, comprehensive reviews indicate that feedback that provides verification and the correct answer is often more beneficial than verification feedback alone (Kluger & DeNisi, 1996).

Experiment 2

The goal of Experiment 2 was to provide additional insight into the negative effects of feedback for learners with some prior knowledge. In line with Experiment 1, we predicted that immediate feedback would have negative effects relative to no feedback. However, we predicted that summative feedback would have neutral or even positive effects relative to no feedback.

Method

Participants. Initial participants were 131 children from second- and third-grade classrooms in three public schools. Of those children, 113 met criteria for participation because they scored below 80% on an equation-solving screening measure. This criteria was adopted from our previous feedback study (Fyfe et al., 2012) and ensured that children had room to learn from the intervention. We used a more lenient inclusion criteria relative to Experiment 1 because all children in this study were given instruction on a correct strategy so it was not necessary that they started at the same initial knowledge level. Data from 12 additional children were excluded for failing to complete all activities. The final sample contained 101 children (M age = 8.2 years, min = 7.0 years, max = 9.8 years; 57 girls, 44 boys).

Materials. The materials were identical to those in Experiment 1 with three exceptions. First, the screening measure included an additional, simpler equation-solving problem (i.e., $7 = _ + 3$) to increase the variability in equation-solving scores. Second, to assess children’s cognitive load we only administered the validated three-item task difficulty scale. Third, to explore a potential reason for the negative effects of feedback, we also measured children’s self-assessment (i.e., whether they considered their performance to reflect negatively on their traits and abilities) using a four-item measure from Kamins and Dweck (1999). Children were asked whether the task made them feel like they were good or not good at solving the problems, a good or a not good student, a nice or a not nice student, and a smart or a not smart

student. Children received one point each time they chose the positive attribute and scores were summed to form an index ranging from 0 to 4. Internal consistency was sufficient ($\alpha = .78$), but there was a restriction in the range of the responses, as 82% of children always selected the positive attribute.

Design. The study had a between-subjects design with children randomly assigned to conditions: no feedback ($n = 33$), immediate feedback ($n = 35$), and summative feedback ($n = 33$). There were no differences between conditions in terms of age, gender, or grade ($ps > .45$).

Procedure. The procedure was identical to Experiment 1 with a few exceptions. First, all children received instruction on a correct problem-solving strategy. Second, for the knowledge check problem, children were not told whether they solved the problem correctly. This ensured that children in the no-feedback condition never received feedback. If they solved the problem correctly, they were told to move on to the next activity. If they solved it incorrectly, general instruction on the equalize strategy was repeated without revealing the correct answer, and they were asked to solve another problem until they solved one correctly. Third, we manipulated both the presence and timing of feedback. The no-feedback condition was identical to Experiment 1. In the immediate-feedback condition, children received trial-by-trial correct-answer feedback, which included right-wrong verification (as in Experiment 1) and also the correct answer. In the summative-feedback condition, children received verification and correct-answer feedback after all 12 problems had been solved. The problems with the child's solutions reappeared on the computer screen, and the experimenter provided correct-answer feedback for each problem.

Data analysis. Two children were missing data on the equation-encoding section of the screening measure and two children failed to provide their date of birth and were missing values for their age. Imputing missing independent variables leads to more precise and unbiased conclusions than omitting participants with missing data (Peugh & Enders, 2004). We used the expectation-maximization algorithm for maximum likelihood estimation via the missing values analysis in SPSS (Schafer & Graham, 2002) to impute the missing encoding scores and ages.

To examine children's performance on the primary outcome measures, we performed a series of ANCOVAs with condition as a between-subjects variable. In particular, condition was dummy coded with immediate feedback and summative feedback entered into the models, and no feedback as the reference group. On several measures, scores were not normally distributed. In those cases, we used binomial logistic regression. Again, condition was dummy coded. In all models, we included children's age and their scores on the screening measure as covariates. Preliminary analyses revealed no interactions with age or screening measure scores so these interaction terms were not retained in the final models. For ANCOVAs, we report partial eta squared as the measure of effect size. For logistic regression, we report odds ratios.

Results

Screening measure. On average, children in the final sample solved 1.0 ($SD = 1.0$) problem correctly (out of 5), encoded 1.7

($SD = 1.2$) problems correctly (out of 4), and only 6% of children provided a relational definition of the equal sign. Performance on the three tasks did not differ as a function of condition, $ps > .55$. We created a composite measure of children's performance by summing z scores across the three tasks. Scores ranged from -2.62 to 6.95 ($M = 0.00$, $SD = 1.97$) and did not differ as a function of condition, $p = .91$.

Knowledge check. The knowledge induction was largely successful. Most children (85%) exhibited knowledge of a correct strategy on the first problem following instruction, 7% needed a second problem, and 2% were successful by the fifth attempt. The remaining 6% of children never used a correct strategy after five attempts with repeated instruction after each problem. Difficulties were often due to poor arithmetic fact knowledge and weak counting skills. For these children, the experiment was stopped and remedial tutoring was provided, as they were clearly not ready to learn about or solve these problems. This resulted in a sample of 95 children (no feedback, $n = 32$; immediate feedback, $n = 32$; summative feedback, $n = 31$).

Intervention measures. As in Experiment 1, for analyses of children's verbal strategy reports we focused on the nine math equivalence problems with operations on both sides of the equal sign. However, the pattern of results was similar if we considered all 12 problems.

Correct strategy use. The frequency of correct strategy use during the intervention was similar for children in the no-feedback ($M = 88\%$, $SE = 4\%$), immediate-feedback ($M = 85\%$, $SE = 4\%$), and summative-feedback conditions ($M = 82\%$, $SE = 4\%$). As in Experiment 1, there was no significant effect of immediate feedback relative to no feedback, $p = .53$. There also was no effect of summative feedback relative to no feedback, $p = .24$. A follow-up analysis revealed no significant difference between the two feedback types, $p = .58$.

Correct strategy generation. The number of different types of correct strategies used was also similar for children in the no-feedback ($M = 1.4$, $SE = 0.1$), immediate-feedback ($M = 1.3$, $SE = 0.1$), and summative-feedback conditions ($M = 1.2$, $SE = 0.1$). As in Experiment 1, there was no significant effect of immediate feedback relative to no feedback, $p = .78$. There was also no effect of summative feedback relative to no feedback, $p = .12$. A follow-up analysis revealed no significant difference between the two feedback types, $p = .20$.

Incorrect strategy generation. The number of different types of incorrect strategies used was highest in the immediate-feedback condition ($M = 1.0$, $SE = 0.2$), next highest in the summative-feedback condition ($M = 0.8$, $SE = 0.2$), and lowest in the no-feedback condition ($M = 0.4$, $SE = 0.2$). As in Experiment 1, there was a significant effect of immediate feedback relative to no feedback, $F(1, 90) = 5.31$, $p = .02$, $\eta_p^2 = .06$. There was a marginal effect of summative feedback relative to no feedback, $F(1, 90) = 3.64$, $p = .06$, $\eta_p^2 = .04$. A follow-up analysis revealed no significant difference between the two feedback types, $p = .69$. Only two children perseverated and used the same incorrect strategy across all nine math equivalence problems. They were both in the no-feedback condition.

Positive versus negative feedback. The majority of children in the feedback conditions received a mix of positive (i.e., that's correct) and negative (i.e., that's incorrect) feedback. A small portion ($n = 8$ in immediate feedback, $n = 2$ in summative

feedback) received positive feedback on all trials, but no children received negative feedback on all trials.

Task difficulty. Children’s ratings of task difficulty were similar across conditions, but somewhat higher in the summative-feedback condition ($M = 2.3, SE = 0.1$) than the immediate- ($M = 2.1, SE = 0.1$) and no-feedback ($M = 2.0, SE = 0.1$) conditions. As in Experiment 1, there was no effect of immediate feedback relative to no feedback, $p = .38$. There was a marginal effect of summative feedback relative to no feedback, $F(1, 89) = 2.98, p = .09, \eta_p^2 = .03$. A follow-up analysis revealed no significant difference between the two feedback types, $p = .40$.

Self-assessment. The percentage of children scoring a 4 out of 4 (i.e., reporting consistently positive self-assessment) was high in the no-feedback (87%), immediate-feedback (78%), and summative-feedback conditions (84%). A logistic regression revealed that there were no significant effects of immediate feedback relative to no feedback, $p = .23$, or summative feedback relative to no feedback, $p = .95$. A follow-up analysis revealed no significant difference between the two feedback types, $p = .24$.

Intervention summary. As in Experiment 1, feedback primarily had neutral effects for children with knowledge of a correct strategy during the intervention. In particular, immediate correct-answer feedback did not impact the frequency of correct strategy use, it facilitated the generation of more incorrect strategies, and it led to similar ratings of task difficulty relative to no feedback. As expected, the summative-feedback condition was not reliably different from the no-feedback condition on any measure during the intervention.

Posttest. To evaluate children’s performance on the posttest, we conducted two separate analyses for procedural knowledge and conceptual knowledge.

Procedural knowledge. Children’s percentage correct on the procedural knowledge scale at posttest was highest in the no-feedback condition ($M = 83%, SE = 5%$), somewhat lower in the immediate-feedback condition ($M = 78%, SE = 5%$), and even lower in the summative-feedback condition ($M = 71%, SE = 5%$). However, procedural knowledge scores were not normally distributed. Across conditions, 40% of children solved all eight of the items correctly. Thus, we used binomial logistic regression to predict the log of the odds of scoring 100%. The results are displayed in Figure 4. Consistent with Experiment 1, there was a significant, negative effect of immediate feedback, $\hat{\beta} = -1.16, z = 2.17, Wald(1, N = 95) = 4.70, p = .03, OR = 0.31$. There was also a significant, negative effect of summative feedback, $\hat{\beta} = -2.07, z = 3.48, Wald(1, N = 95) = 12.10, p = .001, OR = 0.13$. Children in the immediate- and summative-feedback condition were less likely than children in the no-feedback condition to score 100% on the posttest. A follow-up analysis revealed no significant difference between the two feedback types, $p = .13$.

Conceptual knowledge. Findings were similar for conceptual knowledge. Children’s percentage correct on the conceptual knowledge scale at posttest was highest in the no-feedback condition ($M = 53%, SE = 4%$), and lower in the immediate-feedback ($M = 41%, SE = 4%$) and summative-feedback ($M = 40%, SE = 4%$) conditions (see Figure 5). There were significant negative effects of immediate feedback, $F(1, 84) = 4.80, p = .03, \eta_p^2 = .05$, and summative feedback, $F(1, 84) = 5.03, p = .03, \eta_p^2 = .06$,



Figure 4. Procedural knowledge at posttest by condition (Experiment 2).

relative to no feedback. A follow-up analysis revealed no significant difference between the two feedback types, $p = .92$.

Discussion

Experiment 2 was largely consistent with Experiment 1 and supported our first hypothesis. Children with induced knowledge of a correct strategy benefitted more from no feedback than from immediate correct-answer feedback. The negative effect of immediate correct-answer feedback occurred on procedural and conceptual knowledge at posttest. Further, the intervention results mirrored those from Experiment 1 and provided further insight into the effects of feedback during learning. In particular, immediate correct-answer feedback facilitated the generation of incorrect strategies relative to the no-feedback condition, but had no impact on correct strategy generation. Feedback did not impact reported task difficulty or self-assessment, failing to support the potential role of these factors in explaining the negative effects of feedback.

In contrast to our second hypothesis, the effects of summative feedback also were negative. In particular, children who received summative correct-answer feedback exhibited lower procedural and conceptual knowledge than children who received no feedback. Indeed, the two feedback types (immediate and summative) did not differ significantly from one another on the posttest. One possibility is that both types of feedback trigger cognitive or affective responses that interfere with learning. Thus, whether these responses are elicited may be more important than when they are elicited (e.g., during or after problem solving).

General Discussion

The current study is the first to provide causal evidence that differences in prior knowledge can lead to varying effects of feedback during mathematics problem solving. In Experiment 1, for children with no knowledge of a correct strategy, immediate verification feedback led to higher procedural knowledge than no feedback. In contrast, children with induced knowledge of a correct strategy exhibited higher procedural knowledge if they did not receive feedback. In Experiment 2, children with induced knowl-

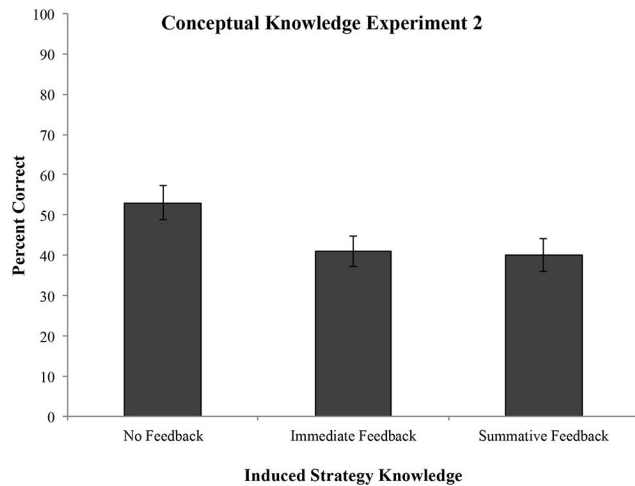


Figure 5. Conceptual knowledge at posttest by condition (Experiment 2). Scores are estimated marginal means. Error bars represent standard errors.

edge of a correct strategy exhibited higher procedural and conceptual knowledge if they did not receive correct-answer feedback (whether it was immediate or summative). The results confirm that feedback can have negative, not just neutral, effects for learners with some prior domain knowledge.

The Role of Prior Knowledge

These findings contribute to the feedback literature in several ways. For example, they address a call to explore the impact of feedback in relation to individual differences generally (Hattie & Gan, 2011) and prior knowledge specifically (Mason & Bruning, 2001; Shute, 2008). Here, we used a prefamiliarization technique, in which some children were exposed to a correct strategy and others were not (Petersen & McNeil, 2013; Rey & Buchwald, 2011). This avoids confounding variables, allows for random assignment, and establishes a causal relation. We also examined a specific type of prior knowledge, knowledge of domain-specific solution strategies, and specified the level of knowledge at which the moderation occurred. We found that only true novices benefitted from feedback during problem solving. Children with moderate knowledge (i.e., used correct and incorrect strategies inconsistently) benefitted from no feedback.

This finding is consistent with research demonstrating that one instructional method is often not best for all learners (Cronbach & Snow, 1977), and it highlights the need to consider individual differences. Indeed, instructional interventions “are likely to be different for different participants. If individual differences are not examined, average treatment differences may mask and thus miss important special effects” (Snow, 1996, p. 545). For example, expertise reversal effects occur when instructional techniques that are effective for novices lose their benefits for more experienced learners (Kalyuga, 2007). Further, the reversal is often related to levels of instructional guidance, such that low-knowledge learners benefit from strong guidance and support, but higher knowledge learners benefit from little to no guidance and support (Kalyuga, Ayres, Chandler, & Sweller, 2003). The current results provide another example of this phenomenon. In particular, low-

knowledge learners benefitted from the provision of verification feedback, but learners with higher prior knowledge actually learned more when no feedback was provided.

The Positive Effects of Feedback

A key result of this work is the strong, positive effect of verification feedback for the no-knowledge group. Indeed, this result is consistent with previous studies demonstrating substantial benefits on learning and development that can occur from the provision of minimal feedback (e.g., Bohlmann & Fenson, 2005; Brainerd, 1972). In Experiment 1, simply telling the no-knowledge children that their answers were right or wrong allowed them to go from solving zero problems correctly at pretest to solving half of the problems correctly on the posttest, both a significant and meaningful increase. Further, the effects of verification feedback were so positive for the no-knowledge group that there were no statistical differences between no-knowledge children who received feedback and strategy-knowledge children who received feedback.

The intervention results help explain why feedback had positive effects for these novices and also inform several theorized functions of feedback. First, verification feedback prevented no-knowledge children from perseverating on the same incorrect strategy, as has been found in previous research (e.g., Bohlmann & Fenson, 2005; Fyfe et al., 2012). That is, feedback encouraged children to entertain alternative approaches to the problems and may have reduced mindlessness, which is being committed to a “single, rigid perspective and . . . oblivious to alternative ways of knowing” (Langer, 2000, p. 220). Second, verification feedback facilitated children’s generation of at least one correct strategy. Indeed, discovering new problem-solving procedures is a key source of cognitive change and can be a strong predictor of subsequent performance (e.g., Rittle-Johnson, 2006; Siegler & Shipley, 1995). However, the findings suggest that verification feedback may not play these positive roles once children know a correct strategy.

The Negative Effects of Feedback

A second key result of this work is the negative effects of verification and correct-answer feedback for the strategy-knowledge group. In particular, problem solving alone was more effective for learners with some prior knowledge than problem solving with feedback—even minimally intrusive verification feedback. More important, these learners were not experts in the domain. Thus, the differences cannot be explained by more general expert–novice differences. Rather these learners were briefly exposed to one correct strategy, and most children continued to use a mix of both correct and incorrect strategies during problem solving. In the following, we outline potential mechanisms underlying the negative effects for the strategy-knowledge group.

First, we have direct evidence that immediate verification and correct-answer feedback increased the use of various incorrect strategies relative to no feedback, which is consistent with prior work (Fyfe et al., 2012). For the no-knowledge group, this negative effect was offset by a positive result, namely the introduction of a correct strategy into their repertoire. For the strategy-knowledge group, the use and strengthening of different incorrect

strategies likely had detrimental effects. Incorrect strategies compete with existing strategies and can reduce the frequency of correct strategy use (Siegler & Shipley, 1995). Thus, on the posttest, these children could select from a number of different strategies, including the correct one they knew, but also a number of different incorrect ones. Further, previous research has suggested that, in some cases, frequent shifts in strategy use are negatively related to learning (e.g., Coyle & Bjorklund, 1997; McGilly & Siegler, 1989). Thus, the strengthening of different incorrect strategies may help explain why the strategy-knowledge children were hindered by immediate feedback.

However, the results from Experiment 2 suggest that there are other mechanisms at work, too. Summative feedback was provided after problem solving and had no reliable impact on strategy generation relative to no feedback; yet, it still resulted in lower learning. Thus, there must be other aspects of feedback that negatively impact learners with some prior knowledge.

One likely possibility is that feedback results in the processing of redundant information for learners with higher prior knowledge and overloads their cognitive resources. Monitoring and evaluating feedback takes place in working memory, a short-term system that enables individuals to control, maintain, and regulate a limited amount of task-relevant information (Miyake & Shah, 1999). When there are high demands on working memory, the system can overload and hinder learning (Sweller, van Merriënboer, & Paas, 1998). Some forms of instructional guidance are thought to cause cognitive overload for higher knowledge learners because the information provided is redundant with their existing knowledge. The redundant information is still likely to be processed in working memory, which takes away from the resources that could be devoted to more germane tasks (Sweller et al., 2011). Indeed, the redundancy principle is often used to explain expertise reversal effects (see Kalyuga, 2007). The idea is that low-knowledge learners need instructional guidance to make progress, but once learners gain sufficient knowledge, the guidance becomes redundant and burdensome to process.

In the current study, it seems safe to assume that the feedback was not redundant for the no-knowledge group. However, for the strategy-knowledge group, the verification and correct-answer feedback messages provided some already-known information. For example, these children solved many problems correctly using the instructed strategy. On these trials, it seems likely that the feedback was redundant and detracting from task-relevant processing. Further, the information would be redundant regardless of whether it was provided after each problem or after the whole problem set. Although children's ratings of task difficulty do not provide evidence for this interpretation, cognitive load is a broad construct that can be measured in multiple ways (see Paas et al., 2003), and appears more difficult to assess in children.

A second possibility is that feedback reduces self-confidence in higher knowledge learners and ultimately hinders learning (Kluger & DeNisi, 1996). Learners with some prior knowledge likely have some expectation of performing well (e.g., Kluger & Adler, 1993), but have not mastered the task. This may lead to a heightened sensitivity to incorrect responses. For example, feedback on incorrect trials may have produced ego threat (i.e., a threat to one's positive self-image), and decreased children's confidence in their use of the taught strategy. This may have led them to revert to old (incorrect) strategies (either during problem solving or on the

posttest). Learners with low prior knowledge may be less susceptible to this attention on the self as they can attribute incorrect responses to their lack of knowledge or experience with no threat to their abilities. Although children's self-assessment responses do not provide evidence for this interpretation, there was a restricted range of scores, which may have limited the usefulness of the measure. Future research should use varied response scales or experimentally test the role of ego threat. For example, one could manipulate expectations by telling some children that the task is hard and meant for older children. These children should expect to receive negative feedback and not feel threatened by it, in which case they may benefit from it relative to no feedback.

Limitations, Future Directions, and Conclusions

Despite the positive contributions of the current study, several limitations suggest directions for future research. More work is needed to better understand the negative effects of feedback and the potential contributions of strategy changes, the redundancy effect, and attention on the self. For example, one possible way to test the role of attention on the self is to manipulate the source of the feedback provided. Feedback from more impersonal sources (e.g., answer key, computer) may be less threatening to one's self-image relative to personal sources (e.g., teacher), and ultimately result in a positive effect relative to no feedback. A related issue is the need to examine how long these negative effects last and whether they influence future learning. In previous work, we found that the negative effects of correct-answer feedback persisted 2 weeks later (Fyfe et al., 2012). Unfortunately, in the current study we only assessed children's knowledge on an immediate posttest. In Experiment 1, we intended to administer a delayed retention test, but scheduling issues in the schools prevented us from doing so.

Future research also should test the generalizability of these results to different tasks, settings, and feedbacks schedules. Across four experiments (current study and Fyfe et al., 2012), researchers have found that verification and correct-answer feedback can have negative effects relative to no feedback, but only in the context of elementary schoolchildren learning to solve math equivalence problems. Previous research also has found negative effects of feedback, but primarily with adults in non-problem-solving domains (see Bangert-Drowns et al., 1991). Further, the one-on-one setting may have played a key role. For example, the evaluative nature of feedback may be particularly salient in a one-on-one tutoring context relative to a classroom context in which individualized attention is reduced. Also, because of the one-on-one aspect, we were able to provide feedback immediately after each problem or immediately after the problem set, which may not be feasible in classrooms with children finishing tasks at different times. A longer delay in feedback may have led to different results. Indeed, at least one study suggested that summative feedback provided the next day led to greater benefits than summative feedback right after the task (Butler et al., 2007). This study was with adults completing a memory task, but does suggest that the length of the feedback delay may matter.

Finally, more work is needed to examine the impact of various feedback types. Negative effects of feedback for higher knowledge learners have been found with feedback that varies in content (i.e., focused on answers vs. focused on strategies), amount (i.e., veri-

fication vs. verification plus the correct answer), and timing (i.e., immediate vs. summative) (current study and Fyfe et al., 2012). However, all versions of our feedback have provided relatively minimal input. Other types of feedback provide elaborated information, such as a conceptual rationale of the correct answer or a hint on a correct problem-solving strategy. For example, Narciss and Huth (2006) found positive effects of “bug-related feedback” in a subtraction task, which flags the specific error made, provides a hint on the correct strategy, and (if errors persist) presents a step-by-step solution strategy along with the correct answer. On the one hand, elaborated feedback may be more beneficial as it provides more information for error correction and understanding. On the other hand, elaborated feedback may be more harmful because the larger amount of information requires more cognitive processing, which may result in increased cognitive interference. Indeed, some studies have found that including more information in the feedback message leads to lower learning (e.g., Phye, 1979; Wentling, 1973).

In conclusion, the present study provides causal evidence for a specific moderator that can help explain both positive and negative effects of some types of feedback. In particular, children with no knowledge of a correct strategy benefitted from verification feedback during problem solving. In contrast, children with induced knowledge of a correct strategy learned more from problem solving alone. The latter result is consistent with recent research on the need for “productive struggle” during learning (Kapur, 2012; Schwartz, Chase, Oppezzo, & Chin, 2011). The idea is that students benefit from periods of exploration during which they engage with relevant problems with minimal external guidance. This need for productive struggle also has been recognized by the National Council of Teachers of Mathematics (2014). Too often,

teachers jump in to rescue students by breaking down the task and guiding students step by step through the difficulties. Although well-intentioned, such rescuing undermines the efforts of students, lowers the cognitive demand of the task, and deprives students of opportunities to engage fully in making sense of the mathematics. (National Council of Teachers of Mathematics, 2014, p. 48)

Providing minimal feedback to higher knowledge learners may be another form of “rescuing” that ultimately hinders learning.

References

- Alfieri, L., Brooks, P. J., Aldrich, N. J., & Tenenbaum, H. R. (2011). Does discovery-based instruction enhance learning? *Journal of Educational Psychology, 103*, 1–18. <http://dx.doi.org/10.1037/a0021017>
- Alibali, M. W. (1999). How children change their minds: Strategy change can be gradual or abrupt. *Developmental Psychology, 35*, 127–145. <http://dx.doi.org/10.1037/0012-1649.35.1.127>
- Alibali, M. W., Phillips, K. M. O., & Fischer, A. D. (2009). Learning new problem-solving strategies leads to changes in problem representation. *Cognitive Development, 24*, 89–101. <http://dx.doi.org/10.1016/j.cogdev.2008.12.005>
- Bangert-Drowns, R. L., Kulik, C.-L. C., Kulik, J. A., & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research, 61*, 213–238. <http://dx.doi.org/10.3102/00346543061002213>
- Bohlmann, N. L., & Fenson, L. (2005). The effects of feedback on perseverative errors in preschool aged children. *Journal of Cognition and Development, 6*, 119–131. http://dx.doi.org/10.1207/s15327647jcd0601_7
- Brainerd, C. J. (1972). Reinforcement and reversibility in quantity conservation acquisition. *Psychonomic Science, 27*, 114–116. <http://dx.doi.org/10.3758/BF03328907>
- Butler, A. C., Karpicke, J. D., & Roediger, H. L., III. (2007). The effect of type and timing of feedback on learning from multiple-choice tests. *Journal of Experimental Psychology: Applied, 13*, 273–281. <http://dx.doi.org/10.1037/1076-898X.13.4.273>
- Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research, 65*, 245–281. <http://dx.doi.org/10.3102/00346543065003245>
- Chesney, D. L., McNeil, N. M., Matthews, P. G., Byrd, C. E., Petersen, L. A., Wheeler, M. C., . . . Dunwiddie, A. E. (2014). Organization matters: Mental organization of addition knowledge relates to understanding math equivalence in symbolic form. *Cognitive Development, 30*, 30–46. <http://dx.doi.org/10.1016/j.cogdev.2014.01.001>
- Clariana, R. B., Wagner, D., & Roher Murphy, L. C. (2000). Applying a connectionist description of feedback timing. *Educational Technology Research and Development, 48*, 5–22. <http://dx.doi.org/10.1007/BF02319855>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.
- Coyle, T. R., & Bjorklund, D. F. (1997). Age differences in, and consequences of, multiple- and variable-strategy use on a multitrial sort-recall task. *Developmental Psychology, 33*, 372–380. <http://dx.doi.org/10.1037/0012-1649.33.2.372>
- Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods: A handbook for research on interactions*. New York, NY: Irvington.
- Evans, M. A., Barraball, L., & Eberle, T. (1998). Parental response to miscues during child-to-parent book reading. *Journal of Applied Developmental Psychology, 19*, 67–84. [http://dx.doi.org/10.1016/S0193-3973\(99\)80028-8](http://dx.doi.org/10.1016/S0193-3973(99)80028-8)
- Falkner, K. P., Levi, L., & Carpenter, T. P. (1999). Children’s understanding of equality: A foundation for algebra. *Teaching Children Mathematics, 6*, 232–236.
- Frantom, C. G., Green, K. E., & Hoffman, E. R. (2002). Measure development: The children’s attitudes toward technology scale (CATS). *Journal of Educational Computing Research, 26*, 249–263. <http://dx.doi.org/10.2190/DWAF-8LEQ-74TN-BL37>
- Fyfe, E. R., DeCaro, M. S., & Rittle-Johnson, B. (2015). When feedback is cognitively-demanding: The importance of working memory capacity. *Instructional Science, 43*, 73–91. <http://dx.doi.org/10.1007/s11251-014-9323-8>
- Fyfe, E. R., Rittle-Johnson, B., & DeCaro, M. S. (2012). The effects of feedback during exploratory mathematics problem solving: Prior knowledge matters. *Journal of Educational Psychology, 104*, 1094–1108. <http://dx.doi.org/10.1037/a0028389>
- Gaddes, W. H., & Crockett, D. J. (1975). The Spreen–Benton aphasia tests, normative data as a measure of normal language development. *Brain and Language, 2*, 257–280. [http://dx.doi.org/10.1016/S0093-934X\(75\)80070-8](http://dx.doi.org/10.1016/S0093-934X(75)80070-8)
- Gielen, S., Peeters, E., Dochy, F., Onghena, P., & Struyven, K. (2010). Improving the effectiveness of peer feedback on learning. *Learning and Instruction, 20*, 304–315. <http://dx.doi.org/10.1016/j.learninstruc.2009.08.007>
- Hart, S. G., & Staveland, L. E. (1988). Development of NASA–TLX: Results of experimental and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139–183). Amsterdam, The Netherlands: North-Holland. [http://dx.doi.org/10.1016/S0166-4115\(08\)62386-9](http://dx.doi.org/10.1016/S0166-4115(08)62386-9)
- Hattie, J., & Gan, M. (2011). Instruction based on feedback. In R. Mayer & P. Alexander (Eds.), *Handbook of research on learning and instruction* (pp. 249–271). New York, NY: Routledge.

- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77, 81–112. <http://dx.doi.org/10.3102/003465430298487>
- Hattikudur, S., & Alibali, M. W. (2010). Learning about the equal sign: Does comparing with inequality symbols help? *Journal of Experimental Child Psychology*, 107, 15–30. <http://dx.doi.org/10.1016/j.jecp.2010.03.004>
- Hofer, S., Nussbaumer, D., & Schneider, M. (2011, August/September). Practice without feedback can increase the adaptivity of strategy choices. In J. Torbeyns (Chair), *Strategy flexibility: Analyzing its related structures and processes*. Symposium presented at the EARLI Conference, Exeter, England.
- Hoover-Dempsey, K. V., Battiato, A. C., Walker, J. T., Reed, R. P., DeJong, J. M., & Jones, K. P. (2001). Parental involvement in homework. *Educational Psychologist*, 36, 195–209. http://dx.doi.org/10.1207/S15326985EP3603_5
- Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review*, 19, 509–539. <http://dx.doi.org/10.1007/s10648-007-9054-3>
- Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist*, 38, 23–31. http://dx.doi.org/10.1207/S15326985EP3801_4
- Kamins, M. L., & Dweck, C. S. (1999). Person versus process praise and criticism: Implications for contingent self-worth and coping. *Developmental Psychology*, 35, 835–847. <http://dx.doi.org/10.1037/0012-1649.35.3.835>
- Kapur, M. (2012). Productive failure in learning the concept of variance. *Instructional Science*, 40, 651–672. <http://dx.doi.org/10.1007/s11251-012-9209-6>
- Kieran, C. (1981). Concepts associated with the equality symbol. *Educational Studies in Mathematics*, 12, 317–326. <http://dx.doi.org/10.1007/BF00311062>
- Kluger, A. N., & Adler, S. (1993). Person-versus computer-mediated feedback. *Computers in Human Behavior*, 9, 1–16. [http://dx.doi.org/10.1016/0747-5632\(93\)90017-M](http://dx.doi.org/10.1016/0747-5632(93)90017-M)
- Kluger, A. N., & DeNisi, A. (1996). Effects of feedback intervention on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119, 254–284. <http://dx.doi.org/10.1037/0033-2909.119.2.254>
- Knuth, E. J., Stephens, A. C., McNeil, N. M., & Alibali, M. W. (2006). Does understanding the equal sign matter? Evidence from solving equations. *Journal for Research in Mathematics Education*, 37, 297–312.
- Krause, U.-M., Stark, R., & Mandl, H. (2009). The effects of cooperative learning and feedback on e-learning in statistics. *Learning and Instruction*, 19, 158–170. <http://dx.doi.org/10.1016/j.learninstruc.2008.03.003>
- Kulhavy, R. W. (1977). Feedback in written instruction. *Review of Educational Research*, 47, 211–232. <http://dx.doi.org/10.3102/00346543047002211>
- Langer, E. J. (2000). Mindful learning. *Current Directions in Psychological Science*, 9, 220–223. <http://dx.doi.org/10.1111/1467-8721.00099>
- Luwel, K., Foustana, A., Papadatos, Y., & Verschaffel, L. (2011). The role of intelligence and feedback in children's strategy competence. *Journal of Experimental Child Psychology*, 108, 61–76. <http://dx.doi.org/10.1016/j.jecp.2010.06.001>
- Mason, J. R., & Bruning, R. (2001). *Providing feedback in computer-based instruction: What research tells us*. Lincoln: University of Nebraska, Center of Instructional Innovation. Retrieved from <http://dwb4.unl.edu/dwb/Research/MB/MasonBruning.html>
- Matthews, P., & Rittle-Johnson, B. (2009). In pursuit of knowledge: Comparing self-explanations, concepts, and procedures as pedagogical tools. *Journal of Experimental Child Psychology*, 104, 1–21. <http://dx.doi.org/10.1016/j.jecp.2008.08.004>
- Matthews, P. G., Rittle-Johnson, B., McEldoon, K., & Taylor, R. (2012). Measure for measure: What combining diverse measures reveals about children's understanding of the equal sign as an indicator of mathematical equality. *Journal for Research in Mathematics Education*, 43, 220–254.
- McGilly, K., & Siegler, R. S. (1989). How children choose among serial recall strategies. *Child Development*, 60, 172–182. <http://dx.doi.org/10.2307/1131083>
- McNeil, N. M. (2008). Limitations to teaching children $2 + 2 = 4$: Typical arithmetic problems can hinder learning of mathematical equivalence. *Child Development*, 79, 1524–1537. <http://dx.doi.org/10.1111/j.1467-8624.2008.01203.x>
- McNeil, N. M., & Alibali, M. W. (2005). Why won't you change your mind? Knowledge of operational patterns hinders learning and performance on equations. *Child Development*, 76, 883–899. <http://dx.doi.org/10.1111/j.1467-8624.2005.00884.x>
- McNeil, N. M., Fyfe, E. R., Petersen, L. A., Dunwiddie, A. E., & Brletic-Shiple, H. (2011). Benefits of practicing $4 = 2 + 2$: Nontraditional problem formats facilitate children's understanding of mathematical equivalence. *Child Development*, 82, 1620–1633. <http://dx.doi.org/10.1111/j.1467-8624.2011.01622.x>
- Miyake, A., & Shah, P. (Eds.). (1999). *Models of WM: Mechanisms of active maintenance and executive control*. New York, NY: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9781139174909>
- Moreno, R. (2004). Decreasing cognitive load for novice students: Effects of explanatory versus corrective feedback in discovery-based multimedia. *Instructional Science*, 32, 99–113. <http://dx.doi.org/10.1023/B:TRUC.0000021811.66966.1d>
- Mory, E. H. (2004). Feedback research revisited. In D. Jonassen (Ed.), *Handbook of research on educational communications and technology: A project for the Association for Educational Communications and Technology* (2nd ed., pp. 745–783). Mahwah, NJ: Erlbaum.
- Narciss, S., & Huth, K. (2004). How to design informative tutoring feedback for multi-media learning. In H. M. Niegemann, D. Leutner, & R. Brunken (Eds.), *Instructional design for multimedia learning* (pp. 181–195). Munster, NY: Waxman.
- Narciss, S., & Huth, K. (2006). Fostering achievement and motivation with bug-related tutoring feedback in a computer-based training for written subtraction. *Learning and Instruction*, 16, 310–322. <http://dx.doi.org/10.1016/j.learninstruc.2006.07.003>
- National Council of Teachers of Mathematics. (2014). *Principles to action: Ensuring mathematical success for all*. Reston, VA: Author.
- Nihalani, P. K., Mayrath, M., & Robinson, D. H. (2011). When feedback harms and collaboration helps in computer simulation environments: An expertise reversal effect. *Journal of Educational Psychology*, 103, 776–785. <http://dx.doi.org/10.1037/a0025276>
- Nussbaumer, D., Schneider, M., & Stern, E. (2014). The influence of feedback on the flexibility of strategy choices in algebra problem solving. In P. Bello, M. Guarini, M. McShane, & B. Scasslati (Eds.), *Proceedings of the thirty-sixth annual conference of the Cognitive Science Society* (pp. 1102–1107). Quebec City, Canada: Cognitive Science Society.
- Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist*, 38, 63–71. http://dx.doi.org/10.1207/S15326985EP3801_8
- Petersen, L. A., & McNeil, N. M. (2013). Effects of perceptually rich manipulatives on preschoolers' counting performance: Established knowledge counts. *Child Development*, 84, 1020–1033. <http://dx.doi.org/10.1111/cdev.12028>
- 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, 525–556. <http://dx.doi.org/10.3102/00346543074004525>

- Phye, G. D. (1979). The processing of informative feedback about multiple-choice test performance. *Contemporary Educational Psychology, 4*, 381–394. [http://dx.doi.org/10.1016/0361-476X\(79\)90057-2](http://dx.doi.org/10.1016/0361-476X(79)90057-2)
- Pianta, R. C., Belsky, J., Houts, R., & Morrison, F. (2007). Opportunities to learn in America's elementary classrooms. *Science, 315*, 1795–1796. <http://dx.doi.org/10.1126/science.1139719>
- Rey, G. D., & Buchwald, F. (2011). The expertise reversal effect: Cognitive load and motivational explanations. *Journal of Experimental Psychology: Applied, 17*, 33–48. <http://dx.doi.org/10.1037/a0022243>
- Rittle-Johnson, B. (2006). Promoting transfer: Effects of self-explanation and direct instruction. *Child Development, 77*, 1–15. <http://dx.doi.org/10.1111/j.1467-8624.2006.00852.x>
- Rittle-Johnson, B., & Alibali, M. (1999). Conceptual and procedural knowledge of mathematics: Does one lead to the other? *Journal of Educational Psychology, 91*, 175–189. <http://dx.doi.org/10.1037/0022-0663.91.1.175>
- Rittle-Johnson, B., Matthews, P. G., Taylor, R. S., & McEldoon, K. (2011). Assessing knowledge of mathematical equivalence: A construct modeling approach. *Journal of Educational Psychology, 103*, 85–104. <http://dx.doi.org/10.1037/a0021334>
- Ryan, R. M. (1982). Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of Personality and Social Psychology, 43*, 450–461. <http://dx.doi.org/10.1037/0022-3514.43.3.450>
- Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*, 147–177. <http://dx.doi.org/10.1037/1082-989X.7.2.147>
- Schwartz, D. L., Chase, C. C., Oppezzo, M. A., & Chin, D. B. (2011). Practicing versus inventing with contrasting cases: The effects of telling first on learning and transfer. *Journal of Educational Psychology, 103*, 759–775. <http://dx.doi.org/10.1037/a0025140>
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*, 153–189. <http://dx.doi.org/10.3102/0034654307313795>
- Siegler, R. S., & Shipley, C. (1995). Variation, selection, and cognitive change. In T. J. Simon & G. S. Halford (Eds.), *Developing cognitive competence: New approaches to process modeling* (pp. 31–76). Hillsdale, NJ: Erlbaum.
- Smith, T. A., & Kimball, D. R. (2010). Learning from feedback: Spacing and the delay-retention effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 80–95. <http://dx.doi.org/10.1037/a0017407>
- Snow, R. E. (1996). Aptitude development and education. *Psychology, Public Policy, and Law, 2*, 536–560. <http://dx.doi.org/10.1037/1076-8971.2.3-4.536>
- Sweller, J., Ayres, P., & Kalyuga, S. (2011). The redundancy effect. In J. M. Spector & S. P. Lajoie, *Cognitive load theory* (pp. 141–154). New York, NY: Springer. http://dx.doi.org/10.1007/978-1-4419-8126-4_11
- Sweller, J., van Merriënboer, J., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*, 251–296. <http://dx.doi.org/10.1023/A:1022193728205>
- Tobias, S. (2010). The expertise reversal effect and aptitude treatment interaction research [Commentary]. *Instructional Science, 38*, 309–314. <http://dx.doi.org/10.1007/s11251-009-9103-z>
- Weaver, J. F. (1973). The symmetric property of the equality relation and young children's ability to solve open addition and subtraction sentences. *Journal for Research in Mathematics Education, 4*, 45–56. <http://dx.doi.org/10.2307/749023>
- Wechsler, D. (2003). *Wechsler Intelligence Scale for Children—WISC-IV* (4th ed.). San Antonio, TX: Harcourt Assessment.
- Wentling, T. L. (1973). Mastery versus nonmastery instruction with varying test item feedback treatments. *Journal of Educational Psychology, 65*, 50–58. <http://dx.doi.org/10.1037/h0034820>

Appendix

Cognitive Load Validation

Rationale

Cognitive load is rarely measured in young children and we are not aware of any validated scales for use with elementary schoolchildren. We developed a nine-item measure to be suitable for young children. These items were developed based on the theoretical construct and definition of cognitive load, the language limitations of young children, and pilot work with a group of elementary schoolchildren. The measure included three subscales intended to tap distinct aspects of cognitive load: mental effort, mental frustration, and task difficulty. Each of the three subscales was adapted from existing items used with older children or adults. The mental effort and mental frustration scales were based on modified items from the NASA Task Load Index (Hart & Staveland, 1988), a measure used in previous studies

with adults to assess cognitive load (Rey & Buchwald, 2011). The task difficulty scale was based on a common subjective rating scale used to assess cognitive load with adolescents and adults (see Paas, Tuovinen, Tabbers, & Van Gerven, 2003). We checked the validity of the measure in several ways. We examined its similarity with a common rating scale used with adults and whether it predicted learning outcomes. We also measured its association with relevant variables. For example, we expected cognitive load to negatively relate to motivation, as theory suggests that tasks that produce cognitive overload can reduce motivation. We also expected cognitive load to have no relation to other cognitive variables that assess different aspects of cognitive processing (e.g., working memory, retrieval fluency). Results are outlined in the following.

(Appendix continues)

Table A1
Cognitive Load Items

Construct tapped	Item
Mental effort	1. I had to think hard to do this math work. 2. I had to keep track of a lot of things at once to do this math work. 3. I had to think about a lot of things to do this math work.
Frustration	1. I was stressed and irritated when I did this math work. 2. <i>When I did this math work I felt calm and relaxed.</i> 3. I was discouraged and annoyed when I did this math work.
Task difficulty	1. <i>This math work was presented in an easy way to understand.</i> 2. Compared to other math work I've done, this math work was hard. 3. This math work was very confusing.

Note. Italicized items are reverse scored.

Participants

The results are based on the final sample in Experiment 1, which contained 108 second- and third-grade children (M age = 8.4 years, min = 7.2 years, max = 9.8 years; 67 girls, 41 boys).

Measures

Cognitive load. Children’s cognitive load was assessed using a nine-item measure we developed to be suitable for young children (see Table A1). The measure includes three 3-item subscales intended to tap distinct aspects of cognitive load: mental effort, mental frustration, and task difficulty. Children responded to each item by circling their answer on a 4-point scale: 1 (*strongly disagree*), 2 (*disagree*), 3 (*agree*), 4 (*strongly agree*). This response scale has been used in previous research with elementary schoolchildren (Frantom, Green, & Hoffman, 2002). For each item, the children’s response was assigned a number from one to four and scores were formed by averaging their responses across the three subscale items. A 10th item was administered for validation purposes. It was adopted from prior adult studies in the cognitive load literature (see Paas et al., 2003). It read: “How easy or difficult was this math task to understand?” Children responded on a 7-point scale ranging from 1 (*extremely easy*) to 7 (*extremely difficult*). Children were assigned a score from one to seven. These items were administered during a one-on-one tutoring session immediately following a mathematics problem-solving activity.

Motivation. We administered three items from the interest and enjoyment scale of the Intrinsic Motivation Inventory (Ryan, 1982). The items were as follow: “I enjoyed solving the math

problems very much.” “These math problems were fun to do.” “The math problems were very interesting.” Children responded to each item by circling their answer on a 4-point scale ranging from 1 (*strongly disagree*) to 4 (*strongly agree*). For each item, the children’s response was assigned a number from one to four and motivation scores were formed by averaging their responses across the three items. Motivation was assessed immediately following the cognitive load items.

Working memory capacity. We measured working memory capacity, which supports learners’ ability to actively select, regulate, and process task-relevant information, using the backward digit-span task (Wechsler, 2003). Children were read a series of numbers at a rate of one per second and were asked to repeat the numbers in reverse order. Number series length began at two and ended at a maximum of eight. There were two items per series length. The task was discontinued when a child recalled both items in a series of a given length incorrectly. Children received one point for each series recalled correctly in backward order. Working memory was assessed immediately after the posttest.

Retrieval fluency. We also measured retrieval fluency (Gadde & Crockett, 1975)—the controlled search and retrieval of information from long-term memory. Children were asked to name as many items from a category (i.e., “animals” and “things to eat”) as possible within a 1-min span. Children received one point for each distinct item named in a category. Scores from each category were averaged together to form a fluency score. Fluency was assessed immediately following the working memory capacity task.

(Appendix continues)

Results

Descriptive statistics. Table A2 contains the descriptive statistics for the cognitive load items. The task difficulty scale had desirable qualities. The mean was near the middle of the response scale, the scores were sufficiently variable, and the distribution was relatively symmetric with a skewness value close to zero. Further, each individual item was related to the total scale score as indicated by moderate item-total correlations. Neither the mental effort nor the frustration scale exhibited these positive qualities to the same degree. We also explored an aggregate scale in which we used all nine items.

Evidence for reliability. Internal consistency, as assessed by Cronbach’s alpha, was high for the task difficulty scale ($\alpha = .77$), somewhat lower for the aggregate scale ($\alpha = .72$), and low for the mental effort scale ($\alpha = .33$) and frustration scale ($\alpha = .58$). Further, the task difficulty items also were all positively and significantly correlated with each other: Items 1 and 2, $r(106) = .48, p < .001$, Items 1 and 3, $r(106) = .51, p < .001$, Items 2 and 3, $r(106) = .57, p < .001$. For the mental effort scale, the interitem correlations were .00, .20, and .21, only two of which were statistically significant. For the frustration scale, the correlations were .17, .35, and .42, only two of which were statistically significant.

Evidence for validity. We examined the relation between the children’s subjective ratings on the three cognitive load scales (and the aggregate scale) with their ratings on an existing measure of cognitive load. We also examined the relation between these

Table A2
Descriptive Statistics for Cognitive Load Items

Item	Item-total correlation	α	M^a	SD	Skewness
Mental effort scale	—	.33	3.07	0.58	-0.41
Item 1	.13	—	3.19	0.86	-0.84
Item 2	.14	—	2.94	0.93	-0.58
Item 3	.29	—	3.07	0.88	-0.73
Frustration scale	—	.58	1.89	0.68	0.63
Item 1	.50	—	2.06	1.01	0.49
Item 2	.32	—	1.97	0.87	0.58
Item 3	.37	—	1.65	0.89	1.41
Task difficulty scale	—	.77	2.29	0.85	0.24
Item 1	.56	—	2.18	0.99	0.39
Item 2	.61	—	2.44	1.04	0.07
Item 3	.63	—	2.25	1.04	0.29
Aggregate scale	—	.72	2.42	0.53	0.06

^a Out of four.

Table A3
Correlations Between Cognitive Load Scales and Relevant Variables

	Adult CL item	Posttest score	Intrinsic motivation	WM capacity	Retrieval fluency
Adult CL item	—	-.28**	-.46**	-.12	.03
Mental effort	.02	-.04	.15	-.07	.03
Frustration	.39**	-.18	-.37**	-.07	-.06
Task difficulty	.64**	-.23*	-.31**	-.13	-.12
Aggregate	.52**	-.21*	-.27**	-.13	-.08

Note. Adult CL item = cognitive load item adopted from prior adult studies; posttest score = children’s total score on an assessment of math equivalence understanding; WM capacity = working memory capacity.
* $p < .05$. ** $p < .01$.

ratings and other relevant variables. Correlations are shown in Table A3. As shown in the table, ratings on the task difficulty scale had a strong, positive correlation with ratings on the existing cognitive load item (see Adult CL on Table A3) demonstrating good convergent validity. Ratings on the frustration scale also were correlated with the existing measure, but ratings on the effort scale were not. Further, the task difficulty scale and the Adult CL item were negatively correlated with scores on the posttest, supporting the idea that children who found the task more difficult during the intervention indeed knew less at posttest. The task difficulty scale and the Adult CL item also were negative correlated with intrinsic motivation scores, supporting the idea that children who found the task more difficult also were less motivated by the task during the intervention. More important, task difficulty scores were unrelated to different cognitive constructs (i.e., working memory capacity and retrieval fluency) demonstrating some discriminant validity.

Summary. Overall, several pieces of evidence support the reliability and validity of the task difficulty scale for assessing children’s subjective cognitive load. However, the same was not true for the mental effort scale and the frustration scale. These latter scales should be dropped. The aggregate scale that included all nine items together functioned somewhat similar to the task difficulty scale alone. However, it had a somewhat lower alpha and for purposes of parsimony, the three-item task difficulty scale was adopted.

Received December 19, 2014
Revision received April 21, 2015
Accepted April 22, 2015 ■