

The benefits of computer-generated feedback for mathematics problem solving

Emily R. Fyfe and Bethany Rittle-Johnson

Vanderbilt University

Journal article published in *Journal of Experimental Child Psychology* in 2016

Full citation:

Fyfe, E. R., & Rittle-Johnson, B. (2016). The benefits of computer-generated feedback for mathematics problem solving. *Journal of Experimental Child Psychology*, 147, 140-151. doi:10.1016/j.jecp.2016.03.009

Author Note

Emily R. Fyfe and Bethany Rittle-Johnson, Department of Psychology and Human Development, Vanderbilt University. Fyfe is now at the University of Wisconsin-Madison and is supported by U.S. Department of Education, IES Research Training Program under grant number R305B130007, as part of the Wisconsin Center for Education Research Postdoctoral Training Program in Mathematical Thinking, Learning, and Instruction.

The research described in this paper was supported in part by a Graduate Research Fellowship from the National Science Foundation to Fyfe, National Science Foundation grant DRL-9746565 to Rittle-Johnson, a dissertation research award from the American Psychological Association to Fyfe, and Institute of Education Sciences, U. S. Department of Education, training grant R305B080025. The authors thank Emilie Hall, Emily Litzow, and Ashli-Ann Douglas for their help with data collection and coding as well as the teachers and children at Sylvan Park Elementary and Saint Henry Catholic School for their participation.

Address correspondence to Emily R. Fyfe, 970 Educational Sciences Building, 1025 West Johnson Street, Madison WI 53706. Email: efyfe@wisc.edu.

Abstract

The goal of the current research was to better understand when and why feedback has positive effects on learning and to identify features of feedback that may improve its efficacy. In a randomized experiment, second-grade children received instruction on a correct problem-solving strategy and then solved a set of relevant problems. Children were assigned to receive no feedback, immediate feedback, or summative feedback from the computer. On a posttest the following day, feedback resulted in higher scores relative to no feedback for children who started with low prior knowledge. Immediate feedback was particularly effective, facilitating mastery of the material for children with both low and high prior knowledge. Results suggest that minimal computer-generated feedback can be a powerful form of guidance during problem solving.

Key words: feedback, problem solving, computer learning, mathematics learning

The benefits of computer-generated feedback for mathematics problem solving

Feedback often has powerful, positive effects for children across development. For example, feedback has been shown to improve performance for preschoolers on a card sort task (Bohlmann & Fenson, 2005), middle-school students on a writing assignment (Gielen et al., 2010), and undergraduates on a multiple-choice test of general knowledge (Butler & Roediger, 2008). However, the effects of feedback are not universally beneficial (Mory, 2004). The goal of the current research is to better understand when feedback has positive effects and to identify features of feedback that may improve its efficacy. Specifically, we manipulated the presence and timing of feedback to experimentally test their impact on children's mathematics learning.

The Mixed Effects of Feedback

In most learning contexts, the purpose of feedback is to provide information that the learner can use to confirm or modify prior knowledge. This feedback can promote the correction of errors (Kulhavy, 1977) and increase motivation (Mory, 2004). In many cases, feedback is helpful as intended and improves learning and performance. Indeed, meta-analyses consistently show that, on average, feedback has a positive effect relative to no feedback (Bangert-Drowns et al., 1991; Kluger & DeNisi, 1996). However, the impact of feedback varies, and under some circumstances, feedback can have neutral or negative effects (see Hattie & Gan, 2011). For example, negative effects have occurred for both right/wrong feedback (Pashler et al., 2005) and correct-answer feedback (Hays, Kornell, & Bjork, 2010) on adults' word learning.

Recent research suggests that learners who already have some knowledge in the domain are especially likely to experience negative effects of feedback. For example, in Fyfe, Rittle-Johnson, and DeCaro (2012), second- and third-grade children solved novel math problems. Some children received trial-by-trial right/wrong feedback, while others received no feedback.

For children with low prior knowledge on the pretest, feedback resulted in *higher* posttest scores than no feedback. However, for children with higher prior knowledge, feedback resulted in *lower* posttest scores than no feedback. Similar results were found when prior knowledge was manipulated via instruction on a correct strategy (Fyfe & Rittle-Johnson, 2015). Given that feedback *can* hinder learning under certain circumstances, more research is needed to understand when and why this occurs and to identify features of feedback that may improve its efficacy.

The Timing of Feedback

The timing of feedback may be one feature that influences the efficacy of feedback. Some researchers believe that feedback should be given immediately after a response in order to eliminate incorrect ways of thinking and reinforce correct ones (Skinner, 1954). Further, immediate feedback may provide motivation to practice, as progress can be easily monitored (Shute, 2008). However, others believe delaying feedback is more beneficial. First, it may prevent learners from becoming over-reliant on the immediate presentation of feedback, which in turn may increase the need to exert effort on one's own response (Bangert-Drowns et al., 1991). Second, delaying feedback allows for the strength of initially incorrect responses to dissipate, which may make processing correct responses easier (Kulhavy, 1977). Finally, delaying feedback allows for spaced presentation of information (Butler, Karpicke, & Roediger, 2007).

Several meta-analyses point to the advantages of immediate feedback, particularly in computer based instruction (Azevedo & Bernard, 1995) and applied classroom studies (Kulik & Kulik, 1988). Indeed, multiple experiments have demonstrated the superiority of immediate feedback over delayed feedback for the acquisition of verbal materials (Beeson, 1973; Brosvic, Dihoff, & Epstein, & Cook, 2006; Dihoff, Brosvic, & Epstein, 2003). However, a substantial body of research has found no difference between immediate and delayed feedback (e.g., Nakata,

2015; Smits, Boon, Sluijsmans, & Van Gog, 2008) or significant advantages of delayed feedback (Bangert-Drowns et al., 1991; Butler et al., 2007; Butler & Roediger, 2008; Clariana et al., 2000; Kulhavy, 1977; Metcalfe, Kornell, & Finn, 2009; Smith & Kimball, 2010). For example, Butler and Roediger (2008) had undergraduate students study general knowledge passages and take a multiple-choice test. Feedback after each response resulted in a lower proportion of correct responses on a one-week posttest than delayed feedback after all test questions were completed.

Recent work has found advantages of delaying feedback even after controlling for key confounds. For example, Metcalfe et al. (2009) found that delayed feedback was more effective than immediate feedback for sixth-grade students' vocabulary learning after controlling for the shorter retention interval between delayed feedback and the time of testing (but see Nakata, 2015). Recently, Mullet et al. (2014) found benefits of delayed feedback after controlling for time spent processing the feedback. Specifically, undergraduate engineering students had to view the feedback on their weekly homework assignments in order to get credit, regardless of when the feedback was provided. Students who received feedback one week after the assignments scored higher on a final exam than students who received feedback right after each assignment.

Thus, in general the evidence on the timing of feedback is mixed, although more and more research is finding benefits of delayed feedback. However, there are still several gaps in the literature that need to be addressed. First, the majority of research is in the context of adults learning from multiple-choice tests and the effects of feedback on their recall of information. Although this work is informative, it may not generalize to children learning mathematics and their ability to generate a solution strategy to a class of novel problems.

Second, there has been no systematic, experimental investigation of the timing of feedback in relation to learners' prior knowledge. As mentioned above, prior knowledge often

predicts whether learners benefit from the presence versus absence of feedback (e.g., Fyfe et al., 2012; Gielen et al., 2010; Krause et al., 2009). Thus, it seems plausible that prior knowledge may also predict learning from immediate versus delayed feedback. Indeed, several researchers have suggested that low-knowledge learners may benefit from immediate feedback as they need to correct initial errors, but high-knowledge learners may benefit from delayed feedback as they need to process problems deeply with little intrusion (e.g., Gaynor, 1981).

Third, little attention has been paid to how the timing of feedback impacts learners' affective states. A leading theory suggests that feedback is less likely to be effective when it directs attention to the self as opposed to the task, because attention on the self can produce affective reactions that interfere with task-relevant processing (Kluger & DeNisi, 1996). For example, negative feedback can produce ego-threat, which may reduce one's confidence or motivation to continue. One possibility is that immediate trial-by-trial feedback results in higher negative affect because it is provided during the learning task. An alternative possibility is that delayed, summative feedback results in higher negative affect because it draws a lot of attention to the self all at once, with no subsequent task on which to redeem one's self-image.

Current Study

The goal of the current study was to address these gaps in the literature. Specifically, we manipulated the presence and timing of feedback to experimentally test their impact on mathematics learning for children with varying levels of prior knowledge. We based our study design on our previous work (Fyfe & Rittle-Johnson, 2015). In that study, children who received prior instruction on the math problems exhibited better learning on the posttest when they received practice *without* feedback than practice with feedback. Further, the negative effect of feedback did not depend on children's prior knowledge on a pretest. Here, we used the same

basic method and assessments. For example, the instruction script and practice problems were identical to those in the previous study, as were the pretest and posttest items. However, we made key changes to enhance the external validity of the results.

First, the tutor's presence was removed during problem solving and the feedback was provided solely by the computer. The use of computers and computer tutor systems in classrooms is increasing rapidly (Greaves & Hayes, 2008), and these systems often provide targeted practice with feedback. Indeed, the oft-cited benefit of computer programs is the ability to give immediate feedback on every student's responses. Thus, it is imperative to understand how the presence and timing of computer-generated feedback affects mathematics learning. Second, the feedback in this study did not provide an explicit right/wrong judgment. It only contained the correct answer so as to focus students' attention on the problem solution as opposed to whether the learner was right or wrong. Third, the posttest was administered the next day, rather than immediately following the task, so as to measure more stable knowledge change.

In the experiment, we examined the impact of feedback for children learning to solve math equivalence problems (e.g., $3+4+5=3+ \underline{\quad}$), which require an understanding that both sides of an equation represent the same quantity. We hypothesized that feedback would result in higher posttest scores relative to no feedback, particularly for children with low prior knowledge.

Method

Participants

Initial participants were 88 second-grade children from one public and one private school. Of these children, 77 met criteria for participation because they scored below 80% on a problem-solving screening measure. This ensured that children had room to learn from the

intervention. Data from two children were excluded for failing to complete all activities. The final sample contained 75 children (M age = 8.2 yrs, $range = 7.4-9.2$ yrs; 41 girls; 34 boys).

Design

The study had a between-subjects design with children randomly assigned to conditions: no feedback ($n=24$), immediate feedback ($n=25$), and summative feedback ($n=26$). There were no significant differences between conditions in terms of age or gender, $ps>.4$.

Materials

Screening Measure. The screening measure was three tasks that tap understanding of math equivalence (from McNeil et al., 2011). For *equation solving*, children solved five math equivalence problems. For *equation encoding*, children reconstructed four math equivalence problems after viewing each for five seconds to assess how they mentally represented the structure of the problem. They received one point for each accurate reconstruction (up to four points). For *defining the equal sign*, children provided a written definition of the equal sign and received one point if they provided a relational definition (e.g., “the same amount”).

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

Self-Assessment. We obtained children’s subjective ratings of self-assessment to explore whether they considered their performance to reflect negatively on their traits. We adapted a measure used with kindergarten students in Kamins and Dweck (1999). Children responded to each of four items (e.g., “The problem-solving task made me feel like I was a smart student.”) on

a 4-point scale ranging from *strongly disagree* to *strongly agree*. Scores on each item were averaged to form a single score out of four ($\alpha=.80$).

Posttest. The posttest, adapted from past work (Rittle-Johnson et al., 2011), was a broader measure that assessed equation-solving success. It included 8 items that assessed children's use of correct strategies to solve math equivalence problems ($\alpha=.87$). Half of the items were similar to those presented during the intervention, and half differed on a key problem feature, such as inclusion of subtraction (i.e., $5+6-3=5+ _$). We administered several additional measures that were not informative for the current results and are thus not discussed further.

Coding. We coded children's problem-solving strategies on the screening measure, intervention problems, and posttest, based on their numerical answers (e.g., for the problem $2+7=6+ _$, an answer of 15 indicated an incorrect "add all" strategy and an answer of 3 indicated a correct strategy). Responses within one of the correct answer were coded as correct. To establish inter-rater reliability, a second rater (who was blind to the initial strategy code) coded children's strategies on 30% of all problems. Inter-rater agreement was high ($\kappa=.98$).

Procedure

Children completed the screening measure in their classrooms in a 10-minute session. Those who met the inclusion criteria then completed a one-on-one tutoring intervention in a single session lasting approximately 25 minutes. This session was conducted in a quiet area at the school with one of two trained tutors, and included strategy instruction and problem solving.

Strategy Instruction. All children received instruction on a correct problem-solving strategy with four math equivalence problems with the blank at the end (e.g., $3+4+2=3+ _$). Children were instructed on the commonly used equalize strategy, which involves adding the numbers on one side of the equal sign and then counting up from the number on the other side to

get the same amount. Children were asked to answer questions to ensure they were attending to instruction. To check that the instruction worked, all children then solved a math equivalence problem on their own. If they solved the problem correctly, they proceeded to the next activity. If they solved it incorrectly, instruction on the equalize strategy was repeated without revealing the correct answer, and they were asked to solve another problem until they solved one correctly (although the children were not told this criteria). We set the protocol such that after three failed attempts the experiment was discontinued and children received more remedial tutoring. Two children were excluded for this reason resulting in a sample of 73 children.

Problem Solving. Children were then asked to solve 12 math problems presented one at a time on a computer (see *Materials* for a description of the problems). First, the tutor described the task and whether or not the computer would provide feedback. Second, during the problem-solving task, the tutor's presence was removed. Specifically, children were told they would work on the computer by themselves for this portion of the session so that they could work at their own pace and not worry about the tutor. Then, the tutor sat a short distance away and engaged in a different task (e.g., read a book) until the child had completed the problem-solving task.

In the *no-feedback* condition, children did not receive feedback and were simply directed to click when they were ready for the next problem. In the *immediate-feedback* condition, children received trial-by-trial correct-answer feedback after each problem. In the *summative-feedback* condition, children received correct-answer feedback after all 12 problems had been solved. The problems with the child's answers reappeared one at a time on the computer screen along with the feedback message. The problems and correct answers remained on the screen while the next problem appeared (up to four problems at a time), which allowed some spontaneous comparison across problems during the provision of feedback, much like a

summative answer key. In both feedback conditions, the feedback was presented visually on the computer screen for each problem (e.g., “10 is the correct answer.”). Importantly, it did not contain an explicit right/wrong verification, it only contained the correct answer. Although the right/wrong judgment was implicit (via comparison of the child’s answer with the correct answer), there were no explicit signals from the computer (e.g., check mark, error noise, etc.).

Following problem solving, children rated their self-assessment, and then returned to class. The following day, children completed the posttest. Due to absences, one child in the no-feedback condition completed the posttest 2 days later and two children in the immediate-feedback condition completed the posttest 4 and 5 days later.

Data Analysis

Children’s scores on most measures were analyzed using ANCOVAs. Scores that were not normally distributed were converted into a dichotomous outcome (i.e., 1 for success and 0 otherwise) and logistic regression was used to predict the odds of success. In all models, condition was the dependent variable. It was dummy coded with immediate feedback and summative feedback entered into the models, and no feedback as the reference group. Children’s age and score on the screening measure were entered as covariates. To test if the effects of condition depended on prior knowledge, the interaction between condition and screening measure was included. Thus, each model included two condition variables, age (mean centered), screening measure score (mean centered), and two condition by screening measure interactions.

Results

Screening Measure

On average, children in the final sample ($n=73$) solved 1.4 ($SD=1.3$) problems correctly (out of 5), encoded 1.8 ($SD=1.4$) problems correctly (out of 4), and 14% provided a relational

definition of the equal sign. Performance did not differ by condition on any task, $ps > .5$. As in prior work (McNeil et al., 2011; Fyfe & Rittle-Johnson, 2015), we created a composite measure by summing z-scores across the three tasks ($M=0.00$, $SD=2.09$, $range=-2.75$ to 6.15). The composite score served as a prior knowledge covariate in subsequent analyses.

Intervention Problems

We analyzed performance on the intervention problems to examine how feedback impacted the on-going task. Across all 12 problems the frequency of correct strategy use was similar in the immediate-feedback ($M=80\%$, $SD=19\%$), summative-feedback ($M=78\%$, $SD=29\%$), and no-feedback ($M=73\%$, $SD=29\%$) conditions. An ANCOVA revealed no main effect of immediate-feedback, $p=.16$, or summative-feedback, $p=.14$, relative to no-feedback. There was no summative-feedback by prior knowledge interaction, $p=.40$; but, there was a marginal immediate-feedback by prior knowledge interaction, $F(1, 66) = 3.83$, $p = .055$, $\eta_p^2 = .06$.

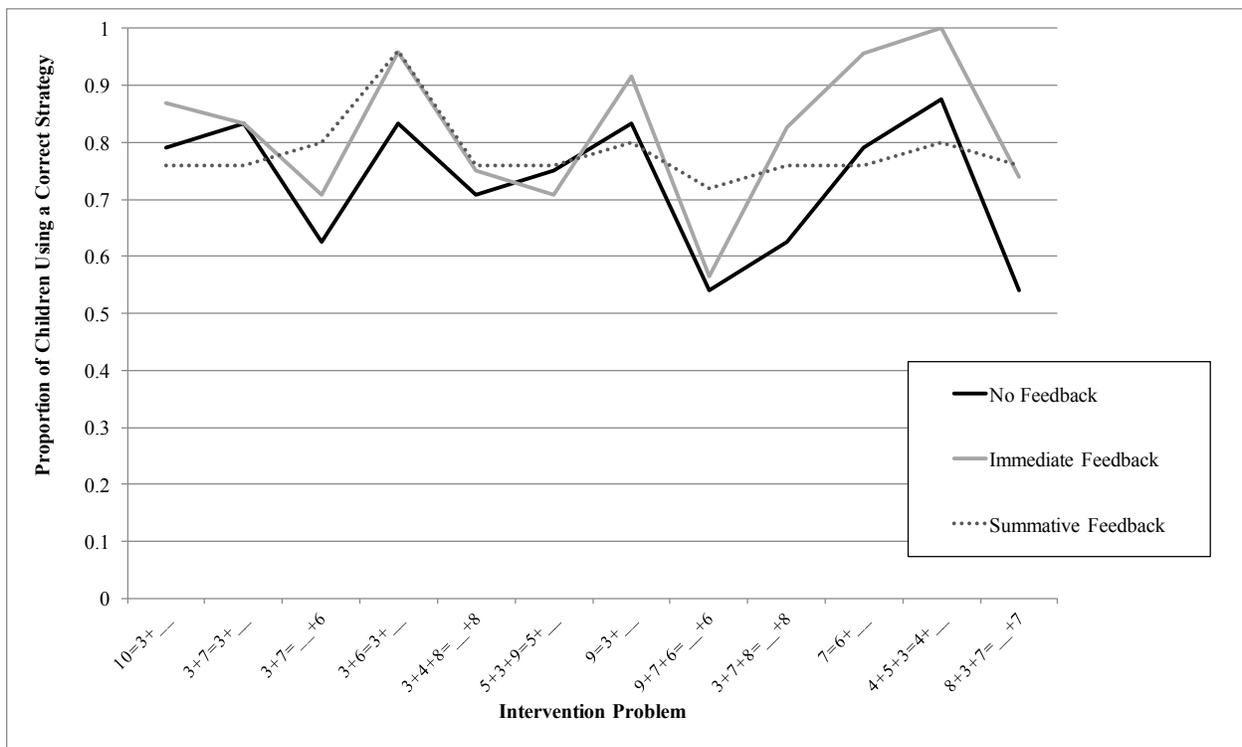
To follow up the interaction, we tested the effect of immediate-feedback for children with lower and higher prior knowledge. Screening measure scores were centered at one standard deviation below the mean in one ANCOVA and one standard deviation above the mean in a second ANCOVA (see Aiken & West, 1991). For children with low prior knowledge, there was a main effect of immediate-feedback, $F(1, 66) = 5.50$, $p = .02$, $\eta_p^2 = .08$. The model estimates a ~20% differential between immediate- and no-feedback ($M = 71\%$ vs. 50%). But, for children with high prior knowledge, there was no main effect of immediate-feedback, $F(1, 66) = 0.11$, $p = .74$, $\eta_p^2 = .00$. The model estimates a ~3% differential favoring no-feedback ($M = 89\%$ vs. 92%). Thus, immediate trial-by-trial feedback boosted the problem-solving performance of children with low prior knowledge, but had minimal impact for children with higher prior knowledge.

Descriptively, a trial-by-trial plot of intervention performance suggests differences were strongest on the last few items (see Figure 1). Indeed, if we split the intervention problems into three blocks (first four problems, middle four problems, and last four problems), significant effects only emerged on the last block. On these last four problems, the raw frequency of correct strategy use was high in the immediate-feedback condition ($M=88\%$, $SD=18\%$), but lower in the summative- ($M=77\%$, $SD=35\%$), and no-feedback ($M=71\%$, $SD=33\%$) conditions. There was a significant main effect of immediate-feedback, $F(1, 66) = 6.87, p = .01, \eta_p^2 = .10$, and no effect of summative-feedback, $p=.13$. Again, a significant immediate-feedback by prior knowledge interaction, $F(1, 66) = 6.16, p = .02, \eta_p^2 = .09$, indicated that the positive effect of immediate feedback was present for low-knowledge, but not high-knowledge, children. Immediate feedback seemed to help children maintain strong performance on difficult problems. For example, children who did not receive feedback during the task scored higher on the first, simpler block of problems than on the last, more difficult block of problems ($M = 80\%$ vs. 74%). However, children who received trial-by-trial feedback maintained high scores from the first block to the last block ($M = 85\%$ vs. 88%), despite the increase in problem difficulty.

Although most children used a correct strategy on the majority of problems, their errors were informative for examining the types of incorrect strategies they employed. Across all children, the most common incorrect strategies were to add the numbers before the equal sign (11% of all trials) and to add all the numbers in the problem (5% of all trials). For example, for the problem $3 + 7 = _ + 6$, an answer of 10 indicates an add-before-equal sign strategy and an answer of 16 indicates an add-all-numbers strategy. Children in the immediate-feedback condition used these two incorrect strategies less often ($M=10\%$, $SD=27\%$) than children in the summative- ($M=17\%$, $SD=26\%$) and no-feedback conditions ($M=21\%$, $SD=27\%$). An ANCOVA

confirmed this revealing a main effect of immediate-feedback, $F(1, 66) = 4.21, p = .04, \eta_p^2 = .06$, no main effect of summative-feedback, $p = .19$, and no interactions with prior knowledge, $ps > .10$. Overall, these two strategies accounted for 61% of all the errors made. This was particularly true for children who did not receive feedback during the task (accounting for 68% of errors in summative- and no-feedback, but only 46% of errors in immediate-feedback).

Figure 1. *Trial-by-trial performance on each intervention problem by condition*



Self-Assessment

Children's ratings of positive self-assessment after the problem-solving task were similar in the immediate-feedback ($M=3.5$ out of 4, $SD=0.4$) and no-feedback ($M=3.4$, $SD=0.4$) conditions, but lower in the summative-feedback condition ($M=3.1$, $SD=0.5$). An ANCOVA revealed no main effect of immediate-feedback, $p=.73$, but a significant main effect of summative-feedback relative to no-feedback $F(1, 66) = 4.33, p = .04, \eta_p^2 = .06$. There were no

condition-by-prior knowledge interactions, $ps > .14$. A follow-up analysis revealed a significant difference between the two feedback types, $F(1, 66) = 6.08, p = .02, \eta_p^2 = .09$.

Across all children, ratings of self-assessment were not correlated with performance on the intervention problems, $r(71) = .14, p = .26$, or on the posttest, $r(71) = .10, p = .41$. Looking by condition, the correlations were negligible in the no-feedback condition (intervention problems, $r(22) = .08, p = .72$, and posttest, $r(22) = .01, p = .95$) and in the immediate-feedback condition (intervention, $r(22) = -.08, p = .72$, and posttest, $r(22) = -.09, p = .69$). However, the correlations were moderate, though non-significant, in the summative-feedback condition (intervention problems, $r(23) = .30, p = .14$, and posttest, $r(23) = .30, p = .14$). This suggests the ratings of self may have played a larger role when the feedback was provided all at once.

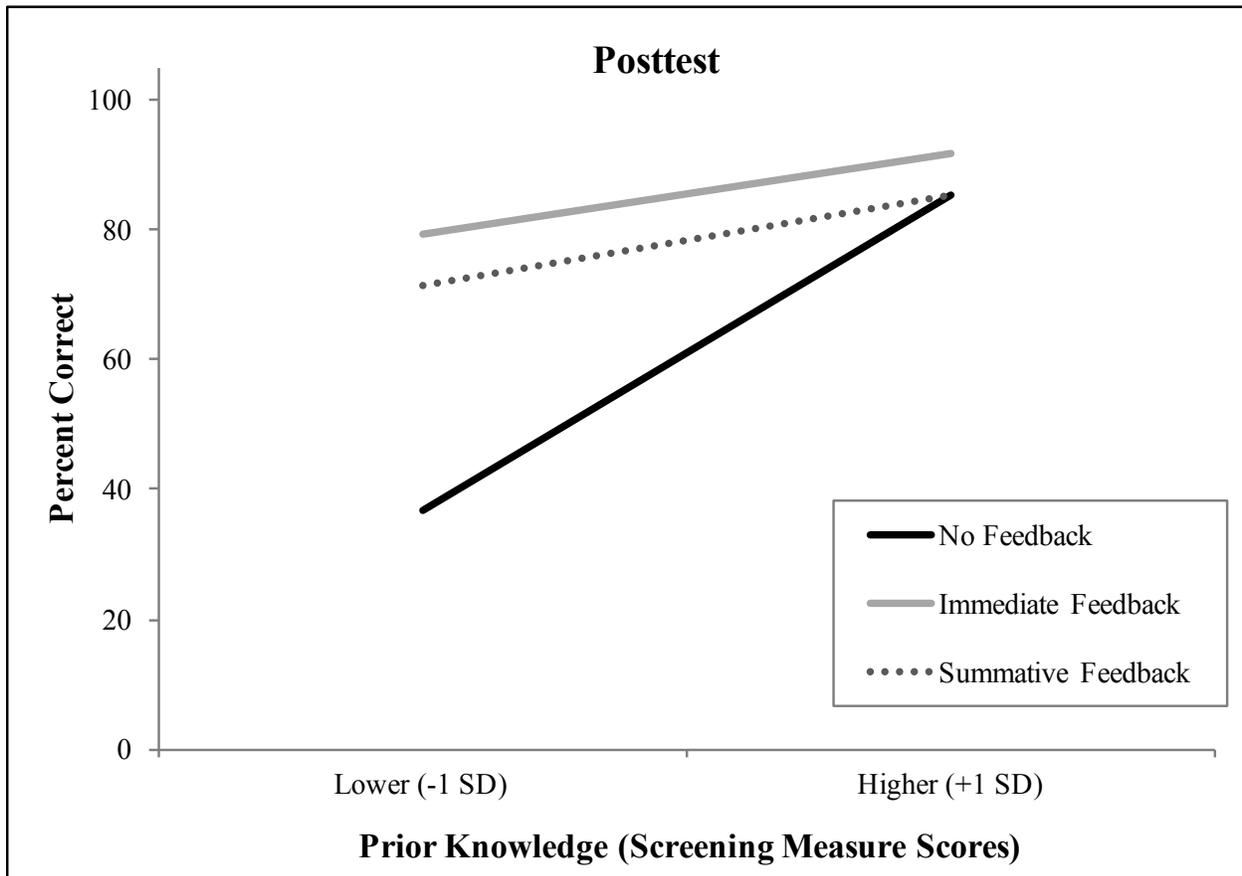
Posttest

On the next-day posttest, percent correct was highest with immediate-feedback ($M=86\%$, $SD=22\%$), lower with summative-feedback ($M=78\%$, $SD=27\%$), and lowest with no-feedback ($M=65\%$, $SD=38\%$). An ANCOVA revealed main effects of immediate-feedback, $F(1, 66) = 10.02, p = .002, \eta_p^2 = .13$, and summative-feedback, $F(1, 66) = 5.06, p = .03, \eta_p^2 = .07$, relative to no-feedback. However, prior knowledge interacted with both immediate-feedback, $F(1, 66) = 5.78, p = .02, \eta_p^2 = .08$, and summative-feedback, $F(1, 66) = 4.56, p = .04, \eta_p^2 = .07$.

To follow up the interactions, two ANCOVAs tested the effect of condition for children with lower and higher prior knowledge (see Figure 2). For children with low prior knowledge, there were large effects of both immediate-feedback, $F(1, 66) = 14.61, p < .000, \eta_p^2 = .18$, and summative-feedback, $F(1, 66) = 9.46, p = .003, \eta_p^2 = .13$. The model estimates a ~35–45% differential between the feedback conditions and no-feedback (see Figure 2). But, for children with high prior knowledge, there were no main effects of immediate-feedback, $F(1, 66) = 0.39, p$

= .53, $\eta_p^2 = .01$, or summative-feedback, $F(1, 66) = 0.00$, $p = .99$, $\eta_p^2 = .00$. The model estimates a 0–5% differential between the feedback conditions and no feedback (see Figure 2).

Figure 2. Posttest scores by condition and prior knowledge.



Note. Estimated scores are plotted at plus/minus one standard deviation from the mean.

However, scores on the posttest were high and not normally distributed as the majority of children (75%) solved more than half of the items correctly with a full 47% of children solving all of the items correctly. We used logistic regression to predict the log of the odds of scoring 100% correct. More children in the immediate-feedback condition (63%) scored 100% on the posttest relative to the summative-feedback (40%) and no-feedback (38%) conditions. There was

a main effect of immediate-feedback, $\beta = 1.74$, $p = .03$, $OR = 5.72$, but no effect of summative-feedback, $\beta = 0.49$, $p = .48$, $OR = 1.63$, relative to no-feedback. A follow-up analysis revealed a marginal difference between the two feedback types, $\beta = 1.25$, $p = .08$, $OR = 3.50$. Prior knowledge did not interact with immediate-feedback or summative-feedback, $ps > .10$.

The results are similar, though less robust, if we use less stringent criteria. For example, in a logistic regression predicting the log of the odds of scoring 75% or higher, there is a marginal effect of immediate-feedback, $\beta = 1.79$, $p = .08$, $OR = 6.01$, no effect of summative-feedback, $\beta = 0.51$, $p = .49$, $OR = 1.66$, and no interactions with prior knowledge, $ps > .15$.

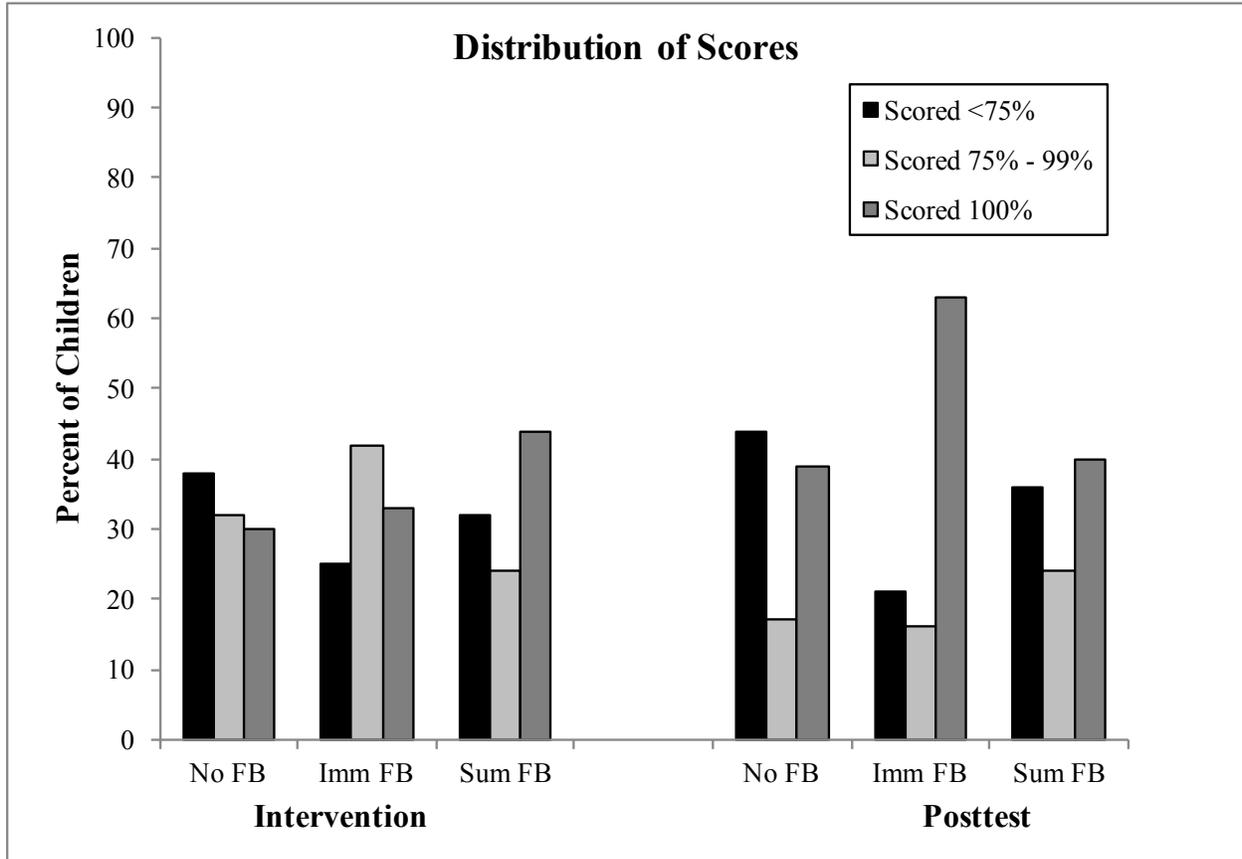
Together, these results suggest that immediate and summative feedback improve posttest scores relative to no feedback for low-knowledge learners. However, only immediate feedback improves mastery of the material, and this is true for both low- and high-knowledge learners. To better understand these condition effects, we further explored performance in two ways. First, we split the posttest into two subscales: learning and transfer (see Table 1). On the learning items, scores were similar with immediate- ($M = 88\%$, $SD = 19\%$) and summative-feedback ($M = 85\%$, $SD = 27\%$), and both were higher than no-feedback ($M = 72\%$, $SD = 35\%$). But, on the transfer items, scores were highest with immediate-feedback ($M = 83\%$, $SD = 29\%$), lower with summative-feedback ($M = 70\%$, $SD = 36\%$), and lowest with no-feedback ($M = 58\%$, $SD = 45\%$). Indeed, for learning scores, there were significant main effects of both immediate-feedback, $F(1, 66) = 6.52$, $p = .01$, $\eta_p^2 = .09$, and summative-feedback, $F(1, 66) = 5.60$, $p = .02$, $\eta_p^2 = .08$. For transfer scores, the main effect was significant for immediate-feedback, $F(1, 66) = 8.01$, $p = .006$, $\eta_p^2 = .11$, but not for summative-feedback, $F(1, 66) = 2.39$, $p = .13$, $\eta_p^2 = .04$.

Table 1

Posttest Learning Subscale ($\alpha = .72$)	Posttest Transfer Subscale ($\alpha = .86$)
Structurally similar to problems presented during the intervention	Contained novel feature: new position of the blank or included subtraction
$8 = 6 + \underline{\quad}$	$\underline{\quad} + 2 = 6 + 4$
$3 + 4 = \underline{\quad} + 5$	$8 + \underline{\quad} = 8 + 6 + 4$
$3 + 7 + 6 = \underline{\quad} + 6$	$5 + 6 - 3 = 5 + \underline{\quad}$
$7 + 6 + 4 = 7 + \underline{\quad}$	$5 - 2 + 4 = \underline{\quad} + 4$

Second, we compared the distribution of scores on the intervention problem-solving task and on the posttest (see Figure 3). Descriptively, an interesting pattern emerged such that the biggest change occurred in the immediate-feedback condition. Immediate feedback seemed to push children who exhibited moderate performance during the intervention toward levels of mastery on the posttest. In contrast, for the summative-feedback condition, the distribution of scores was nearly identical at intervention and posttest. There was some change in the no-feedback condition such that the extremes were more common (relatively low or relatively high performance) at posttest. These patterns suggest that children who did not receive feedback during the intervention task either got it or not – and this was reflected on the posttest as well. However, children who received immediate trial-by-trial feedback were able to learn during the task and exhibit their better understanding of the problems on the next-day posttest.

Figure 3. Distribution of scores by condition at intervention and at posttest.



Note. No FB = No-feedback condition. Imm FB = Immediate-feedback condition. Sum FB = Summative-feedback condition.

Discussion

The results from the current study show positive effects of computer-generated feedback on mathematics learning for children who received prior instruction on the problems. During the tutoring session, immediate feedback boosted problem-solving performance and decreased the use of common incorrect strategies for low-knowledge children. Immediate feedback did not impact ratings of self-assessment, but summative feedback led to lower ratings of self-assessment. On the next-day posttest, both immediate and summative feedback resulted in higher

equation-solving success than no feedback for low-knowledge children. However, only immediate feedback facilitated mastery for both low- and high-knowledge children.

Despite increasing evidence in favor of delaying the presentation of feedback, the present findings indicate that immediate feedback may be more effective for promoting children's mathematics problem solving (see also Brosvic et al., 2006). In addition to facilitating learning, immediate feedback facilitated transfer to novel problems and mastery of the material. Several pieces of evidence suggest that these benefits may be attributable to the trial-by-trial nature of the feedback and children's opportunity to use the feedback on subsequent problems. First, immediate feedback improved performance on the last block of training problems, but not the first block. This suggests children were able to learn from the initial feedback and maintain high performance on the later, more difficult problems. Second, immediate feedback reduced the use of the incorrect 'add-all' and 'add-to-equal' strategies. That is, when children in the immediate feedback condition made errors, they were less likely to use common incorrect strategies that stem from entrenched misconceptions (McNeil & Alibali, 2005) and more likely to try something different. Third, only immediate feedback resulted in a big change in the distribution of children at mastery. This suggests a progression of knowledge such that children started out variable, learned from the feedback, and were able to exhibit mastery by the posttest. This is in contrast to the no-feedback and summative-feedback groups which were more stable.

Indeed, with one exception, children in the summative-feedback condition performed similarly to children in the no-feedback condition, indicating that summative feedback had minimal benefits. One reason may have been the lack of an opportunity to use the feedback right away. Unlike immediate feedback, there was no subsequent problem on which to try a new strategy until the next-day posttest. Another reason may have been increased attention on the

self. Summative feedback led to lower ratings of positive self-assessment relative to no feedback and immediate feedback. Further, these ratings in the summative feedback condition were moderately, negatively correlated with posttest performance (although these correlations did not reach standard levels of significance). Thus, it seems possible that receiving feedback on all of the problems at once increased attention on the self, produced a negative affective response (e.g., ego-threat), and influenced children's general confidence or motivation to learn.

Although these two reasons may play a role in explaining why summative feedback was minimally beneficial, there are other possibilities related to the design of the current experiment. First, the summative feedback may have been less visually and cognitively overwhelming if each problem had been presented individually (as opposed to remaining on the screen). Second, summative feedback may have been more effective given a longer retention interval between the instructional session and posttest. Indeed, Butler et al. (2007) found that the benefits of delayed feedback over immediate feedback emerged on a one-week posttest, but not on a one-day posttest. Further, experiments with motor tasks have consistently found that while immediate feedback results in more efficient learning, delayed feedback results in improved retention and fewer errors on subsequent assessments (Schmidt & Bjork, 1992). Third, summative feedback in the current study was still somewhat immediate in the sense that it was given right after the practice task. Several studies have found advantages to providing summative feedback the next day rather than after the task (e.g., Bardwell, 1981; Butler et al., 2007; but see Dihoff et al., 2004). More generally, future research needs to parse the differences between trial-by-trial versus summative and immediate versus delayed. Often, "delayed" feedback is relative to the conditions used in a particular experiment making it difficult to compare across studies.

Future research should also continue to examine how feedback relates to learners' prior knowledge. The current findings are consistent with previous work demonstrating that feedback often has stronger, positive effects for children with low prior knowledge on a pretest than children with higher prior knowledge. For example, feedback on middle-school students' writing assignments improved performance for learners with low prior knowledge on a pretest, but not for learners with higher prior knowledge (Gielen et al., 2010). Similarly, undergraduate students with high prior knowledge of statistics performed just as well on a posttest whether they received correct-answer feedback during training or not (Krause et al., 2009). Indeed, the biggest benefits of feedback in this study occurred for low-knowledge children during training and at posttest.

However, these results are slightly at odds with our previous study. In Fyfe and Rittle-Johnson (2015), children who received prior instruction exhibited better learning without feedback. Further, this negative effect of feedback did not depend on children's prior knowledge on the pretest. Here, children who received prior instruction exhibited better learning with feedback, though this was more pronounced for children with low prior knowledge. Why were the effects of feedback positive in this study? One possibility is that feedback is more effective in environments that reduce attention on the self. There are three critical differences between the current study and the previous study: (1) the presence of the human tutor during problem solving was removed and feedback was provided solely by the computer, (2) the feedback message only contained the correct answer; it did not contain an explicit right/wrong judgment, and (3) the posttest was administered the following day as opposed to immediately after the learning task. The first two changes reduce attention to self and evaluation, while the third change allows for any effects that may arise from attention on the self to dissipate prior to testing. These differing results are potentially consistent with the hypothesis that feedback will lead to larger gains when

it draws attention to the task as opposed to the self (Kluger & DeNisi, 1996), though future experimental research is needed to test the causal nature of this claim.

In conclusion, the present study provides experimental evidence for the benefits of computer-generated feedback on children's mathematics learning. Specifically, for children given instruction on the target problems, both immediate and summative feedback resulted in higher performance on a next-day posttest relative to a no-feedback control for children with lower prior knowledge. Further, immediate feedback increased transfer and mastery of the material for both low- and high-knowledge children. These results reinforce the notion that even minimal feedback on a computer can be a powerful form of guidance during problem solving.

References

- Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.
- Azevedo, R., & Bernard, R. M. (1995). A meta-analysis of the effects of feedback in computer-based instruction. *Journal of Educational Computing Research, 13*, 111-127.
- Bangert-Drowns, R., Kulik, C., Kulik, J., & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research, 61*, 213.
doi:10.2307/1170535
- Bardwell, R. (1981). Feedback: How does it function? *Journal of Experimental Education, 50*, 4-9.
- Beeson, R. O. (1973) Immediate knowledge of results and test performance. *The Journal of Educational Research, 66*, 224-226.
- Bohlmann, N., & Fenson, L. (2005). The effects of feedback on perseverative errors in preschool aged children. *Journal of Cognition and Development, 6*, 119-131.
doi:10.1207/s15327647jcd0601_7
- Brosvic, G., Dihoff, R., Epstein, M., & Cook, M. (2006). Feedback facilitates the acquisition and retention of numerical fact series by elementary school students with mathematics learning disabilities. *The Psychological Record, 56*, 35-54.
- Butler, A., & Roediger, H. (2008). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Memory & Cognition, 36*, 604-616.
doi:10.3758/MC.36.3.604

- Butler, A., Karpicke, J., & Roediger, H. (2007). The effect of type and timing of feedback on learning from multiple-choice tests. *Journal of Experimental Psychology: Applied*, *13*, 273–281. doi:10.1037/1076-898X.13.4.273
- Clariana, R. B., Wagner, D., & Roher Murphy, L. C. (2000). Applying a connectionist description of feedback timing. *Educational Technology Research and Development*, *48*, 5–22. doi:10.1007/BF02319855
- Dihoff, R., Brosvic, G., & Epstein, M. (2003). The role of feedback during academic testing: The delay retention test revisited. *The Psychological Record*, *53*, 533-548.
- Dihoff, R. E., Brosvic, G. M., Epstein, M. L., & Cook, M. J. (2004). Provision of feedback during preparation for academic testing: Learning is enhanced by immediate but not delayed feedback. *The Psychological Record*, *54*, 207–231.
- Fyfe, E., & Rittle-Johnson, B. (2015). Feedback both helps and hinders learning: The causal role of prior knowledge. *Journal of Educational Psychology*. doi:10.1037/edu0000053
- Fyfe, E., Rittle-Johnson, B., & DeCaro, M. (2012). The effects of feedback during exploratory mathematics problem solving: Prior knowledge matters. *Journal of Educational Psychology*, *104*, 1094-1108. doi:10.1037/a0028389
- Gaynor, P. (1981). The effect of feedback delay on retention of computer-based mathematical material. *Journal of Computer-Based Instruction*, *8*, 28-34.
- Gielen, S., Peeters, E., Dochy, F., Onghena, P., & Struyven, K. (2010). Improving the effectiveness of peer feedback on learning. *Learning and Instruction*, *20*, 304-315. doi:10.1016/j.learninstruc.2009/08.007
- Greaves, T., & Hayes, J. (2008). *America's digital schools 2008: Six trends to watch*. Shelton, CT: MDR.

- 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.
- Hays, M. J., Kornell, N., & Bjork, R. A. (2010). The costs and benefits of providing feedback during learning. *Psychonomic Bulletin & Review*, *17*, 797–801.
doi:10.3758/PBR.17.6.797
- Kamins, M., & Dweck, C. (1999). Person versus process praise and criticism: Implications for contingent self-worth and coping. *Developmental Psychology*, *35*, 835-847.
doi:10.1037/0012-1649.35.3.835
- Karabenick, S., & Knapp, J. (1988). Effects of computer privacy on help seeking. *Journal of Applied Social Psychology*, *16*, 461-72.
- Kluger, A., & 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., Stark, R., & Mandl, H. (2009). The effects of cooperative learning and feedback on e-learning in statistics. *Learning and Instruction*, *19*, 158-70.
doi:10.1016/j.learninstruc.2008.03.003
- Kulhavy, R. (1977). Feedback in written instruction. *Review of Educational Research*, *47*, 211-232. doi:10.2307/1170128
- Kulik, J. A., and Kulik, C. C. (1988). Timing of feedback and verbal learning. *Review of Educational Research*, *58*, 79-97.
- McNeil, N., Fyfe, E., Petersen, L., Dunwiddie, A., & Brletic-Shiple, H. (2011). Benefits of practicing $4=2+2$: Nontraditional problem formats facilitate children's understanding of

- mathematical equivalence. *Child Development*, 82, 1620-1633. doi:10.1111/j.1467-8624.2011.01622.x
- Metcalfe, J., Kornell, N., & Finn, B. (2009). Delayed versus immediate feedback in children's and adults' vocabulary learning. *Memory & Cognition*, 37, 1077-1087. doi:10.3758/MC.37.8.1077
- Mory, E. (2004). Feedback research revisited. In D. Jonassen (Ed.), *Handbook of research on educational communications and technology* (2nd. pp. 745-783). Mahwah, NJ: Erlbaum.
- Mullet, H., Butler, A., Verdin, B., von Borries, R., & Marsh, E. J. (2014). Delaying feedback promotes transfer of knowledge despite student preferences to receive feedback immediately. *Journal of Applied Research in Memory and Cognition*, 3, 222-229. doi:10.1016/j.jarmac.2014.05.001
- Nakata, T. (2015). Effects of feedback timing on second language vocabulary learning: Does delaying feedback increase learning? *Language Teaching Research*, 19, 416-434. doi:10.1177/1362168814541721
- Pashler, H., Cepeda, N., Wixted, J., & Rohrer, D. (2005). When does feedback facilitate learning of words? *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 31, 3-8. doi:10.1037/0278-7393.31.1.3
- Rittle-Johnson, B., Matthews, P., Taylor, R., & McEldoon, K. (2011). Assessing knowledge of mathematical equivalence: A construct modeling approach. *Journal of Educational Psychology*, 103, 85-104. doi:10.1037/a0021334
- Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science*, 3, 207-217. doi:10.1111/j.1467-9280.1992.tb00029.x

- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78, 153–189.
doi:10.3102/0034654307313795
- Skinner, B. (1954). The science of learning and the art of teaching. *Harvard Educational Review*, 24, 86-97.
- Smith, T. A., & Kimball, D. R. (2010). Learning from feedback: Spacing and the delay-retention effect. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 36, 80–95. doi:10.1037/a0017407
- Smits, M., Boon, J., Sluijsmans, D., & Van Gog, T. (2008). Content and timing of feedback in a web-based learning environment: Effects on learning as a function of prior knowledge. *Interactive Learning Environments*, 16, 183-193.