Instr Sci
DOI 10.1007/s11251-014-9323-8

When feedback is cognitively-demanding: the importance of working memory capacity

Emily R. Fyfe · Marci S. DeCaro · Bethany Rittle-Johnson

Received: 3 March 2013 / Accepted: 29 April 2014

Abstract Feedback is generally considered a beneficial learning tool, and providing feedback is a recommended instructional practice. However, there are a variety of feedback types with little guidance on how to choose the most effective one. We examined individual differences in working memory capacity as a potential moderator of feedback type. Second- and third-grade children ($N = 64$) solved unfamiliar math problems prior to receiving instruction. Children received verification feedback on their answers (outcome-feedback) or on their strategies (strategy-feedback). Working memory capacity moderated the effect of feedback type on procedural transfer—the ability to solve novel problems. Children with lower working memory capacity benefitted less from strategy-feedback than outcome-feedback, whereas children with higher working memory capacity benefitted similarly from the two types of feedback. Results suggest the need to consider the cognitive demands of different feedback types. Problem solving can be optimized by considering both characteristics of the learner and the learning environment.

Keywords Feedback · Working memory · Problem solving · Mathematics learning · Cognitive load

The role of feedback during learning and problem solving has been studied extensively (e.g., Mory 2004; Hattie and Gan 2011; Shute 2008). In learning contexts, the purpose of feedback is to provide information that the learner can use to confirm, reject, or modify

E. R. Fyfe (✉)
Department of Psychology and Human Development, Vanderbilt University, 230 Appleton Place, Peabody #552, Nashville, TN 37203, USA
e-mail: emily.r.fyfe@vanderbilt.edu

M. S. DeCaro
Department of Psychological and Brain Sciences, University of Louisville, Louisville, KY, USA

B. Rittle-Johnson
Department of Psychology and Human Development, Vanderbilt University, Nashville, TN, USA

prior knowledge. In general, the effects of feedback are powerful and positive, leading many researchers to endorse the provision of feedback (e.g., Alfieri et al. 2011; Steedly et al. 2008). Indeed, Hattie and Timperley (2007) identify feedback as one of the "highest influences on achievement" in the classroom (p. 83). However, feedback effects show considerable variability (Kluger and DeNisi 1996), indicating some types of feedback are more powerful than others. In response to these findings, researchers have called for future work to specify key moderators of feedback type, including characteristics of the learner (Hattie and Gan 2011; Mory 2004).

The type of feedback provided (e.g., focused on outcomes, strategies, effort, speed, etc.) guides learners attention and may narrow the type of information the learner processes and potentially corrects (Kluger and DeNisi 1996). Consequently, different feedback types may place different demands on learners' cognitive resources (e.g., Moreno 2004). In this case, individual learner differences may impact feedback's effectiveness. For example, working memory (WM) capacity supports learners' ability to actively select, regulate, and process task-relevant information such as feedback (Alloway 2006). Further, WM capacity varies across individuals, with some people demonstrating higher WM capacity than others (Alloway 2006; Conway et al. 2005). Thus, individual differences in WM capacity likely constrain how feedback is processed and what can be learned from different feedback types.

We examined the role of working memory in learning from feedback. Specifically, we examined the cognitive demands of two different feedback types and how differences in WM capacity impact their effects. We focused on two types of feedback that vary in content: outcome-feedback, which focuses on accuracy of answers, and strategy-feedback, which focuses on how answers are obtained. These factors were studied in the context of exploratory problem solving, in which elementary-school children solved unfamiliar mathematics problems prior to instruction. Although WM is thought to play an important role in the utility of feedback (e.g., Schooler and Anderson 1990; Moreno 2004), empirical evaluations are lacking.

Theoretical motivation

By investigating WM capacity as a potential moderator of feedback type, we can better understand why certain types of feedback are more or less effective than others. Monitoring and evaluating feedback rely on WM resources. Indeed, theoretical models of feedback are specifically concerned with internal cognitive processes, such as WM, that might affect how feedback is perceived and used (e.g., Clariana et al. 2000; Kulhavy and Stock 1989). For example, Kulhavy and Stock (1989) suggest that at each point in the task (i.e., initial question, feedback, later question), the learner engages in cognitive activity to process the input and generate a response. This processing includes relating feedback to the initial response, integrating information with prior knowledge, and evaluating one's performance. All of this processing relies extensively on WM resources.

According to cognitive load theory, when instructional techniques place high demands on WM, the system can become overloaded, and learning suffers (Sweller 1998). Cognitive load represents the total burden a task imposes on the learners' cognitive system and it is thus a multidimensional construct that includes a variety of factors, including mental effort, fatigue, and frustration (Paas et al. 2003). Sweller et al. (1998) differentiate three different types of cognitive load. Intrinsic load is a result of the complexity of the to-be-learned content. Extrinsic load is a result of unnecessary cognitive processing caused by suboptimal instructional designs. Finally, germane load is a result of effortful, productive learning processes. According to cognitive load theory, a central goal of instructional design is to limit the total cognitive load imposed on the learner in an attempt to enhance learning.

To the extent that feedback can be productively processed within WM, it will likely result in germane cognitive load. However, feedback may also lead to extraneous processing and overburden WM resources. For example, types of feedback that are overly detailed, unfamiliar, or difficult to process may tax WM resources to a greater extent, leading to higher total cognitive load and less effective learning. Further, certain learners, such as those with lower WM capacity, may be more susceptible to these taxing effects. We discuss feedback type and WM capacity in turn, focusing on their relevance to mathematics problem solving.

Feedback type

In research and in practice, there are a variety of feedback "types," as feedback can be used for a variety of purposes (Mory 2004). We focus on corrective feedback in a problem-solving setting, which is used to help learners detect errors and generate correct alternatives (Dempsey et al. 1993). In particular, we examine two types of corrective feedback that vary in content: outcome-feedback and strategy-feedback. *Outcome-feedback* is focused on the accuracy of the learner's response, whereas *strategy-feedback* is focused on how the learner obtained the response. In the context of mathematics problem solving, outcomes refer to numerical answers and strategies refer to domain-specific problem-solving procedures. This distinction is salient in problem solving, in which the use of different strategies can lead to the same outcome and the use of the same strategy can lead to different outcomes. For example, consider the following problem: $3 + 4 + 5 = 5 + \_\_$. One learner may use a correct grouping strategy by canceling the 5s, which appear on both sides of the equal sign, and adding the 3 and 4. Another learner may use a more naïve strategy by always adding two random numbers (see Rittle-Johnson 2006). In this case, the two different strategies may result in the same outcome. The reverse is also true; for example, one correct strategy can lead to different answers if a calculation error is made.

In addition to varying in content, corrective feedback can also vary in the amount of information provided. For example, feedback can simply verify the correctness of a response (i.e., verification feedback), provide the correct response, or provide additional explanation (Dempsey et al. 1993). We examined the cognitive demands of the content of the feedback message (the type of information), so we kept the *amount* of information consistent. We opted to provide verification feedback for several reasons. First, verification feedback is used in several existing studies on the effects of feedback during mathematics problem solving (e.g., Alibali 1999; Hofer et al. 2011). Second, we were interested in children's exploration of novel problems and strategy generation. Recent evidence demonstrated the powerful effects of verification feedback on strategy discovery and transfer for elementary-school children (Baroody et al. 2013). Further, numerous studies suggest full explanations are more effective *after* problem solving with minimal feedback (e.g., Kapur 2011; DeCaro and Rittle-Johnson 2012). Finally, given concerns that extensive feedback might overwhelm WM capacity, especially in lower-capacity learners, it seemed important to constrain the amount of information in the feedback. Thus, we employed verification feedback on outcomes or on strategies.

Feedback focused on outcomes is one of the most common types of feedback in both educational and research contexts (e.g., Hattie and Timperley 2007; Pianta et al. 2007). Outcome-feedback generally benefits learning when compared to no-feedback conditions (e.g., Kluger and DeNisi 1996). Research suggests that it functions primarily by helping learners correct inaccurate information (e.g., Anderson et al. 1972; Phye and Bender 1989). That is, outcome-feedback helps learners identify their errors and search for more plausible alternatives. In these studies, outcome-feedback has its greatest impact when learners' answers are incorrect. Further, it helps prevent learners from making the same error multiple times.

Despite these positive effects, many researchers recommend focusing feedback more specifically on learners' strategies (e.g., Earley et al. 1990; Clifford 1986; Kamins and Dweck 1999; Luwel et al. 2011). Strategy-feedback concerns the processes that generate outcomes, as opposed to the outcomes themselves. During problem solving, learners must form hypotheses about which strategies are effective. Feedback that focuses on these strategies might help learners reject erroneous hypotheses and provide cues for further strategy searching. Indeed, Earley et al. (1990)examined outcome-feedback and strategy-feedback for undergraduates buying and selling stocks for hypothetical companies. Strategy-feedback was a more "direct and powerful way of shaping an individual's task strategy" (p. 103) than outcome-feedback and resulted in higher-quality information search. Similarly Luwel et al. (2011) examined children's performance on a numerosity judgment task (e.g., how many blocks are green?), which could be solved using one of two correct strategies. They found that strategy-feedback led to greater improvements in adaptive strategy selection than outcome-feedback. Finally, feedback on middle-school students' writing strategies led to better maintenance and generalization of these strategies relative to a no-feedback control (Schunk and Swartz 1993).

By providing information about strategies that apply across problems, strategy-feedback may also promote more transferrable knowledge than outcome-feedback. The benefits of outcome-feedback are often limited to the specific task on which the feedback was provided (e.g., Thompson 1998), whereas strategy-feedback has been shown to improve transfer (Schunk and Swartz 1993). This benefit may be salient for mathematics, in which outcome-feedback depends on specific numbers, but strategy-feedback applies to a wider range of problems.

Cognitive demands of outcome feedback and strategy feedback

Though strategy-feedback has potential advantages relative to outcome-feedback, it has potential consequences as well. In particular, strategy-feedback may tax WM resources to a greater extent than outcome-feedback and lead to greater cognitive load. Indeed, past research has found that different types of feedback can impose different levels of cognitive load (e.g., Corbalan et al. 2010; Moreno 2004). For example, Corbalan et al. (2010) investigated undergraduates' performance on complex linear algebra problems that required solving a variety of steps to obtain a final solution. Students reported lower levels of cognitive load when they received more frequent outcome feedback (after each step) than feedback on the final step only.

There are a number of reasons why strategy-feedback may require more WM resources than outcome-feedback. First, strategy-feedback is less familiar than outcome-feedback because teachers provide strategy-feedback less often than outcome-feedback. For example, in a documentation of classroom practice, Pianta et al. (2007) noted that feedback often referred to the correctness of answers. Thus, learners are more familiar with outcome-feedback, and as familiarity with a technique increases, WM resources are used to a lesser extent (Anderson 1982; Kirschner et al. 2006). Second, because math problems often require multi-step procedures to generate single numerical answers, processing strategy-feedback may require more resources and more time than processing outcome-feedback. Finally, research suggests that learning can be more difficult when hypothesis-testing is required (DeCaro et al. 2008). If strategy-feedback promotes hypothesis-testing about more complex information, it will place heavier demands on WM capacity. In general, research on strategy-feedback is sparse, so more work is needed to understand its cognitive demands. For example, in one meta-analysis, the low frequency of studies using strategy-feedback prevented the researchers from examining it empirically (Kluger and DeNisi 1996).

Individual differences in working memory capacity

Due to the potentially high cognitive demands of strategy-feedback, it may not be effective for all learners. Indeed, aptitude-by-treatment interactions suggest that the benefits of an instructional technique often depend on learner characteristics (Cronbach and Snow 1977). For example, both motivation and prior knowledge moderate the effects of feedback, with feedback generally having a more positive effect for learners with mastery achievement goals (e.g., VandeWalle 2003) and for learners with lower prior knowledge (e.g., Krause et al. 2009). Individual differences in working memory capacity may also impact learning from different feedback types. For example, researchers suggest that feedback competes for WM resources (Schooler and Anderson 1990), which may influence how different individuals process the presented material. That is, individuals with higher WM capacity may be more capable of learning from certain types of feedback relative to those leaners with lower WM capacity.

For children with lower WM capacity, the cognitive demands of strategy-feedback may overwhelm limited resources and impede learning, essentially creating too much cognitive load (e.g., Sweller et al. 1998). Thus, lower-capacity children may instead benefit more from familiar outcome-feedback, which may focus their attention on key information without requiring excessive processing. However, for learners with higher WM capacity, strategy-feedback may represent a "desirable difficulty" (Bjork 1994; McDaniel and Butler 2010). That is, strategy-feedback may place greater demands on WM, but not enough to overwhelm their resources. In particular, strategy-feedback may direct higher-capacity learners' attention to more fruitful information (i.e., germane cognitive load), which may promote transfer.

Current study

In summary, some researchers have proposed that strategy-feedback may be more beneficial than outcome-feedback (e.g., Earley et al. 1990; Luwel et al. 2011), but outcome-feedback is more common and is generally effective (cf. Hattie and Timperley 2007).We suggest that the effectiveness of the two feedback types may depend on learners' WM capacity. The cognitive demands of strategy-feedback may be too high for some learners, particularly those with lower WM capacity. Instead, lower capacity learners may benefit more from outcome-feedback. In contrast, strategy-feedback may be as or more effective for higher WM capacity learners. Given our focus on problem-solving and strategy acquisition, we expected these effects to occur on a measure of problem-solving transfer (i.e., the ability to solve novel problems). These findings can inform our understanding of the cognitive demands of different feedback types and reveal a potentially important individual difference that moderates their effectiveness.

We tested this hypothesis with children (M age = 7 years, 11 months) learning about math equivalence problems (e.g., $3 + 4 + 5 = 3 + \_\_$), which requires an understanding that both sides of an equation represent the same quantity. Math equivalence is a fundamental concept in arithmetic and a critical pre-requisite for understanding algebra (MacGregor and Stacey 1997). Unfortunately, many elementary school children solve these problems incorrectly, struggling to understand math equivalence (e.g., McNeil and Alibali 2005). For example, when asked to solve $3 + 4 + 5 = 3 + \_\_$, children often add all the numbers and put 15 in the blank, or add only the numbers before the equal sign and put 12 (e.g., McNeil 2007). These strategies are thought to stem from children's narrow experience with typical arithmetic problems in an "operations = answer" format, on which these strategies work successfully (McNeil and Alibali 2005).

In the current study, elementary school children participated in exploratory problem solving followed by brief instruction. Problem solving was considered exploratory because children were not explicitly instructed on correct procedures. Including opportunities for problem exploration prior to instruction is a recommended best practice in mathematics education (Dewey 1910; Hiebert and Grouws 2007; Lehrer and Kim 2009) and numerous studies support the benefit of this approach (e.g., DeCaro and Rittle-Johnson 2012; Kapur 2011; Schwartz et al. 2011). For example, DeCaro and Rittle-Johnson (2012) found that elementary-school children learned more about math equivalence when they solved problems with feedback before conceptual instruction, rather than vice versa. Problem exploration is thought to give learners a chance to process the problems at a deeper level and prepare them for future instruction (Schwartz et al. 2011). Thus, we adopted this more optimal learning condition (explore then instruct), but manipulated the type of feedback that was provided during exploration.

The current research question was embedded in a larger study examining the effects of feedback more generally (Fyfe et al. 2012). The larger published report examined the effects of providing feedback during problem solving and reported differences in learning outcomes based on learners' prior knowledge. The published report contains two experiments and the study reported here is a reanalysis of one of them. We did not publish findings on WM capacity in Fyfe et al. (2012) because it was not germane to the purpose of that paper, nor did we focus on the differential effects of strategy- versus outcome-feedback. Thus, the rationale, hypotheses, analyses, and conclusions are all distinct. Further, the sample was recruited and measures were administered with the current study's purpose in mind.

Method

Participants

Participants were drawn from a larger study reported in Fyfe et al. (2012). The sample reported here included 64 second- and third-grade children (M age = 7 years, 11 months, range = 6 years, 10 months to 9 years, 10 months; 35 girls; 36 second-graders) from eight different classrooms in two elementary schools (one public, one parochial). The vast majority of children ($n = 60$) were from the public school, as the private school was small and not all children had parent consent. The children were predominantly African American (98 %; 2 % White) and approximately 61 % received free or reduced price lunch. All of these children scored below 80 % on a conceptual and procedural knowledge measure at pretest (10 children were excluded due to this criterion). This criterion ensured that we worked with children who could still learn from the intervention.

Design

Children participated in a pretest, intervention, posttest, and two-week retention test. For the intervention, children were randomly assigned to the strategy-feedback ($n = 31$, 26 girls, 5 boys) or outcome-feedback ($n = 33$, 19 girls, 14 boys) condition. Random assignment was at the individual student level, but each classroom had approximately the same number of children assigned to each of the two conditions.

Materials

Intervention session

During the intervention, 12 math equivalence problems were presented in paper/pencil format. Six problems were 3- and 4- addend problems (e.g., $10 = 3 + \square$, $3 + 7 = \square + 6$), and the other six were 5-addend problems with a repeated addend on either side of the equal sign (e.g., $5 + 3 + 9 = 5 + \square$).

Assessment

The math equivalence assessment, adapted from past work (Matthews et al. 2012; 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 post- and retention test. The two forms differed primarily in the specific numbers used in the items. Items were previously matched to ensure the content and difficulty were comparable across forms, and item evaluation indicated strong evidence for construct and face validity (Matthews et al. 2012; Rittle-Johnson et al. 2011). The assessment included procedural learning, procedural transfer, and conceptual items (see Tables 1 and 2). The procedural items assessed children's use of correct strategies to solve math equivalence problems. The learning items were similar to those used for the intervention and the transfer items differed on key dimensions (e.g., inclusion of subtraction). The conceptual items assessed children's understanding of the equal sign and of the structure of equations. Alpha coefficients are reported for each subscale in Tables 1 and 2. To establish inter-rater reliability on open-ended conceptual items and on problem-solving strategies on the procedural items, a second rater coded 30 % of the responses. Inter-rater agreement was high (kappas = .88–.97).

Working memory

Children's WM capacity was measured using the backward digit span task from the Wechsler Intelligence Scale for Children (WISC-IV) Working Memory index (Wechsler 2003). Children were read a series of numbers at a rate of one per second and were asked to repeat the numbers in reverse order. Number series length began at two and ended at a

Table 1 Procedural items from the math equivalence assessment form 2

Procedural learning items ($\alpha = .67$)	Procedural transfer items ($\alpha = .76$)
$8 = 6 + \square$	$\square + 2 = 6 + 4$
$3 + 4 = \square + 5$	$8 + \square = 8 + 6 + 4$
$3 + 7 + 6 = \square + 6$	$5 + 6 - 3 = 5 + \square$
$7 + 6 + 4 = 7 + \square$	$5 - 2 + 4 = \square + 4$

Cronbach's alphas are from the retention test

Table 2 Conceptual items from the math equivalence assessment form 2

Conceptual knowledge items ($\alpha = .72$)	Scoring criteria
1. Reproduce $4 + 3 + 9 = 4 + \square$ from memory after viewing for 5 s	Reconstruct numerals, operators, equal sign and blank in correct location
2. Reproduce $8 + 6 + 3 = \square + 2$ from memory after viewing for 5 s	Same as above
3. What does the equal sign (=) mean?	Provide relational definition (e.g., the same amount as)
4. What goes in the box to show that 10 cents is the same mount of money as 1 dime?	Select the equal sign (=) from four options
5. Judge $3 = 3$ and $7 = 3 + 4$ as true or false	Judge both equations as true
6. Judge $31 + 16 = 16 + 31$ and $7 + 6 = 6 + 6 + 1$ as true or false	Same as above
7. Is this a good definition of the equal sign? (Given three definitions to rate)	Rate "two amounts are the same" as a good definition of the equal sign
8. Which definition above is the best definition of the equal sign?	Select "two amounts are the same" as best, over "add" and "the answer to the problem"
9. Decide if $6 + 4 = 5 + 5$ is true and explain how you know	Judge equation as true and explain that both sides of the equal sign are the same amount
10. In the statement: 1 dollar = 100 pennies; What does this equal sign mean?	Provide relational definition

Cronbach's alphas are from the retention test

maximum of eight. There were two items per series length. The task was discontinued when a child recalled both items in a series of a given length incorrectly. WM scores consisted of the number of series that the child correctly recalled in backward order.

Cognitive load

Three items measured children's subjective cognitive load during the intervention. Cognitive load is rarely measured in young children, and we are not aware of any validated measures of cognitive load for this age group. Thus, we adapted the language from three existing items used with older children or adults. The first item was a subjective rating of task difficulty adapted from Paas (1992), a measure used frequently to assess cognitive load in adolescents and adults (see Paas et al. 2003). Children were asked, "How easy or hard was it to solve all of those problems?" and responded on a 7-point scale ranging from *very, very easy* to *very, very hard*. The remaining two items were modified from the NASA Task Load Index (Hart and Staveland 1988), a measure used in previous studies with adults to assess cognitive load (e.g., Rey and Buchwald 2011; Zumbach and Mohraz 2008). The two items measured frustration ("I was stressed and irritated when I solved those problems.") and effort ("I had to work hard to solve those problems."). Children indicated their agreement on a 5-point scale. At least one study has directly compared the efficacy of both the NASA-TLX and the Paas (1992) item as measures of cognitive load during learning (Wiebe et al. 2010). They found that both measures were predictive of learning outcomes and sensitive to changes in cognitive load.

Because cognitive load scales have rarely been employed with children, we included three different items to explore how they functioned in a sample of elementary-school children. On the task difficulty item ($M = 5.1$ out of 7.0, $SD = 1.4$, $min = 2$, $max = 7$), approximately 60 % of children selected a 5 or higher. On the frustration item ($M = 3.1$ out of 5.0, $SD = 1.4$, $min = 1$, $max = 5$), an approximately equal percent of children selected each response. Finally, on the mental effort item ($M = 4.1$ out of 5.0, $SD = 1.2$, $min = 1$, $max = 5$), the distribution was more skewed with the majority of children selecting a 4 or 5. We analyzed the three measures of cognitive load separately, in line with past research (e.g., Rey and Buchwald 2011), as each taps a different aspect of cognitive load (effort, frustration, task difficulty). Indeed, cronbach's alpha across the three items was very low ($\alpha = .12$).

Procedure

Children completed a written pretest in their classrooms in one 30-minute session. Within 1 week, those who met the inclusion criteria (scored < 80 % on the conceptual and procedural pretest measures) completed a one-on-one tutoring intervention and an immediate posttest in a single session lasting approximately 50 min. The tutor was one of two female experimenters who were trained to administer the intervention, and the intervention took place in a quiet room at the child's school. The intervention began with an exploratory problem-solving phase. Children in both conditions were asked to solve 12 math equivalence problems presented with paper and pencil. After each problem, children reported either their strategy or their answer and received feedback on that report. This ensured the condition manipulation was clean and strong; only children in the strategy-feedback condition were encouraged to attend to their strategy and only children in the outcome-feedback condition were encouraged to attend to their answer.

In the *strategy-feedback condition*, children reported how they solved each problem and then received verification feedback on that strategy (see Table 3 for example strategy reports). Specifically, the experimenter repeated the child's strategy and stated whether it was a correct or an incorrect strategy. For example, if a child reported using the incorrect add-to-equal strategy on the problem $5 + 3 + 9 = 5 + \_\_$ (see Table 3), the experimenter repeated the child's report: "Good try, but that is not a correct way to solve the problem. Adding these three numbers (points to 5, 3, and 9) and putting the answer is not a correct way to solve this problem." The experimenter restated the strategy just as the child stated it to ensure no extra information was given. The strategy feedback was based solely on the child's strategy and did not depend on the numerical answer. For example, if a child reported using a correct strategy but obtained an incorrect answer (e.g., due to an arithmetic error), the experimenter provided feedback that the strategy was correct and did not comment on the correctness of the numerical answer. Mismatches between the correctness of strategies and numerical answers occurred on less than 3 % of all trials.

In the *outcome-feedback condition*, children reported their numerical answer and then received verification feedback on their answer. Specifically, the experimenter repeated the child's answer and stated whether it was a correct or an incorrect answer. If the answer was correct, the experimenter said, "Good job! You got the right answer. [Child's answer] is the correct answer." If the answer was incorrect, the experimenter said, "Good try, but you did not get the right answer. [Child's answer] is not the correct answer." The outcome feedback was based solely on the children's numerical answers, regardless of how they derived their answers.

Immediately following the exploratory problem-solving phase, children rated their subjective cognitive load using three different items. This rating was followed by a conceptual instruction phase during which all children received brief instruction on the meaning of the equal sign with number sentences as examples (e.g., $3 + 4 = 3 + 4$). The experimenter identified the two sides of the number sentences, provided an explicit definition of the equal sign (as meaning "the same amount as"), and explained how the left and right sides of the number sentences were equal. Children were asked simple questions to ensure they were attending to the instruction, but no solution strategies were discussed. At the end of the intervention, we administered the posttest and then measured children's WM capacity. Approximately 2 weeks after the intervention session, children completed the retention test in their classrooms.

Data analysis

We used analysis of variance models to test our hypotheses. The goal was to determine whether the effect of condition (outcome- vs. strategy-feedback) depended on individual differences in WM capacity. That is, we examined whether WM capacity changed the strength or direction of the relationship between condition and learning outcomes. As mentioned previously, we predicted that lower WM capacity learners may benefit more from outcome-feedback, but that strategy-feedback may be as or more effective for higher WM capacity learners. Thus, we included the primary independent variable (condition), the hypothesized moderator (WM capacity), and their interaction in the models. A significant interaction indicated that WM capacity moderated the effect of condition. Primary dependent variables included children's percent correct on procedural learning, procedural transfer, and conceptual knowledge (at both posttest and retention test). We also examined the three indicators of cognitive load (ratings of effort, frustration, and task difficulty) as secondary dependent measures. For the primary dependent measures, time (posttest and retention test) was included as a within-subject factor in the ANCOVA model. Procedural and conceptual pretest scores and age were included as covariates in all models. Exploratory analyses revealed no interactions with pretest scores or age so these interaction terms were not retained.

Results

Pretest performance

Children had low to moderate knowledge of math equivalence at pretest. Children's performance on the procedural items at pretest ($M = 28$ %, $SD = 20$ %) did not differ as a

Table 3 Strategies used to solve math equivalence problems

Strategy	How did you solve that problem? $(5 + 3 + 9 = 5 + \_\_)$
Correct strategies	
Equalize	5 plus 3 plus 9 is 17 and I know 5 plus 12 is 17
Add-subtract	I added 5, 3, and 9 and got 17, and 17 minus 5 is 12
Grouping	There was already a 5, so I just added 3 and 9
Incorrect strategies	
Add all	I added all of them together
Add-to-equal	I added these three (points to 5, 3, and 9) and put the answer
Carry	There's a 3 here so I put a 3 here
Other	I know 5 plus 3 is 8. Then I just put 7 because it's 1 less

function of condition, $F < 1.1$, or WM capacity, $r = .04$, $p = .73$. Similarly, children's performance on the conceptual items at pretest ($M = 18$ %, $SD = 18$ %) did not differ as a function of condition, $F < 1$, or WM capacity, $r = -.004$, $p = .97$. Even though no differences between conditions were found, children's procedural and conceptual pretest scores were included as covariates in all subsequent analyses to control for prior knowledge. Age was also included as a covariate in all analyses, though it did not differ as a function of condition, $F < 1$, or WM capacity, $r = .03$, $p = .80$.

Posttest and retention test performance

Procedural learning

At posttest, children's performance on the procedural learning items was moderate ($M = 42$ %, $SE = 5$ %), and remained similar 2 weeks later ($M = 36$ %, $SD = 5$ %). There were no effects of time, Fs < 2.4. Further, procedural learning did not differ as a function of condition, WM, or their interaction, Fs < 1. Thus, in general, children solved problems like those presented during the intervention moderately well following the intervention.

Procedural transfer

Children's procedural transfer was also moderate at posttest ($M = 31$ %, $SE = 4$ %) and remained similar 2 weeks later ($M = 25$ %, $SE = 4$ %). There were no effects of time, Fs < 2.2, condition, $F < 1$, or WM, $F < 1$. However, there was a significant condition by WM interaction, $F(1, 57) = 5.46$, $p = .02$, $\eta_p^2 = .09$. To explore this interaction, we examined the impact of condition at one standard deviation below (lower WM) and above (higher WM) the mean (see Fig. 1). This was accomplished by centering WM at one standard deviation below the mean in one model and at one standard deviation above the mean in a separate model, as recommended by Aiken and West (1991). As shown in Fig. 1, for children with lower WM capacity, strategy-feedback resulted in significantly lower transfer performance ($M = 27$ %) than outcome-feedback ($M = 57$ %), $F(1, 57) = 4.84$, $p = .03$, $\eta_p^2 = .08$. In contrast, for children with higher WM capacity, strategy-feedback

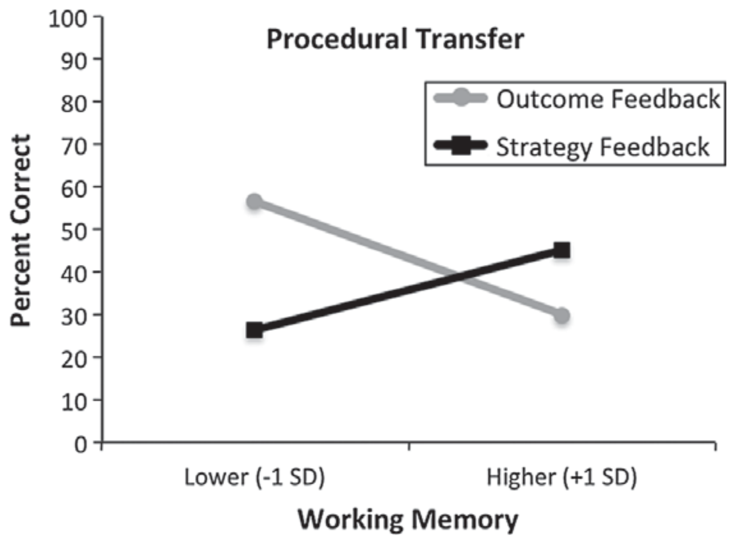

Fig. 1 Procedural transfer scores by feedback condition and working memory capacity Scores are estimated based on posttest and retention test scores. Nonstandardized coefficients are plotted at ± 1 SD from the mean

(M = 45 %) resulted in relatively higher transfer performance than outcome-feedback (M = 30 %), although this effect did not reach significance, $F(1, 57) = 1.24$, $p = .27$, $\eta_p^2 = .02$.

Even though there were no significant effects of time, we verified that the condition by WM interaction was consistent across both posttest and retention test. Specifically, we conducted separate ANCOVAs for each test occasion. The condition by WM interaction was significant at posttest, $F(1, 57) = 4.13$, $p = .05$, $\eta_p^2 = .07$, and marginal at retention test, $F(1, 57) = 3.15$, $p = .08$, $\eta_p^2 = .05$. Overall, WM moderated the impact of condition on procedural transfer, and this effect was consistent across time. Strategy-feedback led to lower transfer than outcome-feedback for children with lower WM capacity, but feedback type did not significantly impact transfer for children with higher WM capacity.

Conceptual knowledge

Children demonstrated good conceptual knowledge at posttest (M = 50 %, SE = 3 %), although it dropped somewhat 2 weeks later (M = 40 %, SE = 3 %), as indicated by a main effect of time, $F(1, 57) = 23.91$, $p < .001$, $\eta_p^2 = .30$. There were no significant interactions with time, $F\text{s} < 2.6$. There was a main effect of WM, $F(1, 57) = 7.55$, $p = .01$, $\eta_p^2 = .12$. Children with higher WM capacity exhibited higher conceptual knowledge than children with lower WM capacity (see Fig. 2). There was no main effect of condition, $F < 1$, nor did condition interact with WM, $F < 1.2$. Even though there were no interactions with time, we confirmed that the main effect of WM was significant at posttest, $F(1, 57) = 10.80$, $p = .002$, $\eta_p^2 = .16$, and marginal at retention test, $F(1, 57) = 3.71$, $p = .06$, $\eta_p^2 = .06$.

Cognitive load

Children in the strategy-feedback condition reported higher levels of effort and frustration than children in the outcome-feedback condition (effort: M = 4.4 out of 5.0, SE = 0.2 vs.

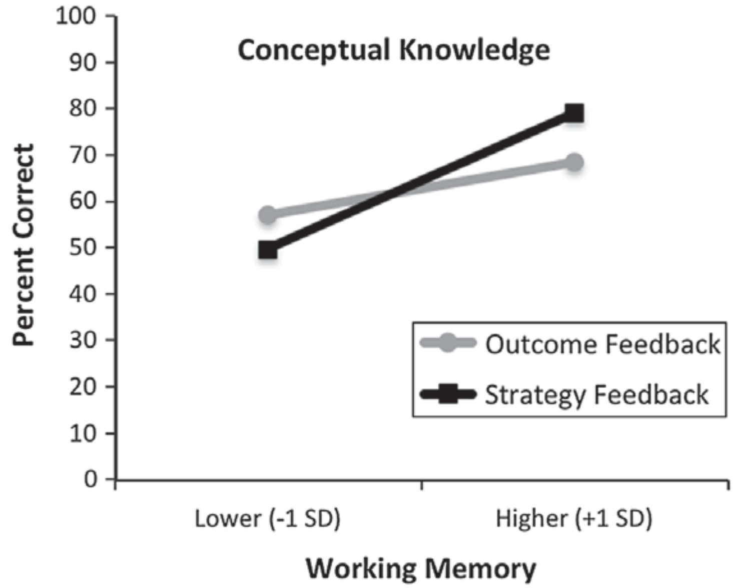

Fig. 2 Conceptual knowledge scores by feedback condition and working memory capacity Scores are estimated based on posttest and retention test scores. Nonstandardized coefficients are plotted at $\pm$ 1 SD from the mean

$M = 3.8$, $SE = 0.2$, $F(1, 56) = 4.31$, $p = .04$, $\eta_p^2 = .07$; frustration: $M = 3.5$ out of 5.0, $SE = 0.2$ vs. $M = 2.8$, $SE = 0.2$, $F(1, 57) = 4.73$, $p = .03$, $\eta_p^2 = .07$). Ratings of task difficulty did not differ by condition, ($M = 5.1$ out of 7.0, $SD = 1.4$), $F(1, 57) = 1.34$, $p = .25$, $\eta_p^2 = .02$. There were no main effects of WM, nor did condition interact with WM for any measure, F's < 1.2. Overall, these results suggest that strategy-feedback resulted in higher perceived effort and frustration than outcome-feedback, regardless of WM capacity.

Discussion

The effects of feedback on learning are powerful, though inconsistent (e.g., Kluger and DeNisi 1996), suggesting that some types of feedback are more effective than others. We examined the effect of feedback type on children's mathematics problem solving and whether individual differences in WM capacity moderated their relative effectiveness. During problem solving, children received verification feedback on their answers (outcome-feedback) or on their strategies (strategy-feedback). In contrast to researchers' suggestions (e.g., Earley et al. 1990; Luwel et al. 2011), we found no evidence that feedback on strategies is more beneficial than feedback on outcomes, and some evidence that it can be detrimental. For children with higher WM capacity, the differences between feedback conditions were not reliable. For children with lower WM capacity, strategy-feedback was less effective than outcome-feedback on subsequent measures of procedural transfer—the ability to solve novel problems—and this was consistent across time. Thus, we provide evidence for an aptitude by treatment interaction in which the effects of feedback type depend on learners' WM capacity (e.g., Cronbach and Snow 1977).

Cognitive load theory suggests one potential reason for the moderating effects of WM capacity. Children who received strategy-feedback reported higher levels of mental effort and frustration than children who received outcome-feedback. Although these ratings did not differ as a function of WM capacity, the increased perceptions of cognitive load may have been more detrimental for children with lower WM capacity. Specifically, the demands of strategy-feedback may have overwhelmed their limited WM resources and thus hindered their ability to process the problems in a deep manner (Sweller 1988). Even if strategy-feedback increased germane load, which is related to productive, effortful processing, the total cognitive load may have consumed too much attention and effort and thus hindered transfer (e.g., Sweller et al. 1998). In contrast, the direct, familiar outcome-feedback may have been processed more easily, freeing up more resources to focus on the problems. Children with higher WM capacity, however, were less impacted by feedback type. Higher-capacity children may have had sufficient resources to process either feedback type without experiencing detrimental effects of excessive cognitive load (Sweller et al. 1998). These findings support a growing number of studies focused on the potential consequences of cognitive overload in learning settings (e.g., Kalyuga, 2007; Mayer and Moreno 2003). As Sweller et al. (1998) note, “any instructional design that flouts or merely ignores WM limitations inevitably is deficient” (p. 253).

Although cognitive load theory provides a plausible explanation of the current results, this explanation should be interpreted with some caution. The empirical support in the current study is weak at best. Indeed, although cognitive load differed by condition, there was no effect of WM capacity nor did it interact with condition. Further, the cognitive load results are exploratory given the lack of a valid scale for this age group. Although we adapted existing items to be suitable for young children, it remains unclear whether eight-year-old children are capable of assessing their mental effort, their frustration, or task difficulty. The measurement of cognitive load is a frequent limitation, even among studies with adults (e.g., de Jong, 2010; Kalyuga 2011). Indeed, Kirschner et al. (2011)identify the subjective measurement of cognitive load as the “ugly side of [cognitive load theory] research” (p. 104). We acknowledge that the current study is no exception, but also note that cognitive load theory is primary centered around the optimal use of WM capacity, which is the key construct considered here.

Indeed, the current results contribute to a body of literature indicating that WM capacity is an important cognitive construct that should be considered in learning contexts (e.g., Alloway 2006; Beilock and DeCaro 2007; Sweller et al. 1998). In the current study, higher WM capacity was associated with higher conceptual knowledge of math equivalence after the intervention, and it also impacted how children responded to feedback type on a measure of procedural transfer. WM is thought to play an important role in a wide range of tasks relevant to learning, including reasoning and problem solving, and it also accounts for a significant portion of variance in general intellectual ability (cf. Conway et al. 2005; Engle et al. 1999). The present study suggests the need to consider individual differences in learners’ WM capacity in relation to the instructional method employed and the cognitive load inherent in learning contexts. Specifically, children with lower WM capacity may experience difficulty, and thus fall behind in learning settings, when feedback demands exceed their limited resources. This finding is consistent with calls to be mindful of lower WM capacity in learning contexts and to simplify instructional practices when possible (e.g., Alloway 2006). Indeed, Alloway (2006) suggests that that the working demands of many classroom activities may be one reason for the poor achievement of low capacity learners. A natural tendency is to provide more information to struggling learners, but this additional information may inadvertently harm learning.

In this study, children received exploratory problem solving followed by brief instruction. A number of researchers in psychology and education have recently endorsed this explore-instruct approach (e.g., Hiebert and Grouws 2007; Kapur 2011; Schwartz et al. 2011). For example, Schwartz et al. (2011) suggest that problem exploration prepares learners for future instruction by promoting attention to key problem features. Prior exploration can also create opportunities for productive failure, in which learners experience difficulty discovering correct solution, but ultimately process the learning material at a deeper level (Kapur 2011). Increasing evidence supports the benefits of this approach (DeCaro and Rittle-Johnson 2012; Kapur 2011, 2012; Schwartz and Martin 2004; Schwartz et al. 2011) and our goal was to consider the effects of feedback type and WM capacity within this particular context

A limitation of the present results concerns differences between the conditions in the current study. Although we focused on the type of feedback provided, the conditions also differed with respect to the type of verbal report asked of the child. In the strategy-feedback condition, children reported how they solved the problem and received feedback on that strategy. In the outcome-feedback condition, children reported their final numerical answer and received feedback on that solution. This difference was intended to focus children's attention on the content of the subsequent feedback message. However, one possibility is that describing one's strategy is more cognitively demanding than reporting one's answer, and that this difference (rather than the difference in the feedback provided) led to the current results. Although we acknowledge this as a limitation, previous research suggests that obtaining verbal reports of strategy use does not influence performance, as indicated by similar patterns of learning when reports are and are not requested (McGilly and Siegler 1990).

A further limitation is the absence of a condition in which children receive both types of feedback. For example, giving children outcome-feedback on some trials and strategy-feedback on others may have provided additional insight into how children respond to the cognitive demands of each feedback type. Also, providing both types of feedback on the same trial may have revealed key advantages (or disadvantages) the combination can offer relative to presenting one feedback type alone. In the current study, we systematically manipulated whether children received outcome-feedback or strategy-feedback in order to tease apart the cognitive demands of the different feedback types (see also Luwel et al. 2011). In future work, we hope to examine how various combinations of outcome- and strategy-feedback influence learning.

Future research is also necessary to examine the cognitive resources required to use different types of feedback effectively. In this study, we focused on two types of verification feedback that focused on numerical answers or domain-specific problem-solving strategies. Yet, feedback can focus on other content relevant to learning and performance, such as metacognitive strategies, effort, or speed (cf. Kluger and DeNisi 1996), and it can include additional information such as the correct answer or elaborate explanations (Dempsey et al. 1993). For example, Moreno (2004) provides some evidence that providing explanatory feedback can reduce cognitive load relative to correct-answer feedback alone. Finally, feedback can be used for a variety of purposes, other than correcting problem-solving errors. Feedback may also be used to increase motivation or productivity (Kim and Hamner 1976), to trigger metacognitive awareness (Kulhavy and Stock 1989), or to alter learners' expectations about their selves or the task (Leary et al. 2009). Clearly, more work is needed to examine the role of WM in learning from the wide variety of feedback types not investigated here.

Additionally, future work should examine the generalizability of the present results across settings, domains, and populations. For example, the brief, experimental one-on-one session is quite dissimilar from many learning settings. Also, feedback may have more pronounced effects in domains with common misconceptions, such as math equivalence, because one of the primary functions of feedback is to correct errors (e.g., Anderson et al. 1972). Finally, our sample included typical children in U.S. elementary schools; thus, these findings may not generalize to at-risk children or children educated in other countries.

Despite these limitations, the current study provides insight into the role of feedback in learning contexts. Specifically, this study demonstrates that the optimal type of feedback can depend on individual differences in WM capacity. Although previous research suggests that feedback on strategies can provide rich and useful information, this information might become cognitively burdensome during problem solving, particularly for those children with lower WM capacity. Thus, the current results caution against recommending strategy-feedback as a universally effective practice. By considering the cognitive processing elicited during learning, as well as the capacity of a learner to benefit from this processing, we can better understand the mechanisms by which instructional settings improve (or detriment) learning and transfer. Problem solving may be optimized when characteristics of both the learner and the learning environments are carefully considered.

Acknowledgments Research supported by National Science Foundation grant DRL-0746565 to Bethany Rittle-Johnson and Institute of Education Sciences, U.S. Department of Education, training grants R305B080008 and R305B080025. Fyfe is supported by a Graduate Research Fellowship from the National Science Foundation. The authors thank Abbey Loehr, Rachel Ross, and Polly Colgan for their assistance with data collection and coding as well Jones Paideia Elementary School and St. Pius Classical Academy for their participation in this research project.

References

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.

Alfieri, L., Brooks, P. J., Aldrich, N. J., & Tenenbaum, H. R. (2011). Does discovery-based instruction enhance learning? *Journal of Educational Psychology, 103*(1), 1–18.

Alibali, M. W. (1999). How children change their minds: Strategy change can be gradual or abrupt. *Developmental Psychology, 35*(1), 127–145.

Alloway, T. P. (2006). How does working memory work in the classroom? *Educational Research and Reviews, 1*(4), 134–139.

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review, 89*, 369–406.

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

Baroody, A. J., Eiland, M. D., Purpura, D. J., & Reid, E. E. (2013). Can computer-assisted discovery learning foster first graders' fluency with the most basic addition combinations? *American Educational Research Journal, 50*(3), 533–573.

Beilock, S., & DeCaro, M. S. (2007). From poor performance to success under stress: Working memory, strategy selection, and mathematical problem solving under pressure. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 33*(6), 983–998.

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about Knowing* (pp. 185–205). Cambridge, MA: MIT Press.

Clariana, R. B., Wagner, D., & Murphy, L. C. (2000). Applying a connectionist description of feedback timing. *Educational Technology Research and Development, 48*(3), 5–22.

Clifford, M. M. (1986). The effects of ability, strategy, and effort attributions for educational, business, and athletic failure. *British Journal of Educational Psychology, 56*, 169–179.

Conway, A., Kane, M., Bunting, M., Hambrick, D., Wilhelm, O., & Engle, R. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review, 12*, 769–786.

Corbalan, G., Paas, F., & Cuypers, H. (2010). Computer-based feedback in linear algebra: Effects on transfer performance and motivation. *Computers & Education, 55*, 692–703.

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

De Jong, T. (2010). Cognitive load theory, educational research, and instructional design: Some food for thought. *Instructional Science, 38*, 105–134. doi:10.1007/s11251-009-9110-0.

DeCaro, M. S., & Rittle-Johnson, B. (2012). Exploring mathematics problems prepares children to learn from instruction. *Journal of Experimental Child Psychology, 113*, 552–568.

DeCaro, M. S., Thomas, R. D., & Beilock, S. L. (2008). Individual differences in category learning: Sometimes less working memory capacity is better than more. *Cognition, 107*, 284–294.

Dempsey, J. V., Driscoll, M. P., & Swindell, L. K. (1993). Text-based feedback. In J. V. Dempsey & G. C. Sales (Eds.), *Interactive instruction and feedback* (pp. 21–54). Englewood Cliffs, NJ: Educational Technology Publications.

Dewey, J. (1910). *How we think*. Lexington, MA: DC Heath.

Earley, P. C., Northcraft, G. B., Lee, C., & Lituchy, T. R. (1990). Impact of process and outcome feedback on the relation of goal setting and task performance. *The Academy of Management Journal, 33*, 87–105.

Engle, R. W., Tulholsky, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent variable approach. *Journal of Experimental Psychology: General, 128*, 309–331.

Fyfe, E. R., Rittle-Johnson, B., & DeCaro, M. S. (2012). The effects of feedback during exploratory mathematics problem solving: Prior knowledge matters. *Journal of Educational Psychology, 104*(4), 1094–1108. doi:10.1037/a0028389.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX: Results of experimental and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human Mental Workload* (pp. 139–183). Amsterdam: North Holland.

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: 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.

Hofer, S., Nussbaumer, D., & Schneider, M. (2011). Practice without feedback can increase the adaptivity of strategy choices. In J. Torbeyns (chair), *Strategy flexibility: Analyzing its related structures and processes.* Symposium presented at the EARLI Conference, UK.

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review, 19*(4), 509–539. doi:10.1007/s10648-007-9054-3.

Kalyuga, S. (2011). Cognitive load theory: How many types of load does it really need? *Educational Psychology Review, 23*, 1–19. doi:10.1007/s10648-010-9150-7.

Kamins, M. L., & Dweck, C. S. (1999). Person versus process praise and criticism: Implications for contingent self-worth and coping. *Developmental Psychology, 35*(3), 835–847.

Kapur, M. (2011). A further study of productive failure in mathematical problem solving: Unpacking the design components. *Instructional Science, 39*(4), 561–579.

Kapur, M. (2012). Productive failure in learning the concept of variance. *Instructional Science, 40*, 651–672.

Kim, J. S., & Hamner, W. C. (1976). Effect of performance feedback and goal setting on productivity and satisfaction in an organizational setting. *Journal of Applied Psychology, 61*(1), 48–57.

Kirschner, P. A., Ayres, P., & Chandler, P. (2011). Contemporary cognitive load theory research: The good, the bad, and the ugly. *Computers in Human Behavior, 27*, 99–105. doi:10.1016/j.chb.2010.06.025.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*, 75–86.

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*(2), 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*(2), 158–170.

Kulhavy, R. W., & Stock, W. A. (1989). Feedback in written instruction: The place of response certitude. *Educational Psychology Review, 1*(4), 279–308.

Leary, M. R., Terry, M. L., Allen, A. B., & Tate, E. B. (2009). The concept of ego threat in social and personality psychology: Is ego threat a viable scientific construct? *Personality and Social Psychology Review, 13*, 151–164.

Lehrer, R., & Kim, M.-J. (2009). Structuring variability by negotiating its measure. *Mathematics Education Research Journal, 21*, 116–133.

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*(1), 61–76. doi:10.1016/j.jecp.2010.06.001.

MacGregor, M., & Stacey, K. (1997). Students' understanding of algebraic notation. *Educational Studies in Mathematics, 33*, 1–19.

Matthews, P. G., Rittle-Johnson, B., McEldoon, K., & Taylor, R. (2012). Measure for measure: What combining diverse measures reveals about children's understanding of the equal sign. *Journal for Research in Mathematics Education, 43*(3), 220–254.

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38*(1), 43–52.

McDaniel, M. A., & Butler, A. C. (2010). A contextual framework for understanding when difficulties are desirable. In A. Benjamin (Ed.), *Successful remembering and successful forgetting: Essays in honor of Robert A. Bjork* (pp. 175–199). NY: Psychology Press.

McGilly, K., & Siegler, R. (1990). The influence of encoding and strategic knowledge on children's choices among serial recall strategies. *Developmental Psychology, 26*, 931–941.

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

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*(4), 883–899. doi:10.1111/j.1467-8624.2005.00884.x.

Moreno, R. (2004). Decreasing cognitive load for novice students: Effects of explanatory versus corrective feedback in discovery-based multimedia. *Instructional Science, 32*, 99–113.

Mory, E. H. (2004). Feedback research revisited. In D. Jonassen, *Handbook of research on educational communications and technology* (2nd. pp. 745–783). NJ: Erlbaum.

Paas, F. (1992). Training strategies for attaining transfer of problem solving skill in statistics: A cognitive load approach. *Journal of Educational Psychology, 84*, 429–434.

Paas, F., Tuovinen, J. E., Tabbers, H., & van Gerven, P. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist, 38*(1), 63–71.

Phye, G. D., & Bender, T. (1989). Feedback complexity and practice: Response pattern analysis in retention and transfer. *Contemporary Educational Psychology, 14*, 97–110.

Pianta, R. C., Belsky, J., Houts, R., & Morrison, F. (2007). Opportunities to learn in America's elementary classrooms [Supplemental material]. *Science, 315*(5820), 1795–1796.

Rey, G. D., & Buchwald, F. (2011). The expertise reversal effect: Cognitive load and motivational explanations. *Journal of Experimental Psychology. Applied, 17*(1), 33–48.

Rittle-Johnson, B. (2006). Promoting transfer: Effects of self-explanation and direct instruction. *Child Development, 77*(1), 1–15. doi:10.1111/j.1467-8624.2006.00852.x.

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*(1), 85–104.

Schooler, L. J., & Anderson, J. R. (1990). The disruptive potential of immediate feedback. *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society* (pp. 702–708). Cambridge, MA.

Schunk, D. H., & Swartz, C. W. (1993). Goals and progress feedback: Effects on self-efficacy and writing achievement. *Contemporary Educational Psychology, 18*(3), 337–354.

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

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*, 153–189.

Steedly, K., Dragoo, K., Arafeh, S., & Luke, S. (2008). Effective mathematics instruction. *Evidence for Education, 3*(1), 1–11. Retrieved from: http://nichcy.org/research/ee/math. Accessed 12 May 2014.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257–285. doi:10.1016/0364-0213(88)90023-7.

Sweller, J., Merrienboer, J., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251–296.

Thompson, T. (1998). Metamemory accuracy: Effects of feedback and the stability of individual differences. *American Journal of Psychology, 111*, 33–42.

VandeWalle, D. (2003). A goal orientation model of feedback-seeking behavior. *Human Resource Management Review, 13*(4), 581–604. doi:10.1016/j.hrmr.2003.11.004.

Wechsler, D. (2003). *Wechsler intelligence scale for children—4th edition (WISC-IV)*. San Antonio, TX: Harcourt Assessment.

Wiebe, E. N., Roberts, E., & Behrend, T. S. (2010). An examination of two mental workload measurement approaches to understanding multimedia learning. *Computers in Human Behavior, 26*(3), 474–481. doi:10.1016/j.chb.2009.12.006.

Zumbach, J., & Mohraz, M. (2008). Cognitive load in hypermedia reading comprehension: Influence of text type and linearity. *Computers in Human Behavior, 24*, 875–887.