Journal of Educational Psychology
2012, Vol. 104, No. 4, 1094–1108


0022-0663/12/$12.00 DOI: 10.1037/a0028389

# The Effects of Feedback During Exploratory Mathematics Problem Solving: Prior Knowledge Matters

Emily R. Fyfe and Bethany Rittle-Johnson
Vanderbilt University

Marci S. DeCaro
University of Louisville

Providing exploratory activities prior to explicit instruction can facilitate learning. However, the level of guidance provided during the exploration has largely gone unstudied. In this study, we examined the effects of 1 form of guidance, feedback, during exploratory mathematics problem solving for children with varying levels of prior domain knowledge. In 2 experiments, 2nd- and 3rd-grade children solved 12 novel mathematical equivalence problems and then received brief conceptual instruction. After solving each problem, they received (a) no feedback, (b) outcome feedback, or (c) strategy feedback. In both experiments, prior knowledge moderated the impact of feedback on children's learning. Children with little prior knowledge of correct solution strategies benefited from feedback during exploration, but children with some prior knowledge of a correct solution strategy benefited more from exploring without feedback. These results suggest that theories of learning need to incorporate the role of prior knowledge and that providing feedback may not always be optimal.

*Keywords:* guided discovery, feedback, expertise reversal effect, mathematical equivalence

Contemporary learning theorists often endorse guided discovery learning, as opposed to discovery or explicit instruction alone, as the best method to facilitate understanding (e.g., Mayer, 2004; Schwartz, Chase, Oppezzo, & Chin, 2011; Wise & O'Neill, 2009). Providing an exploratory activity with subsequent instruction is one form of guided discovery that has been shown to aid learning (e.g., Schwartz & Bransford, 1998). However, the level of guidance provided during the exploratory activity has largely gone unstudied. Feedback is one form of guidance that could potentially boost the efficacy of exploration by guiding the learner's search for information. In two experiments, we examined how and for whom feedback might enhance learning during exploration prior to explicit instruction. We investigated these questions in the context of children exploring mathematical equivalence, a fundamental concept in arithmetic and algebra.

This article was published Online First May 7, 2012.

Emily R. Fyfe and Bethany Rittle-Johnson, Department of Psychology and Human Development, Vanderbilt University; Marci S. DeCaro, Department of Psychological and Brain Sciences, University of Louisville.

This research was supported by National Science Foundation Grant DRL-0746565 to Bethany Rittle-Johnson and by U.S. Department of Education, Institute of Education Sciences Training Grant R305B080025. We thank Laura McLean, Abbey Loehr, Maryphyllis Crean, Lucy Rice, Rachel Ross, and Polly Colgan for their assistance with data collection and coding as well as the staff, teachers, and children at Westmeade Elementary School, Jones Paideia Elementary Magnet School, and St. Pius X Classical Academy for their participation in this research project.

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

## Guided Discovery Learning

An emerging consensus is that people learn best through some form of *guided discovery*, which combines exploration and instruction (e.g., Alfieri, Brooks, Aldrich, & Tenenbaum, 2011; Hmelo-Silver, Duncan, & Chinn, 2007; Lorch et al., 2010; Mayer, 2004). The assumption is that "students need enough freedom to become cognitively active in the process of sense making, and . . . enough guidance so that their cognitive activity results in the construction of useful knowledge" (Mayer, 2004, p. 16). There is currently not a precise definition of guided discovery, largely because the term captures such a broad range of activities including problem-based learning (Barrows & Tamblyn, 1980), inquiry learning (Rutherford, 1964), and constructivist learning (Steffe & Gale, 1995). We adopt the general framework outlined by Alfieri and colleagues (2011) and define guided discovery as exploratory learning tasks that are supplemented with some form of instructional guidance. Learning tasks are exploratory if learners have not received instruction on how to complete them, and instructional guidance encompasses a variety of tools, from in-depth instruction manuals to minimal feedback or coaching. Alfieri et al.'s (2011) recent meta-analysis revealed the superiority of guided discovery over both explicit instruction and unguided discovery learning.

Providing exploratory activities prior to explicit instruction is one form of guided discovery that has been recommended by researchers in education and psychology alike (e.g., Hiebert & Grouws, 2007; Schwartz, Lindgren, & Lewis, 2009), and it is the form that we focus on in this article. For example, a number of mathematics education researchers promote the belief that "each person must struggle with a situation or problem first in order to make sense of the information he or she hears later" (Stigler & Hiebert, 1998, p. 3). Similarly, several researchers in psychology suggest that exploration facilitates the development of differentiated knowledge of the target domain, which prepares people to learn more deeply from future instruction than would be possible otherwise (e.g., Schwartz & Bransford, 1998; Schwartz et al., 2009).

Increasing evidence supports the claim that exploration prior to instruction is beneficial (e.g., DeCaro & Rittle-Johnson, 2011; Schwartz & Bransford, 1998; Schwartz & Martin, 2004; Schwartz et al., 2011). For example, elementary school children learned new math concepts better if they solved unfamiliar problems *before* receiving instruction, rather than vice versa (DeCaro & Rittle-Johnson, 2011). Similarly, middle school students who explored a set of novel density problems before hearing a lecture exhibited better transfer than students who heard the lecture first and practiced the problems after (Schwartz et al., 2011). However, further research is needed to optimize this form of guided discovery. For example, the level of guidance provided during the exploratory phase has largely gone unstudied. In this study, we examined the effects of providing guidance versus no guidance during exploration prior to instruction.

## Feedback as One Form of Guidance

Feedback is touted as one form of guidance that may be particularly effective. *Feedback* is any information about performance or understanding that the learner can use to confirm, reject, or modify prior knowledge (Mory, 2004). Based on their meta-analysis, Alfieri et al. (2011) specifically recommend “providing timely feedback” as an optimal form of guidance (p. 13). Also, a recent review indicates that guided discovery methods that provide feedback during problem solving enable deeper learning than unguided discovery methods (Mayer, 2004). For example, kindergarteners generally struggle with conservation tasks; however, providing feedback to children solving novel conservation tasks improved their explanations on subsequent problems (Brainerd, 1972). Although these studies did not focus on exploration *prior* to instruction, they suggest that feedback during exploratory problem solving can be beneficial.

In addition to these endorsements, there are several reasons to suggest feedback is beneficial. First, decades of research have demonstrated powerful effects of feedback for student achievement in the classroom (see Hattie & Timperley, 2007). Indeed, one meta-analysis comparing students who received feedback interventions to control groups who received no feedback included 470 effects and revealed an average positive effect size of .38 for feedback on student performance measures (e.g., reading errors, arithmetic computations, and so forth; Kluger & DeNisi, 1996). The effects spanned a variety of feedback types, task types, and means of presentation.

Second, past research indicates that the primary function of feedback is to identify errors and encourage the adoption of correct alternatives (e.g., Anderson, Kulhavy, & Andre, 1972; Birenbaum & Tatsuoka, 1987; Kulhavy, 1977). In these studies, participants who received feedback had a higher chance of correcting their initial errors than participants who did not. For example, Phye and Bender (1989) examined the role of feedback on a memory task and found that when feedback was not available, perseveration (i.e., making the same error multiple times) was the most frequent error type—a pattern not demonstrated in the feedback conditions. Also, though feedback may not be necessary for new strategy generation, it has been shown to speed up strategy generation relative to no feedback (Alibali, 1999). Together, these studies indicate that feedback helps learners reject erroneous ideas and search for more plausible alternatives.

Given these positive effects of feedback, it seems likely that it would improve the efficacy of exploration prior to instruction. Yet, all reviews of feedback research note the wide variability of effects on learning (e.g., Hattie & Gan, 2011; Kluger & DeNisi, 1996; Mory, 2004). For example, Kluger and DeNisi (1996) found an average positive effect size for feedback in their meta-analysis, but more than one third of the effects were negative, indicating that feedback in those cases actually decreased performance. The authors noted the need for future research to determine the conditions under which feedback is effective. More pointedly, a recent review of the impact of feedback in instructional settings called for work to explore feedback effects in relation to individual learner characteristics (Hattie & Gan, 2011). To that end, we explored the possibility that feedback enhances exploration for only a certain subset of learners.

## The Role of Prior Knowledge

The feedback literature points to prior knowledge as a key characteristic to consider when evaluating the effects of feedback (e.g., Hannafin, Hannafin, & Dalton, 1993). Indeed, many agree that “to be effective, feedback needs to be . . . compatible with students’ prior knowledge” (Hattie & Timperley, 2007, p. 104). This idea is consistent with past work on aptitude by treatment interactions, which demonstrate that learner characteristics can determine whether an instructional technique will be effective or not (Cronbach & Snow, 1977; Snow, 1978; Snow & Lohman, 1984). In particular, learners with low prior knowledge often need substantial instructional support, whereas those with higher prior knowledge do not (see Kalyuga, 2007, and Tobias, 2009). For example, novices learn more from studying worked examples than from solving problems unaided. But as domain knowledge increases, independent problem solving becomes the superior learning activity (e.g., Kalyuga & Sweller, 2004). Overall, research on aptitude by treatment interactions suggest that learner characteristics generally, and prior knowledge specifically, can determine who will or will not benefit from a given technique.

One previous study has directly investigated the impact of feedback in relation to learners’ prior knowledge (Krause, Stark, & Mandl, 2009). College students studied statistics problems via worked examples and practice problems presented in a computer-based learning environment. Students with low prior knowledge in the domain exhibited better performance on a posttest if they received explicit feedback during training than if they did not. However, students with higher prior knowledge in the domain did not benefit from such feedback.

Additional research on feedback in problem-solving domains points to the particular importance of learners’ prior knowledge of correct solution strategies. Specifically, learners with little prior knowledge of domain-specific solution strategies seem to benefit from feedback, whereas learners with some knowledge of correct strategies may not. In one study, elementary school children practiced solving mathematical equivalence problems (e.g., $3 + 4 + 5 = 3 + \_\_$; Alibali, 1999). All the children had low prior knowledge, as they could not solve any problems correctly on a pretest. In this study, feedback supported the generation of more diverse strategies relative to a no feedback control. In contrast, a recent study found positive effects of problem solving *without* feedback for children with higher prior knowledge (Hofer, Nussbaumer, & Schneider, 2011). High-school students who already knew the target strategies solved algebraic equations with or without feedback. In both conditions, adaptive strategy selection (i.e., choosing the most efficient strategy given the problem type) increased at a similar rate, indicating that feedback was unnecessary. In line with this work, Luwel, Foustana, Papadatos, and Verschaffel (2011) examined children's performance on a numerosity judgment task that could be solved using two different correct strategies. Children who demonstrated neither strategy at pretest benefited greatly from feedback in terms of strategy use, efficiency, and adaptivity. In contrast, feedback had a much weaker effect for children who already knew the strategies at pretest (though these two groups of children also differed in general intelligence).

Synthesizing across these studies suggests that learners with little knowledge of correct solution strategies should benefit from feedback during problem exploration, whereas learners with some prior knowledge of correct strategies may not. A potential explanation for this effect comes from Siegler's (1996) overlapping waves theory. For learners with little to no knowledge of a correct strategy, the guiding effects of feedback may facilitate the acquisition of a correct strategy, which in turn may jumpstart the process of subsequent strategy changes, dampen the strength of existing incorrect strategies (Siegler, & Shipley, 1995), and prepare children to learn from future instruction by enhancing their knowledge of the problem space (e.g., Schwartz et al., 2009). However, for learners who have already acquired one or more correct strategies, problem-solving experience alone may suffice, as it allows for reflection on the applicability and efficiency of existing strategies (e.g., Shrager & Siegler, 1998). Thus, one possibility is that children with little knowledge of correct strategies need feedback to enhance initial acquisition of correct strategies, but learners with some knowledge of correct strategies do not.

Together, research on feedback and research on aptitude by treatment interactions indicate that domain knowledge may play a central role in moderating the effects feedback. It suggests that learners with limited prior knowledge may benefit from feedback, but learners with some prior knowledge, particularly of correct solution strategies, may not. However, past research has not evaluated the impact of feedback for children with variable prior knowledge within the same study, nor has it focused on feedback during exploration prior to instruction.

## Current Study

We examined the effects of feedback during exploration prior to instruction for children with varying levels of prior knowledge. We focused on mathematical equivalence problems (problems with operations on both sides of the equal sign, such as $3 + 4 + 5 = 3 + \_\_$). These problems are not typically included in elementary mathematics curricula (Perry, 1988; Seo & Ginsburg, 2003). A recent analysis revealed that of all instances of the equal sign in a textbook series for Grades 1–6, equations with operations on both sides of the equal sign accounted for just 4% of instances (Rittle-Johnson, Matthews, Taylor, & McEldoon, 2011). Further, decades of research have shown that elementary school children exhibit poor performance on mathematical equivalence problems (e.g., McNeil, 2008; Rittle-Johnson & Alibali, 1999), which often stems from misinterpretations of the equal sign as an operator symbol meaning "get the answer," as opposed to a relational symbol meaning "the same amount" (Kieran, 1981; McNeil & Alibali, 2005). Thus, mathematical equivalence problems are unfamiliar and difficult for elementary school children, providing an apt domain with which to investigate exploratory problem solving.

In the context of mathematics problem solving, two types of feedback seem particularly relevant: *outcome feedback* provides a judgment about the accuracy of the learner's response, whereas *strategy feedback* provides a judgment about how the learner obtained that response. Outcome feedback has been studied extensively and is generally related to positive outcomes (e.g., Kluger & DeNisi, 1996). In contrast, few empirical studies have examined the effects of strategy feedback (e.g., Ahmad, 1988; Luwel et al., 2011). The limited evidence suggests that strategy feedback can improve strategy selection relative to outcome feedback; however, more research is needed to examine its benefits across tasks and outcome measures. Our primary goal was to compare the effects of feedback versus no feedback during exploration prior to instruction. However, two types of feedback were included to explore whether different types differentially impact learning and also to bolster the empirical evaluation of strategy feedback.

In the study, children received a tutoring session that included exploratory problem solving followed by brief instruction. During problem solving, children received (a) no-feedback, (b) outcome feedback, or (c) strategy feedback after solving novel mathematical equivalence problems. After the session, children completed a posttest (immediately and 2 weeks later) that assessed conceptual and procedural knowledge of mathematical equivalence (Rittle-Johnson et al., 2011). *Conceptual knowledge* is an understanding of the principles governing a domain and *procedural knowledge* is the ability to execute action sequences (i.e., domain-specific strategies) to correctly solve problems (e.g., Rittle-Johnson & Alibali, 1999). We incorporated microgenetic methods, such as strategy reports (Siegler & Crowley, 1991), to explore how feedback influenced learning.

We hypothesized that children who received feedback would exhibit better procedural knowledge of mathematical equivalence than children who did not. However, research on feedback (e.g., Alibali, 1999) and research on aptitude by treatment interactions (see Kalyuga, 2007) suggest that learners with lower prior knowledge may benefit more from feedback as a source of guidance; thus, we expected this effect to be larger for children with lower prior knowledge. Differences were predicted in procedural knowledge because feedback was directed at children's problem solving, but we looked at potential differences in conceptual knowledge as well. We also explored why differences in procedural knowledge might occur by examining children's strategy use. Feedback has been shown to facilitate the generation of strategies (Alibali, 1999) and to reduce perseveration (e.g., Phye & Bender, 1989). The results from this study will help us understand if feedback is beneficial during exploratory problem solving prior to instruction, as well as how and for whom it works. Two experiments were conducted with the same basic design to evaluate the replicability of the findings.

## Experiment 1

### Method

**Participants.** Consent was obtained from 115 second- and third-grade children at a public school in middle Tennessee. Of those children, 93 met criteria for participation because they scored at or below 80% on both a conceptual and procedural knowledge measure at pretest. We used a liberal inclusion criterion to examine learning outcomes for children who varied in terms of prior knowledge, but still had room for growth. Six additional children had their data excluded: one for failing to complete the intervention, one for failing to follow directions, and four for missing the retention test. The final sample consisted of 87 children ($M$ age = 8 years 6 months; 52 girls; 35 boys; 44% White, 41% African American, 6% Hispanic, 9% other). Approximately 47% received free or reduced-price lunch. Teacher reports indicated that four of these children were receiving special education services; however, their performance did not differ from the sample norm so their data were included in all final analyses.

**Design.** The experiment had a pretest–intervention–posttest design followed by a 2-week retention test. For the brief tutoring intervention, children were randomly assigned to one of three conditions: strategy feedback ($n$ = 25), outcome feedback ($n$ = 31), or no feedback ($n$ = 31). It is important to note that the conditions were well matched in terms of background characteristics. There were no significant differences among children in the three conditions in terms of age, $F(2, 84) = 0.03$, $p = .97$; gender, $\chi^2(2, N = 87) = 2.29$, $p = .32$; grade, $\chi^2(2, N = 87) = 0.55$, $p = .76$; ethnicity, $\chi^2(6, N = 87) = 8.03$, $p = .24$; special education status, $\chi^2(2, N = 87) = 0.71$, $p = .70$; or free or reduced-price lunch status, $\chi^2(4, N = 87) = 3.08$, $p = .55$.

**Procedure.** Children completed a written pretest in their classrooms in one 30-min session. Within 1 week, those who met our inclusion criteria completed a one-on-one tutoring intervention and immediate posttest in a single session lasting approximately 45 min. This session was conducted in a quiet room at the school with one of two female experimenters. Approximately 2 weeks after the intervention session ($M$ = 14.0 days, $SD$ = 2.7), children completed the written retention test in small-group sessions in their classrooms.

The tutoring intervention began with exploratory problem solving. Children were asked to solve 12 mathematical equivalence problems presented one at a time on a computer screen using E-Prime 2.0 (Schneider, Eschman, & Zuccolotto, 2002). Specifically, they were asked to figure out the number that went in the box to make the number sentence true. The problems increased in difficulty with two exceptions. The first four problems consisted of three- and four-addend problems (e.g., $10 = 3 + \square$, $3 + 7 = 3 + \square$, $3 + 7 = \square + 6$). These were followed by six five-addend problems with a repeated addend on either side of the equal sign (e.g., $5 + 3 + 9 = 5 + \square$, $9 + 7 + 6 = \square + 6$). Two additional problems (the seventh and tenth) were simple three-addend problems (i.e., $9 = 6 + \square$, $7 = 6 + \square$). These were included in the block with the six more difficult problems to ensure children in the two feedback conditions received some positive feedback and to ensure all children were paying attention. After each problem, children reported how they solved the problem and received different kinds of feedback based on their condition.

In the *strategy-feedback condition*, children received feedback on how they solved each problem (e.g., "Good job! That is one correct way to solve that problem"; "Good try, but that is not a correct way to solve the problem"). The strategy feedback was based solely on the correctness of the child's verbal strategy report and did not depend on the correctness of the numerical answer (though these differed on only 3% of trials). For example, if a child reported using a correct strategy but obtained an incorrect answer (e.g., due to an arithmetic error), we provided positive strategy feedback. In the *outcome-feedback condition*, children received feedback on their answer to the problem. This included a judgment about the correctness of the answer as well as the correct response (e.g., "Good job! You got the right answer—$X$ is the correct answer"; "Good try, but you did not get the right answer—$X$ is the correct answer"). The outcome feedback was based solely on the correctness of the child's numerical answer and did not depend on the correctness of the strategy used (though these differed on only 4% of trials). We provided the correct answer because past work suggests this enhances the effects of outcome feedback (Kluger & DeNisi, 1996). For both conditions, feedback was presented verbally by the experimenter and visually on the computer screen. Finally, in the *no-feedback condition*, children did not receive any feedback after solving a problem and were simply told to go to the next one.

After the exploratory problem solving, all children received brief conceptual instruction on the relational function of the equal sign, adapted from past research (Matthews & Rittle-Johnson, 2009). The experimenter provided a definition of the equal sign, using a number sentence as an example. Specifically, $3 + 4 = 3 + 4$ was displayed on the screen while the experimenter identified the two sides of the problem, defined the equal sign as meaning "the same amount as," and explained how the left and right side of the problem were equal. The meaning of the equal sign was reiterated with four additional number sentences (e.g., $4 + 4 = 3 + 5$). Children were asked to answer simple questions and to identify the two sides of the number sentences to ensure they were attending to instruction. No procedures were discussed, and children were not asked to solve any mathematical equivalence problems during the instruction.

Between the exploratory problem solving and instruction, children completed a brief form of the mathematical equivalence assessment (midtest) to gauge the immediate effects of exploration prior to instruction. Additionally, children rated their subjective cognitive load (using two modified items from the NASA Task Load Index (Hart & Staveland, 1988) and completed several other measures not relevant to the current results. Given concerns about the validity of the cognitive load measure for use with children, results for this measure are not reported.

**Assessment and coding.** The mathematical equivalence assessment, adapted from past work (Rittle-Johnson et al., 2011), was administered at pretest, posttest, and retention test. Two parallel forms were used: Form 1 at pretest and Form 2 at posttest and retention test. The assessment included procedural and conceptual knowledge scales (see Table 1 for example items). The procedural knowledge scale (eight items) assessed children's use of correct strategies to solve mathematical equivalence problems (correct strategies and correct answers differed on less than 1% of all trials). Six of the items contained operations on both sides of the equal sign, one item was a simpler problem with a single operation

Table 1
*Example Items From the Procedural and Conceptual Knowledge Scales on the Mathematical Equivalence Assessment*

| Item type | Task | Scoring criteria |
|---|---|---|
| Procedural | | $\alpha = .83$ in Experiment 1; $\alpha = .85$ in Experiment 2 |
| Familiar problems | Solve 1 problem with operation on right side ($8 = 6 + \square$) | Use correct strategy (if strategy is ambiguous, response must be within 1 of correct answer) |
| | Solve 3 problems with operations on both sides, blank on right (e.g., $3 + 4 = \square + 5$) | Same as above |
| Transfer problems | Solve 3 problems with operations on both sides, blank on left or includes subtraction (e.g., $\square + 6 = 8 + 6 + 5$) | Same as above |
| | Solve 1 equation with an unknown variable ($y + 4 = 8 + 2$) | Same as above |
| Conceptual | | $\alpha = .64$ in Experiment 1; $\alpha = .71$ in Experiment 2 |
| Meaning of equal sign | Define equal sign | 1 point for relational definition (e.g., the same amount) |
| | Rate definitions of equal sign as good, not good, or don't know | 1 point for rating "two amounts are the same" as a good definition |
| Structure of equations | Reproduce 3 equivalence problems from memory | 1 point for correctly reconstructing all 3 problems |
| | Indicate whether 5 equations such as $3 = 3$ are true or false | 1 point for correctly recognizing 4 or more equations as true or false |

*Note.* Cronbach's alphas are for posttest. Alphas were somewhat lower at pretest, largely due to floor effects on some items.

on the right side of the equal sign, and one item was a more challenging problem with an unknown variable (i.e., *y*). The conceptual knowledge scale (eight items) assessed two concepts: (a) the meaning of the equal sign, and (b) the structure of equations. A brief version of Form 1 (five items—three conceptual and two procedural items) was used as a midtest during the intervention. The more difficult items were included on the midtest, as they were similar in difficulty to the problems presented during intervention.

We coded the conceptual knowledge items requiring a written explanation (see Table 1 for coding criteria). We established interrater reliability by having a second rater code the written responses of 20% of the children. Interrater agreement was high (exact agreement = 95%–97%; $\kappa$s = 90%–95%). Kappas calculate interrater reliability adjusted for chance (Cohen, 1960). Values above 81% are considered excellent (Landis & Koch, 1977). We also coded children's problem-solving strategies on the procedural knowledge assessment and on the intervention problems (Table 2). On the assessment, strategies were inferred from children's written work. For the intervention, strategies were based on children's verbal reports. Interrater agreement was high (exact agreement = 88%; $\kappa$ = 86%). Although specific strategies were coded on the procedural knowledge assessment, scores were based solely on whether a strategy was correct or incorrect. Interrater agreement on correct strategy versus incorrect strategy use was high (exact agreement = 99%; $\kappa$ = 98%).

**Data analysis.** We used a contrast analysis of variance (ANOVA) model (West, Aiken, & Krull, 1996). In this model, contrast codes represent a categorical variable with more than two levels, which in this case was condition. The condition variable had three groups (no feedback, outcome feedback, strategy feedback), so two coded variables were created. The primary goal was to determine whether any guidance during exploration prior to instruction was beneficial; thus, the first variable (feedback) compared no feedback to the two feedback conditions combined. We also explored whether the type of guidance mattered. Thus, the second variable (feedback type) compared outcome feedback to strategy feedback. Three covariates were also included (children's

Table 2
*Strategies Used to Solve Mathematical Equivalence Problems*

| Strategy | Sample explanation (4 + 5 + 8 = __ + 8) |
|---|---|
| Correct strategies | |
| Equalize[a] | I added 4, 5, and 8 and got 17, and 9 plus 8 is also 17. |
| Add–subtract[a] | I added 4, 5, and 8 and got 17, and 17 minus 8 is 9. |
| Grouping[a] | I took out the 8s, and I added 4 plus 5. |
| Ambiguous | 8 divided by 8 is 0, and 4 plus 5 is 9. |
| Incorrect strategies | |
| Add all[a] | I added the 4, 5, 8, and 8. |
| Add-to-equal[a] | 4 plus 5 is 9, and 9 plus 8 is 17. |
| Add-two | I added the 5 and the 8. |
| Carry[a] | I saw a 4 here, so I wrote a 4 in the blank. |
| Ambiguous | I used 8 plus 8 and then 5. |

[a] Entries represent the strategies demonstrated in the strategy evaluation task presented in Experiment 2.

age and procedural and conceptual knowledge pretest scores). Preliminary analyses indicated that gender was not a significant predictor of the outcome measures so it was not included in the final model. Finally, to evaluate whether condition effects depended on prior knowledge, we included two interaction terms: feedback by prior knowledge and feedback type by prior knowledge. Procedural knowledge at pretest was used as the prior knowledge measure as it is the most relevant type of prior knowledge for learning during problem solving. Preliminary analyses indicated that conceptual knowledge at pretest did not interact with condition. Thus, our statistical model was a contrast-based analysis of covariance with two contrast-coded between-subject variables (feedback, feedback type), three covariates, and two prior knowledge interaction terms. All effects reported with this model are similar if we replace the two contrast-coded variables with a two-degree of freedom "condition" variable followed by post hoc tests.

The assumption of homogeneity of variance for this model was largely supported. For the procedural knowledge variables, Levene's tests indicated equal variances on the posttest and retention test, $F$s $< 1$, though not at midtest, $F(2, 84) = 6.30$, $p = .003$ (likely due to limited number of midtest items). With all three time points in the same (repeated-measures) model, covariance matrices were also homogeneous, Box's M $= 8.01$, $F(12, 30952) = 0.63$, $p = .82$. For the conceptual knowledge variables, Levene's tests indicated equal variances at midtest, posttest, and retention test, $F$s $< 2$. With all three time points in the same (repeated-measures) model, covariance matrices were also homogeneous, Box's M $=$ 7.78, $F(12, 30952) = 0.61$, $p = .83$. Overall, ANOVA models were appropriate for analyzing the data.

## Results

**Pretest.** On the pretest, children answered few procedural ($M$ = 29%, $SD$ = 22%) and conceptual ($M$ = 27%, $SD$ = 19%) items correctly. Children's prior procedural knowledge was of particular interest. As shown in Figure 1, about half of the children were getting two or fewer items correct (out of 8). Most of those children were getting the one simple item correct, while a few were getting two items correct, indicating that they could solve the simple problem as well as one of the more difficult problems. The remaining half of the children were getting three or more items correct, indicating that they could solve multiple difficult problems correctly. Thus, about half of our sample entered with little to no knowledge of how to solve problems with operations on both sides of the equal sign, which were targeted in the intervention. There were no significant differences between conditions on either scale at pretest, $F$s $< 1$.

**Primary outcomes.** To evaluate children's performance on the midtest, posttest and retention test we conducted repeated-measures analyses of covariance (ANCOVAs) with feedback (feedback vs. none) and feedback type (outcome vs. strategy) as between-subject variables and time (midtest, posttest, retention test) as the within-subject variable. The three covariates and two interaction terms were included. Conclusions remain unchanged when the midtest is removed from the model. Procedural knowledge and conceptual knowledge were examined as separate outcomes. Feedback was expected to lead to higher procedural knowledge than no feedback, particularly for children with lower prior knowledge. We explored whether strategy feedback led to higher scores than outcome feedback and whether this effect interacted with prior knowledge. The effect of feedback on children's conceptual knowledge was also examined, though we had no prior predictions.

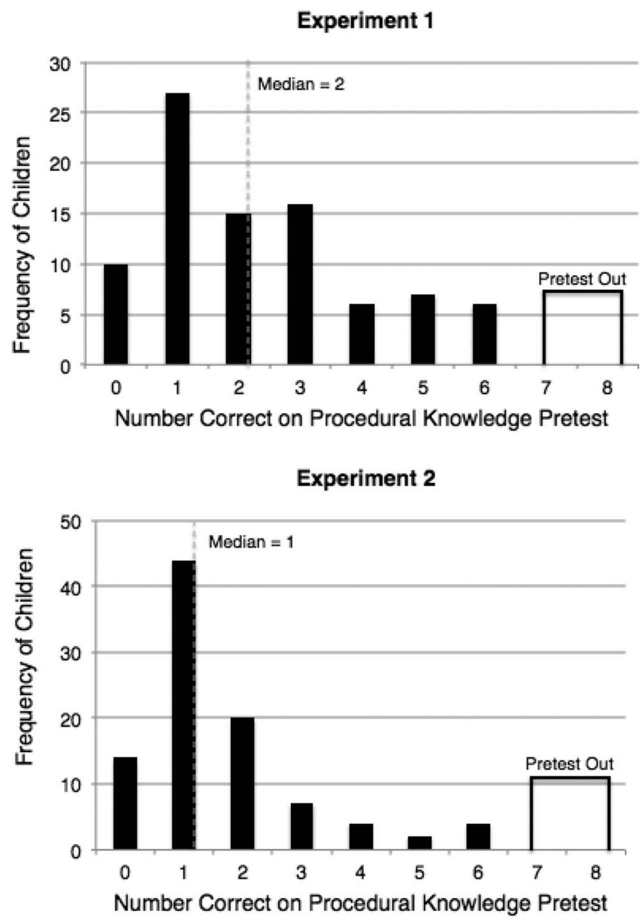


*Figure 1.* Frequency of children's scores on procedural knowledge measure at pretest.

**Procedural knowledge.** Children's procedural knowledge increased from midtest to posttest and remained similar 2 weeks later (see Table 3), $F(2, 158) = 13.64$, $p < .001$, $\eta_p^2 = .15$. There were no main effects of feedback or feedback type, $F$s $< 1$. However, there was a feedback by prior knowledge interaction, $F(1, 79) = 5.70$, $p = .02$, $\eta_p^2 = .07$. Consistent with our predictions, as prior procedural knowledge increased, the benefits of feedback decreased ($B = -1.04$, standard error [$SE$] $= 0.43$). Feedback type did not interact with prior knowledge ($p = .44$).[1]

[1] Procedural knowledge results remain unchanged when the midtest is removed from the model. In the model without the midtest, there were no main effects of feedback or feedback type, nor did feedback type interact with prior knowledge, $F$s $< 2$. Consistent with the full model, there was a feedback by prior knowledge interaction, $F(1, 79) = 3.92$, $p = .051$, $\eta_p^2 =$ .05. As prior knowledge increased, the benefits of feedback decreased ($B =$ $-0.77$, $SE = 0.39$).

Table 3
*Percentage of Correct Answers on Procedural Knowledge Measure in Experiments 1 and 2 by Condition and Prior Knowledge*

| Time | Prior knowledge | No feedback | | Outcome feedback | | Strategy feedback | |
|---|---|---|---|---|---|---|---|
| | | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| Experiment 1 | | | | | | | |
| Pretest | Low | 15 | 9 | 13 | 9 | 13 | 9 |
| | Moderate | 51 | 15 | 54 | 16 | 48 | 14 |
| Midtest | Low | 0 | 0 | 22 | 31 | 23 | 32 |
| | Moderate | 58 | 42 | 50 | 46 | 40 | 46 |
| Posttest | Low | 26 | 31 | 39 | 34 | 32 | 27 |
| | Moderate | 71 | 25 | 61 | 32 | 50 | 27 |
| Retention test | Low | 29 | 32 | 33 | 32 | 33 | 36 |
| | Moderate | 81 | 22 | 71 | 27 | 41 | 32 |
| Experiment 2 | | | | | | | |
| Pretest | Low | 8 | 6 | 9 | 6 | 11 | 5 |
| | Moderate | 38 | 19 | 39 | 19 | 32 | 10 |
| Midtest | Low | 8 | 19 | 23 | 26 | 21 | 34 |
| | Moderate | 43 | 46 | 29 | 32 | 40 | 38 |
| Posttest | Low | 24 | 31 | 38 | 38 | 40 | 28 |
| | Moderate | 49 | 41 | 32 | 23 | 30 | 27 |
| Retention test | Low | 24 | 24 | 31 | 29 | 29 | 31 |
| | Moderate | 54 | 36 | 33 | 28 | 29 | 29 |

*Note.* Children are categorized as having low or moderate prior knowledge on the basis of a median split on the procedural knowledge assessment at pretest; however, in the primary analysis models, prior knowledge was treated as a continuous variable.

To further interpret the interaction, we categorized the children into two groups on the basis of a median split on procedural knowledge at pretest, as described in the pretest section. Children who scored above the median got three or more problems correct on the pretest. Thus, they exhibited moderate prior knowledge of correct strategies, though they continued to use incorrect strategies as well ($n = 35$). Children who scored below the median got two or fewer problems correct on the pretest. Thus, they exhibited little prior knowledge of correct strategies ($n = 52$). We then examined the main effects of feedback for each group (see Figure 2). For the low prior knowledge group, children who received feedback tended to exhibit *higher* procedural knowledge ($M = 34\%$, $SE = 5\%$) than children who did not ($M = 20\%$, $SE = 6\%$), $F(1, 79) = 3.28$, $p = .07$, $\eta_p^2 = .04$. For the moderate prior knowledge group, children who received feedback tended to exhibit *lower* procedural knowledge ($M = 49\%$, $SE = 6\%$) than children who did not ($M = 66\%$, $SE = 8\%$), $F(1, 79) = 3.66$, $p = .06$, $\eta_p^2 = .04$. Overall, feedback during exploration was more beneficial than no feedback for children with low prior knowledge of correct strategies. For children with moderate prior knowledge of correct strategies, the reverse was true, with feedback actually hindering learning relative to no feedback.

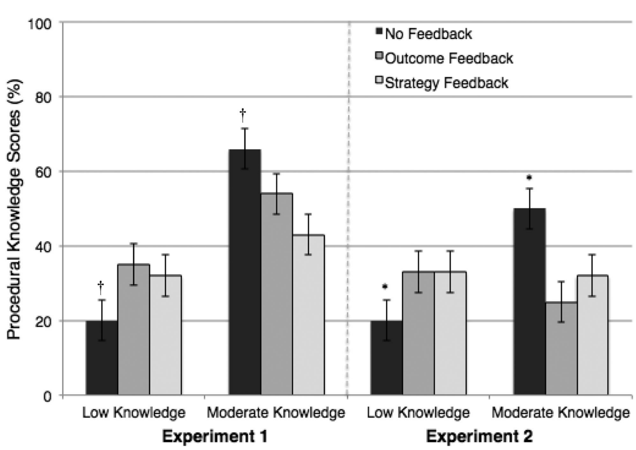


*Figure 2.* Percentage of correct answers on procedural knowledge measure by condition and prior knowledge. Scores are estimated marginal means based on midtest, posttest, and retention test scores combined. Differences are between the no-feedback condition and the two feedback conditions combined: * $p < .05$. † $p < .07$. Error bars represent standard errors.

**Conceptual knowledge.** Children's conceptual knowledge also changed over time. Scores increased from midtest ($M = 20\%$, $SE = 2\%$) to posttest ($M = 55\%$, $SE = 2\%$) and remained similar at retention test ($M = 51\%$, $SE = 3\%$), $F(2, 158) = 89.73$, $p < .001$, $\eta_p^2 = .53$. There were no effects related to feedback versus no feedback. There was a marginal effect of feedback type, $F(1, 79) = 3.56$, $p = .06$, $\eta_p^2 = .04$, which was qualified by a marginal feedback type by prior knowledge interaction, $F(1, 79) = 2.93$, $p = .09$, $\eta_p^2 = .04$. As prior knowledge increased, the benefits of outcome feedback increased relative to strategy feedback ($B = 0.67$, $SE = 0.39$).[2]

To further interpret this marginal interaction, we examined the effect of feedback type for children with low and moderate prior

[2] Conceptual knowledge results remain similar when the midtest is removed from the model. In the model without the midtest, there were no effects related to feedback versus no feedback, $F$s $< 1$, nor did feedback type interact with prior knowledge ($p = .14$). Consistent with the full model, there was a main effect of feedback type, $F(1, 79) = 4.29$, $p = .04$, $\eta_p^2 = .04$. Children who received outcome feedback exhibited higher conceptual knowledge ($M = 58\%$, $SE = 4\%$) than children who received strategy-feedback ($M = 47\%$, $SE = 4\%$).

knowledge separately (on the basis of a median split on procedural knowledge pretest scores, see Figure 3). For the low prior knowledge group, there were no differences between the types of feedback, $F(1, 79) = 0.34$, $p = .56$. For the moderate prior knowledge group, children who received outcome feedback had higher conceptual knowledge ($M = 55\%$, $SE = 5\%$) than children who received strategy feedback ($M = 38\%$, $SE = 6\%$), $F(1, 79) = 5.07$, $p = .03$. These results suggest that outcome feedback, strategy feedback, and no feedback promoted similar levels of conceptual knowledge for children with low prior knowledge; but outcome feedback promoted greater conceptual knowledge than strategy feedback for children with moderate prior knowledge of correct strategies. Note that outcome feedback was not more effective than no feedback for children with moderate knowledge ($p = .56$).

**Intervention activities.** To better understand *how* exploration impacted learning, we explored children's responses during the intervention. Recall, children verbally reported the strategies they used to solve the problems during the intervention. We were interested in how feedback influenced children's strategy use. Past work indicates that the primary function of feedback is to identify errors and encourage the search for plausible alternatives (e.g., Kulhavy, 1977). Thus, promoting the use of different strategies might be one mechanism by which feedback influences exploration. Overall, children used a variety of correct and incorrect solution strategies on the 12 practice problems ($M = 2.9$, $SD = 1.3$).

**Correct strategy use.** Children used zero, one, or two different *correct* strategies ($M = 0.8$, $SD = 0.8$). As shown in Table 4, a qualitative examination indicates that feedback tended to promote the use of more different correct strategies than no feedback. Because of the categorical nature of the data, we used chi-square analyses to examine differences between the feedback and no-feedback conditions. Specifically, the number of children who used zero correct strategies or one or more correct strategies was examined. More children used at least one correct strategy in the feedback conditions (63%) compared with in the no-feedback condition (45%), though this effect was not significant, $\chi^2(1, N = 87) = 2.44$, $p = .12$. This effect was stronger for children with low prior knowledge. For the low prior knowledge group, the number of children generating one or more different correct strategies was significantly higher in the feedback conditions (55%) than in the no-feedback condition (21%), $\chi^2(1, N = 52) = 5.54$, $p = .02$. For the moderate prior knowledge group, the number of children using one or more different correct strategies did not differ significantly in the feedback (74%) and no-feedback conditions (83%), $\chi^2(1, N = 35) = 0.40$, $p = .53$. The pattern of results was relatively similar for both feedback types.

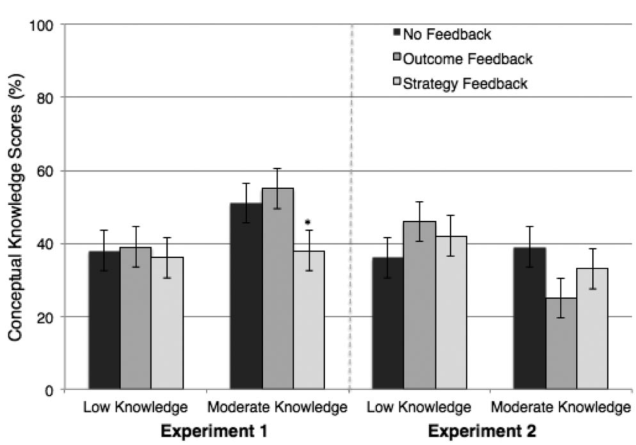


*Figure 3.* Percentage of correct answers on conceptual knowledge measure by condition and prior knowledge. Scores are estimated marginal means based on midtest, posttest, and retention test scores combined. Significant difference is between the outcome feedback and strategy feedback conditions: $^* p < .05$. Error bars represent standard errors.

Table 4
*Proportion of Children Using Zero, One, or Two Different Correct Strategies on the Intervention Problems by Condition and Prior Knowledge in Experiment 1*

| | Feedback type | | |
|---|---|---|---|
| Knowledge/strategies | None | Outcome | Strategy |
| Low prior knowledge | | | |
| No. of correct strategies | | | |
| 0 | .79 | .44 | .47 |
| 1 | .16 | .33 | .47 |
| 2 | .05 | .22 | .07 |
| No. of children | 19 | 18 | 15 |
| Moderate prior knowledge | | | |
| No. of correct strategies | | | |
| 0 | .17 | .23 | .30 |
| 1 | .58 | .31 | .50 |
| 2 | .25 | .46 | .20 |
| No. of children | 12 | 13 | 10 |

**Incorrect strategy use.** The number of different *incorrect* strategies used ranged from 0 to 5 ($M = 2.1$, $SD = 1.3$). To examine differences across conditions, we used our primary ANCOVA model with feedback and feedback type as between-subject variables and number of different incorrect strategies used as the dependent variable. There was a main effect of feedback, $F(1, 79) = 11.22$, $p = .001$, $\eta_p^2 = .12$, but no other effects were significant. Children who received feedback used a greater number of different incorrect strategies ($M = 2.4$, $SE = 0.2$) than children who did not ($M = 1.6$, $SE = 0.2$). The effect was similar in strength for all children. For the low prior knowledge group, children who received feedback used roughly one more incorrect strategy ($M = 2.8$, $SE = 0.2$) than children who did not ($M = 1.9$, $SE = 0.3$). Similarly, for the moderate prior knowledge group, children who received feedback used roughly one more incorrect strategy ($M = 1.9$, $SE = .3$) than children who did not ($M = 1.0$, $SE = 0.3$). The pattern of results was the same for both types of feedback.

There were also differences in perseveration—using the same incorrect strategy on all of the problems. More children perseverated in the no-feedback condition (23%) than in the strategy-feedback (8%) or outcome-feedback (0%) conditions, $\chi^2(2, N = 87) = 8.73$, $p = .01$. Moreover, the effect was more pronounced for children with low prior knowledge. For low prior knowledge children, more children perseverated in the no-feedback condition (32%) than in the strategy-feedback (13%) or outcome-feedback (0%) conditions, $\chi^2(1, N = 54) = 6.53$, $p = .01$. For children with moderate prior knowledge, few children perseverated at all, and there were no significant differences among conditions (8% in no-feedback condition, 0% in strategy- and outcome-feedback conditions), $\chi^2(1, N = 37) = 2.14$, $p = .14$. Overall, children with low prior knowledge used more correct and incorrect strategies if they received feedback and tended to perseverate on the same incorrect strategy if they did not. For moderate-knowledge children, feedback promoted the use of incorrect strategies relative to no feedback, but did not have a similar effect on correct strategies, which may explain why feedback hindered their performance.

## Discussion

In Experiment 1, our primary hypothesis was supported. Feedback during exploration led to higher procedural knowledge than no feedback, but only for children with relatively low prior knowledge of correct strategies. For children with moderate prior knowledge, feedback led to *lower* procedural knowledge relative to no feedback. Our secondary analyses indicated that for children with low prior knowledge, feedback promoted the generation of both correct and incorrect strategies and prevented perseveration relative to no feedback. For children with moderate prior knowledge, feedback promoted the use of incorrect strategies relative to no feedback. Overall, the benefits of providing feedback during exploration prior to instruction depend on prior knowledge. Children with low knowledge of domain-specific strategies benefited from receiving feedback, whereas children with moderate prior knowledge benefited more from exploring without it. Feedback type had little effect, with the exception that outcome feedback tended to lead to higher conceptual knowledge than strategy feedback (but not the no-feedback condition) for children with moderate prior knowledge.

Although we predicted that prior knowledge would moderate the impact of condition on procedural knowledge learning, we did not have a prior reason to expect a reversal, such that feedback would actually harm learning for children with moderate prior knowledge. In addition, several limitations in the design of Experiment 1 constrain the generalization of the findings. First, the condition manipulation was not as clean or as strong as it could have been. For example, all children were asked to report how they solved each problem. Though this resulted in detailed information regarding the strategies used, it inevitably guided all children's attention to some degree to their problem-solving strategies. The strategy-feedback manipulation would be stronger if only children in the strategy-feedback condition were encouraged to attend to their strategy use. Additionally, the feedback provided in both feedback conditions was relatively vague and not specific to the child's response. For example, in the strategy-feedback condition, incorrect strategies were referred to as "not a correct way," which may have been unclear to children. Further, children in both the strategy-feedback and outcome-feedback conditions were told if their target response (strategy or answer, respectively) was correct, but only children in the outcome-feedback were given additional correct information (i.e., the correct answer). The contrast between the two feedback conditions could be improved.

Finally, we also sought to clarify the influences of feedback type during exploration prior to instruction. Given the paucity of research comparing outcome-feedback to strategy-feedback and the slight variation in means for children with moderate prior knowledge in these two conditions, we wanted to confirm that feedback type is not central to children's learning during exploration. To address these concerns, we conducted a second experiment similar to Experiment 1, but with several modifications intended to strengthen the design.

# Experiment 2

Experiment 2 was designed to strengthen the condition manipulation from Experiment 1 and verify the results with an independent sample of children. Our goal was to replicate the finding that children with low prior knowledge benefit from feedback during exploration prior to instruction, whereas children with moderate prior knowledge benefit from no feedback. We strengthened the condition manipulation in three ways. First, to differentiate the conditions, we only had children in the strategy-feedback condition report how they solved each problem. Children in the other conditions were asked to report different information to mimic the interaction with the experimenter (i.e., their answer in the outcome-feedback condition and their completion of the problem in the no-feedback condition). Second, we made the feedback more specific by revoicing the child's response. In the strategy-feedback condition, we restated the child's strategy, and in the outcome-feedback condition, we restated the child's answer. Finally, we did not give the correct answer in the outcome-feedback condition. In Experiment 1, only children in the outcome-feedback condition received additional information (i.e., the correct answer). An alternative solution was to provide children in the strategy-feedback condition with additional information (i.e., a correct strategy). However, telling children how to solve a problem is a form of explicit instruction, and we were interested in the guidance provided *prior* to explicit instruction. So we eliminated the correct answer in the outcome-feedback condition to enhance parallelism across conditions. Consistent with Experiment 1, we expected low-knowledge children to benefit more from feedback relative to no feedback and moderate-knowledge children to benefit more from no feedback, regardless of feedback type.

## Method

**Participants.** Consent was obtained from 111 second- and third-grade children at two schools (one public, one parochial) in middle Tennessee. Of those children, 101 met criteria for participation because they scored at or below 80% on both a conceptual and procedural knowledge measure at pretest. Six additional children had their data excluded: two for failing to complete the intervention and four for missing the retention test. The final sample consisted of 95 children (*M* age = 7 years 11 months; 60 girls; 35 boys; 97% African American; 3% White). Approximately 61% received free or reduced-price lunch. Teacher reports indicated that three of these children were receiving special education services; however, their performance did not differ from the norm so their data was included in all final analyses.

**Design.** The design and procedure were identical to those in Experiment 1 with a few exceptions outlined in this section. As before, children were randomly assigned to one of three conditions: strategy-feedback ($n = 31$), outcome-feedback ($n = 33$), or no-feedback ($n = 31$) condition. The conditions were well matched in terms of background characteristics. There were no differences among children in the three conditions in terms of age, $F(2, 92) = 0.15$, $p = .86$; grade, $\chi^2(2, N = 95) = 0.08$, $p = .96$;

ethnicity, $\chi^2(2, N = 95) = 2.11$, $p = .35$; special education status, $\chi^2(2, N = 95) = 2.11$, $p = .35$; or free or reduced-price lunch status, $\chi^2(4, N = 95) = 4.79$, $p = .31$. Despite random assignment, there was a difference in conditions in terms of gender, $\chi^2$ $(2, N = 95) = 9.06$, $p = .01$. The strategy-feedback condition had fewer boys (16%) than the outcome-feedback (42%) or no-feedback (52%) conditions. Because of this difference, we explored whether gender was a significant predictor of the outcome measures (see "Data analysis" section, which will follow).

**Procedure.** Consistent with Experiment 1, all children received a tutoring session that began with exploratory problem solving followed by brief conceptual instruction. The 12 mathematical equivalence problems from Experiment 1 were used, but they were presented in paper-and-pencil format rather than on a computer screen to simulate a more typical classroom activity. The computer program was still used by the experimenter to record information.

In the *strategy-feedback condition*, children reported how they solved each problem and received feedback on the strategy, which included a revoicing of the strategy report (e.g., "Good job! That is one correct way to solve that problem—[Child's strategy] is a correct way to solve it"; "Good try, but that is not a correct way to solve the problem—[Child's strategy] is not a correct way to solve it"). For example, if a child reported using the add-all strategy (see Table 2), the experimenter repeated the child's report: "Good try, but that is not a correct way to solve the problem. Adding all the numbers together is not a correct way to solve this problem." The experimenter revoiced the strategy just as the child stated it to ensure no additional information was given. If the child was unsure of the strategy used, the experimenter stated: "It is not clear if you used a correct way to solve this problem. Let's try another one. This time, try to remember how you solved the problem," though this occurred on only 2% of trials. In the *outcome-feedback condition*, children reported their numerical answer and received feedback on that answer, which included a revoicing of their answer but not the correct answer (e.g., "Good job! You got the right answer; [child's answer] is the correct answer"; "Good try, but you did not get the right answer; [child's answer] is not the correct answer"). Finally, in the *no-feedback condition*, children reported when they completed a problem and were told to go on.

Again, children completed a brief midtest between the problem solving and instruction. Also, they rated their subjective cognitive load, using two modified items from the NASA Task Load Index as well as a third item adapted from past cognitive load studies with adolescents (e.g., Kalyuga et al., 2004). Given concerns about the validity of the cognitive load measure for use with children, results for this measure are not reported.

**Assessments and coding.** The mathematical equivalence assessment, modified slightly from Experiment 1 to improve psychometric properties, was administered at pretest, posttest, and retention test. Again, a brief version (five items—three conceptual and two procedural items) was used as a midtest during the intervention. We established interrater reliability by having a second rater code the subjective responses of 30% of the children (see Table 1 for coding criteria). Interrater agreement was high for written explanations (exact agreement = 93%–99%, $\kappa$s = 87%–98%) and for strategy codes (exact agreement = 91%, $\kappa$ = 89%).

**Strategy evaluation.** A strategy evaluation task was administered after the posttest and after the retention test to assess children's ratings of correct and incorrect strategies for solving mathematical equivalence problems (Rittle-Johnson & Alibali, 1999). Children were told that students from another school had solved these problems. They were presented with examples of the strategies used and asked to evaluate each strategy as "very smart, kind of smart, or not so smart" (see Table 2 for the strategies demonstrated). However, preliminary analyses indicated no systematic differences among the conditions; thus, we do not report results for this task.

**Data analysis.** The same ANCOVA model from Experiment 1 was employed. We used a contrast-based ANCOVA with two contrast-coded between-subject variables (feedback, feedback type), three covariates (children's age, procedural and conceptual knowledge pretest scores), and two condition by prior knowledge interaction terms (with procedural knowledge pretest scores as the prior knowledge measures). As in Experiment 1, preliminary analyses indicated that gender was not a significant predictor of the outcome measures and that conceptual knowledge pretest scores did not interact with condition. The assumption of homogeneity of variance for this model was largely supported. For the procedural knowledge variables, Levene's tests indicated equal variances at midtest, posttest, and retention test, $F$s $< 2.2$. With all three time points in the same model, the variance–covariance matrices were also homogeneous, Box's M = 20.09, $F(12, 40695) = 1.60$, $p = .09$. For the conceptual knowledge variables, Levene's tests indicated equal variances at midtest and posttest, $F$s $< 2$, though not at the retention test, $F(2, 92) = 4.66$, $p = .01$. With all three time points in the same model, the variance–covariance matrices were also homogeneous, Box's M = 13.68, $F(12, 40694) = 1.09$, $p = .37$. Overall, ANOVA models were appropriate for analyzing the data.

## Results

**Pretest.** On the pretest, children answered few procedural ($M$ = 20%, $SD$ = 18%) and conceptual ($M$ = 19%, $SD$ = 18%) items correctly. Overall, the scores were somewhat lower compared with those in Experiment 1. Children's prior procedural knowledge was of particular interest. As shown in Figure 1, about half of the children were getting one or fewer items correct (out of 8). Most of those children were getting one item correct, and that one item was almost always the simple problem (see Table 1). The remaining half of the children were getting two or more items correct, indicating that they could solve at least one of the more difficult problems correctly. Thus, about half of our sample entered with little to no knowledge of how to solve problems with operations on both sides, which were targeted in the intervention. There were no significant differences between conditions on either scale at pretest, $F$s $< 1$.

**Primary outcomes.** To evaluate performance on the midtest, posttest and retention test, we conducted repeated-measures ANCOVAs with feedback (feedback vs none) and feedback type (outcome vs strategy) as between-subject variables and time (midtest, posttest, retention test) as the within-subject variable. The three covariates and two interaction terms were included. Statistical conclusions remain unchanged when the midtest is removed from the model. In Experiment 2, only results for primary outcomes are reported. Unlike Experiment 1, detailed strategy reports were only available for children in the strategy-feedback condition; thus, an analysis of children's strategy use during exploration was not performed.

**Procedural knowledge.** Children's procedural knowledge scores increased from midtest to posttest and remained similar 2 weeks later (see Table 3), $F(2, 174) = 3.77$, $p = .03$, $\eta_p^2 = .04$. There were no main effects of feedback or feedback type, nor did feedback type interact with prior knowledge, $F$s $< 1$. However, consistent with Experiment 1, there was a feedback by prior knowledge interaction, $F(1, 87) = 4.67$, $p = .03$, $\eta_p^2 = .05$. As prior procedural knowledge increased, the benefits of feedback decreased ($B = -1.06$, $SE = 0.49$).[3]

To further interpret this interaction, we used a median split to categorize children into two groups on the basis of their procedural knowledge at pretest. Children who scored above the median got two or more problems correct on the pretest. Thus, they exhibited moderate prior knowledge of correct strategies, though they continued to use incorrect strategies as well ($n = 58$). Children who scored below the median got one or no problems correct on the pretest. Thus, they exhibited little prior knowledge of correct strategies ($n = 37$). We then examined the main effects of feedback for each group (see Figure 2). For the low prior knowledge group, children who received feedback exhibited significantly higher procedural knowledge ($M = 33\%$, $SE = 4\%$) than children who did not ($M = 20\%$, $SE = 5\%$), $F(1, 87) = 4.00$, $p = .05$, $\eta_p^2 = .04$. For the moderate prior knowledge group, children who received feedback exhibited significantly *lower* procedural knowledge ($M = 28\%$, $SE = 5\%$) than children who did not ($M = 50\%$, $SE = 6\%$), $F(1, 87) = 7.54$, $p = .007$, $\eta_p^2 = .08$. The results replicated the effect found in Experiment 1. Feedback was more beneficial than no feedback for children with low prior knowledge of correct strategies, but for children with moderate knowledge of correct strategies, the reverse was true.

**Conceptual knowledge.** Children's conceptual knowledge scores also increased from midtest ($M = 21\%$, $SE = 2\%$) to posttest ($M = 50\%$, $SE = 2\%$) and stayed similar at retention test ($M = 43\%$, $SE = 2\%$), $F(2, 174) = 67.13$, $p < .001$, $\eta_p^2 = .44$. There were no main effects of feedback or feedback type, nor did feedback type interact with prior knowledge, $F$s $< 1$. There was a marginal feedback by prior knowledge interaction, $F(1, 87) = 3.63$, $p = .06$, $\eta_p^2 = .04$. As prior procedural knowledge increased, the benefits of feedback marginally decreased ($B = -0.70$, $SE = 0.37$). Feedback also interacted with time, $F(2, 174) = 7.14$, $p = .001$, $\eta_p^2 = .08$, such that the benefits of feedback were stronger at the midtest and decreased over time.[4]

To further interpret the marginal interaction, we examined the effect of feedback for low and moderate prior knowledge children separately (on the basis of a median split on procedural knowledge pretest scores; see Figure 3). For the low prior knowledge group, children who received feedback exhibited slightly higher conceptual knowledge ($M = 44\%$, $SE = 3\%$) than children who did not ($M = 37\%$, $SE = 4\%$), $F(1, 87) = 2.56$, $p = .11$, $\eta_p^2 = .03$. For the moderate prior knowledge group, children who received feedback exhibited slightly lower conceptual knowledge ($M = 29\%$, $SE = 3\%$) than children who did not ($M = 39\%$, $SE = 5\%$), $F(1, 87) = 2.60$, $p = .11$, $\eta_p^2 = .03$. Although not reliable, these conceptual knowledge results resemble the pattern of findings found for procedural knowledge across both experiments.

## Discussion

Results from Experiment 2 were consistent with Experiment 1 and supported our primary hypothesis. The effect of feedback prior to instruction was moderated by prior knowledge. For children with relatively low prior knowledge of correct strategies, feedback during exploratory problem solving led to higher procedural knowledge than no feedback. But for children with moderate prior knowledge of correct strategies, feedback led to *lower* procedural knowledge than no feedback. There was a similar, yet weaker effect for children's conceptual knowledge. Feedback type had little effect in general, providing evidence that both types of feedback hinder performance for children with moderate prior knowledge, but help performance for children with low prior knowledge. Overall, Experiment 2 replicated the findings from Experiment 1 with an independent sample of children and supported our primary conclusions.

# General Discussion

Guided discovery facilitates deeper learning than discovery or instruction alone (e.g., Alfieri et al., 2011). For example, providing exploratory activities with subsequent instruction can be beneficial (e.g., DeCaro & Rittle-Johnson, 2011; Schwartz et al., 2011). However, the amount of guidance provided *during* the exploration has largely gone unstudied. We examined the effects of feedback during exploratory problem solving for children with various levels of prior knowledge. In two experiments, children solved unfamiliar mathematical equivalence problems and then received conceptual instruction. During problem solving, some children received feedback (on their answer or on their strategy), whereas others did not. In both experiments, prior knowledge moderated the impact of feedback on procedural knowledge. For children with low prior knowledge of domain-specific strategies, feedback led to higher procedural knowledge than no feedback. But for children with moderate prior knowledge of correct strategies, feedback led to lower procedural knowledge than no feedback. Effects on conceptual knowledge were weak; suggesting feedback during exploration primarily impacts procedural knowledge. Feedback type (outcome vs. strategy) had little effect in general. We discuss these findings in light of past research on the effects of prior knowledge and outline potential explanatory mechanisms. Finally, we consider the implications for guided discovery learning as well as future inquiries.

[3] Procedural knowledge results remain unchanged when the midtest is removed from the model. In the model without the midtest, there were no main effects of feedback or feedback type, nor did feedback type interact with prior knowledge, $F$s $< 1$. Consistent with the full model, there was a feedback by prior knowledge interaction, $F(1, 87) = 4.25$, $p = .04$, $\eta_p^2 = .05$. As prior knowledge increased, the benefits of feedback decreased ($B = -0.92$, $SE = 0.45$).

[4] Conceptual knowledge results remain unchanged when the midtest is removed from the model. In the model without the midtest, there were no main effects of feedback or feedback type, nor did feedback type interact with prior knowledge, $F$s $< 1$. Consistent with the full model, there was a marginal feedback by prior knowledge interaction, $F(1, 87) = 2.97$, $p = .09$, $\eta_p^2 = .03$. As prior knowledge increased, the benefits of feedback decreased ($B = -0.61$, $SE = 0.36$).

## The Central Place of Prior Knowledge

The current study addresses a call in the feedback literature to attend to individual learner characteristics (Hattie & Gan, 2011) and is consistent with the notion that prior knowledge plays a central role in learning and instruction (see Tobias, 2009). Based on previous feedback research (e.g., Krause et al., 2009; Luwel et al., 2011), we expected feedback to be especially effective for learners with low prior knowledge and to have little effect for learners with higher prior knowledge. However, we did not anticipate that feedback during exploration prior to instruction would actually *reduce* learning for children with higher prior knowledge relative to no feedback. This is an unexpected and important finding in the research on feedback, particularly given that children who were harmed by feedback knew at least one correct strategy but used incorrect strategies as well. Thus, learners with only moderate levels of prior knowledge benefit from exploration without feedback, whereas the opposite is true for those with low prior knowledge.

This finding supports past research demonstrating that a single instructional method is often not best for learners with varying levels of prior knowledge. Indeed, a number of findings, termed *expertise reversal effects*, indicate that instructional techniques that are effective for novices lose their benefits for learners with more knowledge in the target domain (see Kalyuga, 2007 and Kalyuga, Ayres, Chandler, & Sweller, 2003, for reviews). In particular, a common conclusion is that "instructional guidance, which may be essential for novices, may have negative consequences for more experienced learners" (Kalyuga et al., 2003, p. 24). For example, we know that low-knowledge, but not high-knowledge, learners benefit more from (a) studying a given solution rather than imagining it (Leahy & Sweller, 2005), (b) seeing worked examples rather than solving open-ended problems (Kalyuga & Sweller, 2004), and (c) having multiple pieces of information presented together rather than in isolation (Kalyuga, Chandler, & Sweller, 1998).

The present results also contribute to a growing body of literature indicating that prior knowledge of domain-specific problem-solving strategies is a key individual difference (e.g., Alibali, 1999; Hofer et al., 2011; Kalyuga & Sweller, 2004; Luwel et al., 2011; McNeil & Alibali, 2005; Rittle-Johnson, Star, & Durkin, 2009; Siegler & Crowley, 1991). For example, Rittle-Johnson, Star, and Durkin (2009) found that middle-school students with no knowledge of algebraic strategies at pretest benefited most from studying examples one at a time or comparing different problem types. However, students who attempted algebraic strategies at pretest benefited most from comparing different solution strategies. Further, Kalyuga and Sweller (2004) successfully developed a rapid knowledge assessment to identify whether students should learn equation solving by studying worked examples or by solving problems unaided. In the assessment, students were asked to identify the next step in a problem solution, and students who skipped intermediate steps were awarded higher scores. Thus, in both cases, prior knowledge of domain-specific strategies was an important predictor of learning outcomes. The current study extends this research to the use of feedback during exploratory problem solving. Children with low prior knowledge of domain-specific strategies need feedback to improve their procedural knowledge. On the other hand, children who already know a correct strategy, even if they use it infrequently, do not need feedback and actually perform better without it.

## Potential Explanatory Mechanisms

Although our evidence indicates that the effects of feedback during exploration depend on prior knowledge, the mechanisms underlying this effect still need to be identified. Why do children with little prior knowledge of correct strategies benefit from feedback, whereas children with moderate knowledge of correct strategies benefit more from exploring without it? Our current data point to the important role feedback can play in strategy generation and selection.

One of the primary roles of feedback in non-problem-solving domains is to help learners identify errors and search for more plausible alternatives (see Mory, 2004). The same was true for children learning to solve unfamiliar mathematics problems. Children who received feedback exhibited greater strategy variability than those who did not, and this was particularly true for the low-knowledge group. Several specific changes in strategy use highlight potential explanatory mechanisms. First, feedback prevented low prior knowledge children from perseverating on the same incorrect strategy, which supports the idea that the main function of feedback is to identify initial errors (Mory, 2004). Second, for children with low prior knowledge, feedback promoted the use of a greater variety of both correct and incorrect strategies relative to no feedback, and past work indicates that cognitive variability is an impetus for cognitive change (e.g., Siegler, 1994; Siegler & Shipley, 1995). That is, thinking of or using a variety of strategies can help learners evaluate alternative approaches and be more receptive to future instruction (e.g., Alibali & Goldin-Meadow, 1993; Perry, Church, & Goldin-Meadow, 1988). Third, feedback facilitated low prior knowledge children's acquisition of at least one correct strategy. Past work indicates that children can and do discover correct strategies independently (e.g., Siegler & Crowley, 1991), but this can take weeks of extended practice. For learners with little to no knowledge of correct strategies, the constraining effects of feedback may have sped up the process of strategy acquisition (see also Alibali, 1999). Strategy acquisition may jumpstart subsequent strategy changes including the strengthening of that correct strategy on later assessments (Siegler, 1996). Together, the changes in strategy use may explain why children with low prior knowledge of correct strategies ultimately learned more when given feedback than when not.

Children's strategy use may also provide insight into why feedback hindered the performance of learners who knew a correct strategy, but used incorrect strategies as well. Recall, for the moderate-knowledge group, feedback promoted the use of a greater number of incorrect strategies relative to no feedback, but had a much weaker effect on correct strategies. Indeed, the number of moderate-knowledge children using one or two different correct strategies did not differ in the feedback and no-feedback conditions. One possibility is that the constraining effects of feedback, which helped the low-knowledge group acquire a correct strategy, was not necessary since these children already knew a correct strategy. Perseveration was also rare for children with moderate prior knowledge, so feedback had very limited impact on reducing children's reliance on the same incorrect strategy. Additionally, because feedback only facilitated the use of incorrect strategies, it may have had a negative impact on their procedural knowledge. If correct strategies compete against incorrect strategies for selection, increasing the number of incorrect strategies could reduce use of correct strategies (Siegler & Shipley, 1995).

Strategy use may also shed light more generally on how exploration prior to instruction impacts learning. Schwartz and Bransford (1998) suggested that exploratory activities facilitate the "development of differentiated knowledge" of the target problem space (p. 510). In past studies, exploration improved knowledge of the structure of the target problems (e.g., DeCaro & Rittle-Johnson, 2011; Schwartz et al., 2011) and the concepts underlying them (e.g., DeCaro & Rittle-Johnson, 2011). Consistent with this work, children in our study generated a wide variety of problem-solving strategies during the exploratory phase. Exploring a problem space may help learners acquire more nuanced knowledge, and thus prepare them to learn from subsequent instruction.

Changes in children's strategy use provide one explanation for the moderating effect of prior knowledge, but there are likely other mechanisms at work. For example, many prior knowledge interactions are explained in terms of the learner's experience of cognitive load (e.g., Kalyuga, 2007; Rey & Buchwald, 2011). For low-knowledge learners, novel tasks can easily overload working memory; thus, they often need external guidance to reduce cognitive load. In contrast, learners with higher prior knowledge can use their existing knowledge structures to complete the task without cognitive overload; thus, they often do not need external guidance. This may help explain why feedback was needed to facilitate learning in low-knowledge children but not in moderate-knowledge children. We measured subjective cognitive load in this study, but cognitive load scales have rarely been employed with children, and even in adults, the ability to provide accurate self-reports of mental effort has been questioned (Schnotz & Kurschner, 2007). Thus, difficulties in gathering empirical evidence in support of the cognitive load mechanism, particularly in children, make this account difficult to verify empirically. It is also possible that differences in motivation would help explain the findings. Children who are more knowledgeable may also be more motivated to learn. In turn, those who are more motivated may thrive in less structured, challenging environments, whereas children who are less motivated may not (e.g., Schnotz, 2010). More work is needed to tease apart these alternative explanations.

## Guided Discovery Learning

The present study also has important implications for research on guided discovery. It suggests that prior knowledge (and other individual differences) should be considered when evaluating guided discovery methods. Too often researchers consider individual differences "an annoyance . . . to be reduced by any possible device," rather than a source of relevant information (Cronbach, 1957, p. 674). In future research, learner characteristics should continue to be examined to assess the generalizability of guided discovery methods and how they can be optimized for certain subsets of learners. To be clear, our findings do not suggest that guided discovery in general is only beneficial for children with low knowledge. Indeed, all of our conditions were considered guided discovery as they contained exploration and instruction. All children seemed to benefit from the instruction, whereas only children with low prior knowledge of domain-specific strategies benefited from the feedback during exploration. One potential reason for this distinction is that the exploration focused on problem solving, whereas the instruction focused on concepts. Prior knowledge of domain-specific strategies may matter more for problem solving.

The results also highlight the need to evaluate different aspects of guided discovery. We examined the guidance provided *during* exploration prior to instruction and found that more was not always better. Unfortunately, even when researchers recognize the benefits of combining discovery and instruction, the usual suggestion is to include more guidance during exploration (e.g., Alfieri et al., 2011; Mayer, 2004). However, our results indicate that there is a time for just exploration. For moderate-knowledge children, combining unguided exploration with subsequent instruction improved procedural knowledge to a greater extent than exploration with feedback. Thus, optimizing learning does not always require an increase in guidance; sometimes it requires the removal of unnecessary information. This latter point is particularly important. If feedback had no effect for children with moderate prior knowledge, it might seem safe to provide feedback to all learners with the goal of helping those with low prior knowledge and doing no harm to those with some prior knowledge. However, the fact that feedback harms learning for moderate-knowledge children highlights the urgency of matching the instructional guidance to the learner.

## Limitations and Future Directions

Despite the positive contributions of the current study, future research is needed. First, studies should continue investigating the effects of feedback type. We did not detect many differences between outcome feedback and strategy feedback, and those we did were weak and inconsistent. It is possible that outcome and strategy feedback influence children's problem solving in similar ways. However, past research suggests that is not the case. For example, Luwel et al. (2011) examined children's use of two different correct strategies for completing a numerosity judgment task and found that strategy feedback led to more adaptive strategy selection than outcome feedback. It may be that strategy feedback is more beneficial when children are choosing among multiple strategies that vary in efficiency, not accuracy.

More work is also needed to verify the generalizability of our results across domains and settings. For example, feedback may have a larger impact for low-knowledge learners in domains with misconceptions, such as mathematical equivalence, because the role of feedback is to facilitate the correction of misconceptions and errors. In domains without misconceptions, feedback may be less necessary. Also, feedback may be most effective in one-on-one tutoring settings, in which feedback is immediate and can influence current performance. The focus of future work should be a more feasible application of feedback in a classroom setting, such as providing answers to a problem set or providing delayed written feedback on a homework assignment or test.

Finally, additional clarifications regarding the distinction among levels of prior knowledge are necessary. For example, future work should address what counts as sufficient prior knowledge so as to determine when feedback is no longer effective. In the current study, children simply needed to use a correct strategy on the target problem type once or twice. More generally, studies that demonstrate treatment by prior knowledge interactions have not identified the precise level of prior knowledge at which the reversal occurs. As more research shows that the effectiveness of instruction depends on prior knowledge, instructors will need guidance on how to choose instructional methods for particular children with particular levels of prior knowledge.

## Conclusion

This study extends research on guided discovery methods in which exploratory activities are provided with subsequent explicit instruction. We examined how and for whom a particular form of guidance, feedback, might enhance learning from the combination of exploration and instruction. Feedback during exploratory problem solving prior to instruction facilitates learning for children with low prior knowledge of a domain. However, children with moderate prior knowledge benefit more from exploring independently without feedback before receiving explicit instruction. Thus, providing feedback may not always be optimal.

## References


Ahmad, M. (1988, January). The effect of computer-based feedback on using cognitive strategies of problem solving. In *Proceedings of selected research papers presented at the annual meeting of the Association for Educational Communications and Technology, Research and Theory Division, New Orleans, LA* (pp. 1–22). Bloomington, IN: Association for Educational Communications and Technology.

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. doi:10.1037/a0021017

Alibali, M. W. (1999). How children change their minds: Strategy change can be gradual or abrupt. *Developmental Psychology, 35,* 127–145. doi:10.1037/0012-1649.35.1.127

Alibali, M. W., & Goldin-Meadow, S. (1993). Gesture-speech mismatch and mechanisms of learning: What the hands reveal about a child's state of mind. *Cognitive Psychology, 25,* 468–523. doi:10.1006/cogp.1993.1012

Anderson, R. C., Kulhavy, R. W., & Andre, T. (1972). Conditions under which feedback facilitates learning from programmed lessons. *Journal of Educational Psychology, 63,* 186–188. doi:10.1037/h0032653

Barrows, H. S., & Tamblyn, R. (1980). *Problem-based learning: An approach to medical education.* New York, NY: Springer.

Birenbaum, M., & Tatsuoka, K. (1987). Effects of "on-line" test feedback on the seriousness of subsequent errors. *Journal of Educational Measurement, 24,* 145–155. doi:10.1111/j.1745-3984.1987.tb00270.x

Brainerd, C. J. (1972). Reinforcement and reversibility in quantity conservation acquisition. *Psychonomic Science, 27,* 114–116.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20,* 37–46. doi:10.1177/001316446002000104

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist, 12,* 671–684. doi:10.1037/h0043943

Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods: A handbook for research on interactions.* New York, NY: Irvington.

DeCaro, M. S., & Rittle-Johnson, B. (2011, March). Preparing to learn from math instruction by solving problems first. In B. Rittle-Johnson & M. S. DeCaro (Chairs), *When are times for telling? Preparing students to learn from instruction. Symposium presented at the Society for Research in Child Development Conference, Montreal, ON, Canada*. Ann Arbor, MI: Society for Research in Child Development.

Hannafin, M. J., Hannafin, K. M., & Dalton, D. W. (1993). Feedback and emerging instructional technologies. In J. V. Demposey, G. C. Dales (Eds.), *Interactive instruction and feedback* (pp. 263–286). Englewood Cliffs, NJ: Educational Technology.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of experimental and theoretical research. *Advances in Psychology, 52,*139–183. doi: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*(1), 81–112. doi:10.3102/003465430298487

Hiebert, J., & Grouws, D. A. (2007). The effects of classroom mathematics teaching on students' learning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 371–404). Charlotte, NC: Information Age.

Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist, 42,* 99–107. doi:10.1080/00461520701263368

Hofer, S., Nussbaumer, D., & Schneider, M. (2011, 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 biennial conference of the European Association for Research on Learning and Instruction Conference, Exeter, United Kingdom.

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review, 19,* 509–539. doi:10.1007/s10648-007-9054-3

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist, 38,* 23–31. doi:10.1207/S15326985EP3801_4

Kalyuga, S., Chandler, P., & Sweller, J. (1998). Levels of expertise and instructional design. *Human Factors, 40,* 1–17. doi:10.1518/001872098779480587

Kalyuga, S., Chandler, P., & Sweller, J. (2004). When redundant on-screen text in multimedia technical instruction can interfere with learning. *Human Factors, 46,* 567–581. doi:10.1518/hfes.46.3.567.50405

Kalyuga, S., & Sweller, J. (2004). Measuring knowledge to optimize cognitive load factors during instruction. *Journal of Educational Psychology, 96,* 558–568. doi:10.1037/0022-0663.96.3.558

Kieran, C. (1981). Concepts associated with the equality symbol. *Educational Studies in Mathematics, 12,* 317–326. doi:10.1007/BF00311062

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. doi:10.1037/0033-2909.119.2.254

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. doi:10.1016/j.learninstruc.2008.03.003

Kulhavy, R. W. (1977). Feedback in written instruction. *Review of Educational Research, 47,* 211–232. doi:10.2307/1170128

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33,* 159–174. doi:10.2307/2529310

Leahy, W., & Sweller, J. (2005). Interactions among the imagination, expertise reversal, and element interactivity effects. *Journal of Experimental Psychology: Applied, 11,* 266–276. doi:10.1037/1076-898X.11.4.266

Lorch, R. F., Jr., Lorch, E. P., Calderhead, W. J., Dunlap, E. E., Hodell, E. C., & Freer, B. D. (2010). Learning the control of variables strategy in higher and lower achieving classrooms: Contributions of explicit instruction and experimentation. *Journal of Educational Psychology, 102,* 90–101. doi:10.1037/a0017972

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. doi:10.1016/j.jecp.2010.06.001

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. doi:10.1016/j.jecp.2008.08.004

Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? The case for guided methods of instruction. *American Psychologist, 59,* 14–19. doi:10.1037/0003-066X.59.1.14

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. doi: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. doi:10.1111/j.1467-8624.2005.00884.x

Mory, E. H. (2004). Feedback research revisited. In D. Jonassen, *Handbook of research on educational communications and technology: A project for the Association for Educational Communications and Technology* (2nd. pp. 745–783). Mahwah, NJ: Erlbaum.

Perry, M. (1988). Problem assignment and learning outcomes in nine fourth-grade mathematics classes. *Elementary School Journal, 88,* 412–426. doi:10.1086/461547

Perry, M., Church, R. B., & Goldin-Meadow, S. (1988). Transitional knowledge in the acquisition of concepts. *Cognitive Development, 3,* 359–400. doi:10.1016/0885-2014(88)90021-4

Phye, G. D., & Bender, T. (1989). Feedback complexity and practice: Response pattern analysis in retention and transfer. *Contemporary Educational Psychology, 14,* 97–110. doi:10.1016/0361-476X(89)90028-3

Rey, G. D., & Buchwald, F. (2011). The expertise reversal effect: Cognitive load and motivational explanations. *Journal of Experimental Psychology: Applied, 17,* 33–48. doi:10.1037/a0022243

Rittle-Johnson, B., & Alibali, M. W. (1999). Conceptual and procedural knowledge of mathematics: Does one lead to the other? *Journal of Educational Psychology, 91,* 175–189. doi: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. doi:10.1037/a0021334

Rittle-Johnson, B., Star, J. R., & Durkin, K. (2009). The importance of prior knowledge when comparing examples: Influences on conceptual and procedural knowledge of equation solving. *Journal of Educational Psychology, 101,* 836–852. doi:10.1037/a0016026

Rutherford, F. J. (1964). The role of inquiry in science teaching. *Journal of Research in Science Teaching, 2,* 80–84. doi:10.1002/tea.3660020204

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime user's guide.* Pittsburgh, PA: Psychology Software Tools.

Schnotz, W. (2010). Reanalyzing the expertise reversal effect. *Instructional Science, 38,* 315–323. doi:10.1007/s11251-009-9104-y

Schnotz, W., & Kurschner, C. (2007). A reconsideration of cognitive load theory. *Educational Psychology Review, 19,* 469–508. doi:10.1007/s10648-007-9053-4

Schwartz, D., & Bransford, J. D. (1998). A time for telling. *Cognition and Instruction, 16,* 475–522. doi:10.1207/s1532690xci1604_4

Schwartz, D., & Martin, T. (2004). Inventing to prepare for future learning: The hidden efficiency of encouraging original student production in statistics. *Cognition and Instruction, 22,* 129–184. doi:10.1207/s1532690xci2202_1

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. doi:10.1037/a0025140

Schwartz, D., Lindgren, R., & Lewis, S. (2009). Constructivist in an age of non-constructivist assessments. In S. Tobias & T. Duffy (Eds.), *Constructivist instruction: Success of failure?* (pp. 34–61). New York, NY: Routledge.

Seo K-H., & Ginsburg, H. P. (2003). "You've got to carefully read the math sentence . . ."; Classroom context and children's interpretation of the equals sign. In A. J. Baroody & A. Dowker (Eds.), *The development of arithmetic concepts and skills* (pp. 161–187). Mahwah, NJ: Erlbaum.

Shrager, J., & Siegler, R. S. (1998). SCADS: A model of children's strategy choices and strategy discoveries. *Psychological Science, 9,* 405–410. doi:10.1111/1467-9280.00076

Siegler, R. S. (1994). Cognitive variability: A key to understanding cognitive development. *Current Directions in Psychological Science, 3,* 1–5. doi:10.1111/1467-8721.ep10769817

Siegler, R. S. (1996). *Emerging minds: The process of change in children's thinking.* New York, NY: Oxford University Press.

Siegler, R. S., & Crowley, K. (1991). The microgenetic method: A direct means for studying cognitive development. *American Psychologist, 46,* 606–620. doi:10.1037/0003-066X.46.6.606

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.

Snow, R. (1978). Theory and method for research on aptitude processes. *Intelligence, 2,* 225–278. doi:10.1016/0160-2896(78)90019-3

Snow, R. E., & Lohman, D. F. (1984). Toward a theory of cognitive aptitude for learning from instruction. *Journal of Educational Psychology, 76,* 347–376. doi:10.1037/0022-0663.76.3.347

Steffe, L., & Gale, J. (Eds.). (1995). *Constructivism in education.* Hillsdale, NJ: Erlbaum.

Stigler, J., & Hiebert, J. (1998, Winter). Teaching is a cultural activity. *American Educator,* 1–10.

Tobias, S. (2009). An eclectic appraisal of the success of failure of constructivist instruction. In S. Tobias & T. D. Duffy (Eds.), *Constructivist theory applied to education: Success of failure?* (pp. 335–350). New York, NY: Routledge, Taylor and Francis.

West, S. G., Aiken, L. S., & Krull, J. L. (1996). Experimental personality designs: Analyzing categorical by continuous variable interactions. *Journal of Personality, 64,* 1–48. doi:10.1111/j.1467-6494.1996.tb00813.x

Wise, A. F., & O'Neill, K. (2009). Beyond more versus less: A reframing of the debate on instructional guidance. In S. Tobias & T. M. Duffy (Eds.), *Constructivist instruction: Success of failure?* (pp. 82–105). New York, NY: Routledge.

Received July 28, 2011
Revision received March 20, 2012
Accepted March 22, 2012 ■