Abstract Title Page

Not included in page count.

Title: The Timing of Feedback on Mathematics Problem Solving in a Classroom Setting

Author(s): Emily R. Fyfe and Bethany Rittle-Johnson, Vanderbilt University

Abstract Body

Limit 4 pages single spaced.

Background / Context:

Description of prior research and its intellectual context.

Feedback is a ubiquitous learning tool that is theorized to help learners detect and correct their errors (Kulhavy, 1977). It is defined as any information about performance that the learner can use to modify prior knowledge (Mory, 2004). Generally, feedback is assumed to be helpful and providing feedback is a recommended practice (Alfieri et al., 2011). Although feedback often improves learning, in some cases it has neutral or negative effects (Kluger & DeNisi, 1996). Some work has been done to understand this variability, but mostly with adults studying test-like material (e.g., cued recall of word pairs) in laboratory settings. We focus on the effects of feedback during problem solving for elementary school students in mathematics classrooms.

Previous research suggests that the effects of feedback may depend on students' prior knowledge and experience (e.g., Krause et al., 2009; Luwel et al., 2011). For example, in two one-on-one tutoring experiments, elementary school children solved novel math problems (Fyfe, Rittle-Johnson, & DeCaro, 2012). Some children received feedback and others did not. For low-knowledge students, feedback facilitated posttest problem solving. But, for students with moderate prior knowledge, feedback hindered accuracy at posttest relative to no feedback. Thus, feedback may negatively impact students before they exhibit mastery in the target domain.

Although informative, there are issues with previous studies that limit their implications for theory and practice. First, it is unclear whether the varying effects of feedback are present in more typical classroom settings. For example, feedback may have stronger effects in one-on-one settings in which students are more likely to attend to feedback and use it to change performance.

Second, it is unclear how teachers can identify which students will benefit from feedback. Previous studies have relied on domain-specific researcher-created pretests to classify low- and high-knowledge learners. One solution is to use student's grade as an indicator of prior knowledge. Grade is a salient characteristic that teachers can use to categorize students. It captures a broader range of math exposure and general math knowledge. Also, modifying the provision of feedback at the grade level is more feasible than at the individual student level.

Finally, it is unclear whether all types of feedback would lead to both positive and negative effects. In previous tutoring studies, feedback was presented immediately after each problem. The immediacy of the feedback may have been responsible for its positive effects. For example, trial-by-trial feedback helped low-knowledge learners quickly detect their error on one problem and generate a correct strategy on the following problem (Fyfe et al., 2012). At the same time, immediate feedback may trigger cognitive and emotional processes for students with higher prior knowledge that interfere with learning. For example, students with more knowledge may feel pressure to solve problems correctly. Feedback that states otherwise may threaten the students' self-esteem and draw attention away from the task at hand (Leary et al., 2009).

Summative feedback, which is delayed until all the problem have been solved, provides a promising alternative that is viable in a classroom setting. Because delayed feedback occurs after the problems have been solved, it may reduce the negative effects of immediate feedback that are due to the redirection of attention during problem solving. However, it may also reduce the positive impact of feedback on strategy generation. Here, we contrast immediate, summative, and no feedback for second- and third-grade students in a classroom setting.

Purpose / Objective / Research Question / Focus of Study:

Description of the focus of the research.

The goal of this study was to examine the effects of feedback in a classroom context for children solving math equivalence problems (problems with operations on both sides of the equal sign). Math equivalence is a fundamental concept in arithmetic and algebra and young children often struggle to solve math equivalence problems correctly (e.g., McNeil, 2008). Unfortunately, elementary curricula do not typically include math equivalence problems (Powell, 2012). In the current study, we taught all students a procedure for solving these problems to ensure students had some exposure to the problems prior to problem solving. During the problem-solving task, we manipulated the presence and timing of feedback. We focused on children in two grade levels to explore the effects of feedback for children with varying levels of knowledge. We expected immediate feedback to have positive effects relative to no feedback for second-graders, but neutral or negative effects for third-graders. In contrast, we expected summative feedback to have positive effects for third-graders, but neutral or negative effects for second-graders.

Setting:

Description of the research location.

We worked with children in 7 second-grade and 7 third-grade classrooms from two suburban elementary schools (one public and one private) in Middle Tennessee.

Population / Participants / Subjects:

Description of the participants in the study: who, how many, key features, or characteristics.

All second- and third-grade children were invited to participate, with a total of 274 children completing an initial screening measure. Data from 31 children were excluded for failing to complete all intervention activities. The final sample contained 243 six-to ten-year-old children (*M* age = 8.3 yrs; 44% Female; 74% White). Close to half were third-graders (*n* = 136) and the remaining students were second-graders (*n* = 107).

Intervention / Program / Practice:

Description of the intervention, program, or practice, including details of administration and duration.
For Track 2, this may include the development and validation of a measurement instrument.

At the beginning of the intervention, children in all conditions received instruction on a correct problem-solving strategy with four math equivalence problems. 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. The instructor demonstrated the procedure on all four problems. She also posed questions for each problem, attempting to call on each child at least once. To ensure the instruction worked, all children then solved two math equivalence problems on their own. After each problem, the instructor restated the correct problem-solving strategy, but did not provide the correct answers.

Children were then asked to solve 12 math problems (e.g., $3 + 7 = 6 + \_\_$) on their own. Each child received his/her own problem-solving packet. In the *no-feedback* condition, children worked on the problems at their own pace and did not receive any feedback on their answers. In the *immediate-feedback* condition, children worked on the problems one at a time and received correct-answer feedback from the instructor after each problem. In the *summative-feedback* condition, children worked on the problems at their own pace and received correct-answer feedback from the instructor after all students had solved all the problems.

Research Design:

Description of the research design.

The study had a pretest-intervention-posttest design. The intervention occurred in children's classrooms during their normal mathematics instruction. Within each classroom, children were randomly assigned to one of three conditions: (1) immediate feedback (n = 82), (2) summative feedback (n = 81), or (3) no feedback (n = 80). Each grade had approximately the same number of children assigned to each condition (second grade: immediate, n = 35; summative, n = 36; no, n = 36; third grade: immediate, n = 47; summative, n = 45; no, n = 44). Research assistants worked with small groups of children assigned to the same condition, allowing for children within the same classroom to be assigned to different conditions.

Data Collection and Analysis:

Description of the methods for collecting and analyzing data.
For Track 2, this may include the use of existing datasets.

Children completed the screening measure in their classrooms in a 10-minute session. The screening measure, taken from McNeil et al., (2011) had children solve four math equivalence problems, define the equal sign, and reconstruct four equations from memory (to measure encoding). Within two weeks, children participated in the intervention during their regularly scheduled math class. Children assigned to the same condition sat in a small group in their classroom and were supervised by a research assistant. Immediately following the intervention, children completed a posttest. The posttest, adapted from past work (Rittle-Johnson et al., 2011), is a broader measure that assesses equation-solving success. It included 8 items that assessed children's accuracy on a wider range of math equivalence problems, including ones with subtraction and ones with the unknown on the left side of the equation (α = .85).

Findings / Results:

Description of the main findings with specific details.

Screening Measure. Second-grade students exhibited low prior knowledge on the screening measure. On average, they solved 0.6 (SD = 1.3) math equivalence problems correctly (out of 4), encoded 1.6 (SD = 1.3) problems correctly (out of 4), and only 18% provided a relational definition of the equal sign. On the other hand, third-grade students exhibited higher prior knowledge. On average, they solved 1.4 (SD = 1.8) problems correctly, encoded 2.5 (SD = 1.3) problems correctly, and 15% provided a relational definition of the equal sign. Indeed, third-grade students outperformed second-grade students on equation solving (p < .001) and equation encoding (p < .001), but not on equal sign definition (p = .52). In line with previous research (Chesney et al., 2014), we created a composite measure of students' performance. Specifically, we summed z-scores from each task. Composite scores ranged from –2.6 to 5.5 (M = 0.0, SD = 1.9). The scores did not differ as a function of condition (p = .58), but differed by grade (p < .001) as third-graders had higher scores than second-graders. Composite scores served as a covariate in subsequent analyses, though the pattern of results was similar if it was not included.

Instruction Check. The strategy instruction was largely successful. Most children (92%) solved both instruction check problems correctly. However, a few solved one (n = 8) or both (n = 11) problems incorrectly, only four of which were third graders. The 11 children who solved both problems incorrectly were excluded from subsequent analyses as their performance suggested they were not ready to learn about or solve these problems. This resulted in a sample of 232 children (no-feedback, n = 77; immediate-feedback, n = 80; summative feedback, n = 75). The pattern of findings is similar if no children are excluded.

Intervention. We examined children's accuracy on the 12 problems presented during the intervention. We conducted an ANCOVA with condition (no feedback, immediate feedback, summative feedback), grade (second, third), and their interaction as between-subject variables. We included student's age and their composite score on the screening measure as covariates.

Overall, children did quite well ($M = 88\%$, $SD = 20\%$). There was a main effect of grade, $F(1, 222) = 4.94$, $p = .03$, $\eta_p^2 = .02$. Third-graders solved more problems correctly ($M = 91\%$, $SE = 2\%$) than second-graders ($M = 83\%$, $SE = 3\%$). There was also a main effect of condition, $F(2, 222) = 4.40$, $p = .01$, $\eta_p^2 = .04$. Pairwise comparisons indicated that children in the immediate-feedback condition solved more problems correctly ($M = 92\%$, $SE = 2\%$) than children in the no-feedback ($M = 83\%$, $SE = 2\%$; $p = .004$) and summative-feedback ($M = 86\%$, $SE = 2\%$; $p = .07$) conditions. There was no significant difference between children in the latter two conditions ($p = .27$). The condition by grade interaction was not significant, $p = .40$.

Posttest. Overall, children did well on the posttest. There was a main effect of grade, $F(1, 222) = 12.76$, $p < .001$, $\eta_p^2 = .05$. Third-graders exhibited higher equation-solving success ($M = 91\%$, $SE = 3\%$) than second-graders ($M = 71\%$, $SE = 3\%$). There was no main effect of condition ($p = .46$), but there was a condition by grade interaction, $F(2, 222) = 4.32$, $p = .01$, $\eta_p^2 = .04$ (see Figure 1). In second grade, there was a significant effect of condition, $F(2, 222) = 3.84$, $p = .02$, $\eta_p^2 = .03$. Children in the no-feedback condition exhibited lower equation-solving success than children in the immediate-feedback ($p = .01$) and summative-feedback conditions ($p = .03$). In third grade, the effect of condition was not significant, $p = .43$. Inspection of Figure 1 suggests somewhat higher accuracy in the no-feedback condition, but not reliably so. To ensure the neutral effect of condition in third grade was not due to a ceiling effect, we excluded children who performed near mastery on the screening measure (solved 3 or 4 of the 4 problems correctly) and reran the analysis. Conclusions remained unchanged in this restricted sample.

Conclusions:

Description of conclusions, recommendations, and limitations based on findings.

In the current study, we examined the presence and timing of feedback in a classroom context for children solving math problems. The effects depended on student's grade level. At posttest, feedback had positive effects for second-grade students relative to no-feedback. In contrast, feedback condition had neutral effects for third-grade students. The results are consistent with previous research suggesting that the effects of feedback vary across students with different levels of knowledge (e.g., Fyfe et al., 2012; Krause et al., 2009). Specifically, students with lower knowledge benefit from the provision of feedback during problem solving. In contrast, students with higher prior knowledge do just as well if not better when no feedback is provided. Here, we extended this finding to elementary school children in a classroom setting.

The findings support the notion that learner characteristics may be as, if not more important than characteristics of the feedback itself (c.f. Cronbach & Snow, 1977). For example, in previous studies, the moderating role of prior knowledge occurred regardless of the content of feedback (Fyfe et al., 2012). Specifically, feedback about students' answers and feedback about students' strategies produced similar results. Here, the moderating role of grade level occurred regardless of the timing of feedback. Feedback provided immediately after each problem produced similar results as feedback provided after all problems had been solved.

Despite these positive contributions, several limitations suggest directions for future research. For example, more work is needed to test the generalizability of these results to different tasks and age groups. Additionally, future research is needed to explain why feedback has neutral or negative effects for students with higher prior knowledge.

Appendices

Not included in page count.

Appendix A. References

References are to be in APA version 6 format.

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.

Chesney, D. L., McNeil, N. M., Matthews, P. G., Byrd, C. E., Petersen, L. A., Wheeler, M. C., Fyfe, E. R., & Dunwiddie, A. E. (2014). Organization matters: Mental organization of addition knowledge relates to understanding math equivalence in symbolic form. *Cognitive Development.* doi:10.1016/j.cogdev.2014.01.001

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

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

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*, 158-70.

Kulhavy, R. W. (1977). Feedback in written instruction. *Review of Educational Research, 47*(2), 211-232. doi: 10.2307/1170128.

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.

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.

McNeil, N. M. (2008). Limitations to teaching children 2 + 2 = 4: Typical arithmetic problems can hinder learning of mathematical equivalence. *Child Development, 79*, 1524-1537.

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

Powell, S. R. (2012). Equations and the equal sign in elementary mathematics textbooks.

Elementary School Journal, 112, 627-648.

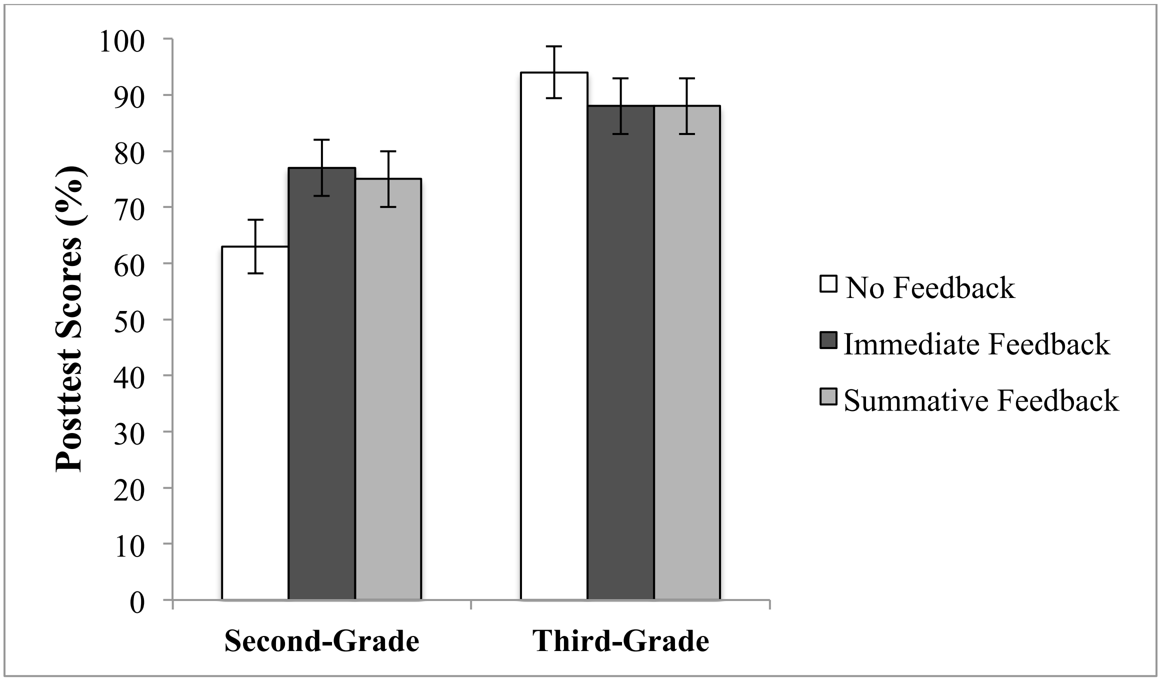
Rittle-Johnson, B., Matthews, P. G., Taylor, R. S., & McEldoon, K. (2011). Assessing knowledge of mathematical equivalence: A construct modeling approach. *Journal of Educational Psychology, 103*(1), 85-104.

Appendix B. Tables and Figures

Not included in page count.

Figure 1

Percent Correct at Posttest by Condition and Grade



Note. Scores are estimated marginal means. Error bars represent standard errors.