Tradeoffs between Immediate and Future Learning

Eliane Stampfer and Kenneth R. Koedinger
Human-Computer Interaction Institute, Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh, PA, 15217, USA
[stampfer, krk]@cs.cmu.edu

Abstract. We compared two types of feedback in an intelligent tutoring system for fraction-addition: *correctness feedback* indicated when each step was right or wrong, and *grounded feedback* showed fraction bars as conceptual scaffolds, requiring more student interpretation. Correctness students solved the tutor problems more efficiently and had better immediate target learning. Grounded feedback students improved more between the post-test and delayed post-test, suggesting that grounding may better prepare students for future learning.

Keywords: Intelligent Tutoring Systems; Fraction Addition; Metacognition.

Introduction

Prior work on tutoring systems (Nathan, 1998) found benefits for learning with easy-to-interpret conceptual feedback instead of correctness feedback. This feedback is *grounded* in students' prior knowledge and evokes conversational grounding: the tutoring system dynamically translates a student's intention (e.g., make equivalent fractions) to a visual representation (fraction bars) that may help the student see if that action matched the original intention (is the same amount colored in?). It lets the tutoring system ask "did you mean that?" and allows students to evaluate their own work. However, research indicates that interpreting representations is non-trivial and may overwhelm students and hinder learning (Ainsworth, Bibby, & Wood, 2002). This experiment aims to evaluate one form of grounded feedback to determine if the benefits of student evaluation of their own work outweigh the costs of interpreting an additional representation.

Tutor Design

Instead of immediate correctness feedback, the grounded feedback tutor displays a fraction bar for each fraction, dynamically updating the converted and sum fractions as the student inputs them (Figure 1, left). The bars are the same size, and corresponding bars are aligned vertically to encourage comparison. In a pilot study, students used the bars to find and fix errors (Stampfer, Long, Aleven, & Koedinger, 2011). Without correctness feedback, students may flounder and waste time on mistakes. However, the grounded feedback may also encourage sense-making and deeper conceptual thinking.

The correctness feedback tutor colors input green if correct, and red otherwise (Figure 1, right). Also, the correctness tutor prevents both inefficient strategies (e.g., converting addends with like denominators) and the erasure of correct inputs. Since grounded feedback is intended to foster evaluation skills, students may try inefficient strategies and erase correct inputs, thus creating situations where those evaluation skills are necessary. All tutors had on-demand text hints (not shown below), and required students to solve each problem correctly before moving on.

Methods and Participants

All six 5th grade classes from a public school near Pittsburgh elected to participate in the study (138 students). Two classes each were high, average, and low achieving. Students took a pretest, completed a short tutorial on using the tutor software and adding fractions, then solved up to 20 fraction addition problems with a randomly assigned software tutor (for up to 40 minutes). Afterwards, students took a post-test, and did a delayed post-test two weeks later. Test forms were matched and counter-balanced.

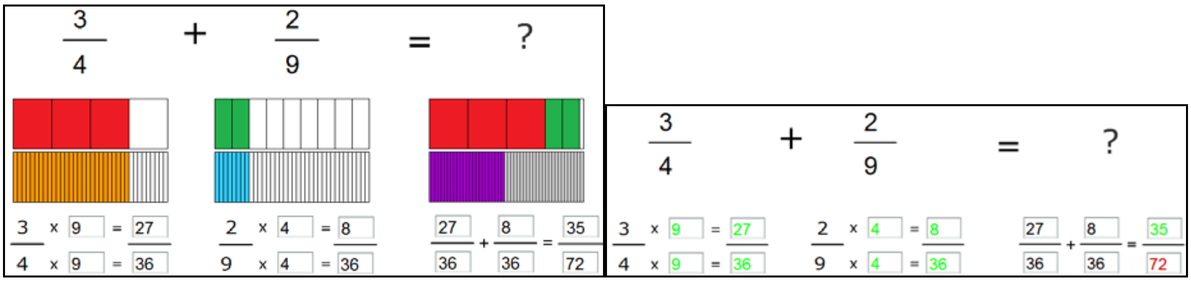

Figure 1. Grounded Feedback (left) and Correctness Feedback (right) with sample input

About half of the grounded feedback students were *Guided:* before starting the fraction-addition problems, they practiced comparing fractions with fraction bars and were told “the answer rectangle will have two colors, showing the fractions you are adding.” The rest of the grounded feedback students were *Discovery:* they did not get any tutoring on how to use or interpret the feedback.

Results and Discussion

Discovery and Guided had similar tutor performance. The additional Guided training did not reduce time taken or hints requested per problem, or how often students marked incorrect work as complete. Therefore, Discovery and Guided are collapsed into one Grounded Feedback (GF) condition. GF students solved fewer tutor problems and requested more hints per problem than the Correctness (C) students (mean problems: C 17.7, GF 9.2, $p<.01$ with full pretest score and tutorial time as significant covariates; hints per problem: C 2.47, GF 7.68, $p<.01$ with the above covariates).

Due to a randomization error, C had more students from the high-achieving classes than the low-achieving ones, explaining their higher pretest scores (Table 1). Though differences on the full post-test are not significant when the full pretest score and the class achievement level are covariates, an ANCOVA on the target post-test scores shows differences between C and GF ($p=.036$, target pre-test scores and class achievement level are covariates).[1] Greater efficiency and immediate learning for C is consistent with prior work on cognitive load (Kirschner, Sweller, & Clark, 2006).

The school curriculum between the post and delayed-post tests included fractions, explaining those gains. The greater improvement from post to delayed post by the GF condition is marginally significant ($p=.057$) in an ANCOVA on delayed post-test target scores with target section of post-test and class achievement as covariates.[2] Correctness score differences from post to delayed are not significant, indicating that classroom instruction hardly benefitted Correctness students. Additionally, differences between conditions on the delayed post-test (full and target section) are not significant.

Table 1: Test score means (standard deviations). Full test out of 13, Target section out of 4.

Condition	Target Only Pretest	Target Only Post-test	Target Only Delayed Post
Correctness	1.23 (1.05)	1.89 (1.34)	1.93 (1.24)
GF	1.12 (.917)	1.44 (.921)	1.85 (1.25)
Condition	Full Pretest	Full Post-test	Delayed Post
Correctness	5.47 (3.09)	6.62 (3.39)	6.86 (3.24)
GF	4.49 (2.47)	5.63 (2.60)	6.80 (3.00)

[1] Condition differences at Full Post-test are not significant (with full pre-test score and class achievement as covariates).
[2] Greater GF improvements from post to delayed is also significant on the full test ($p=.035$, with full post-test score and class achievement as covariates).

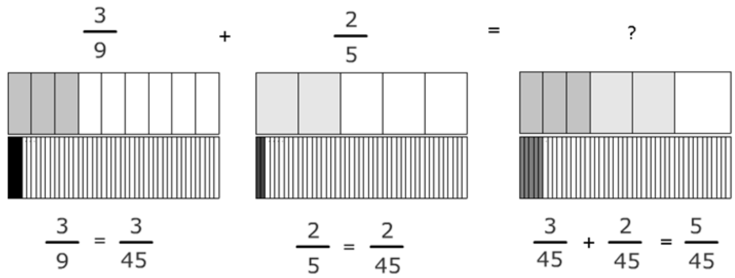

Figure 2: Sample work that students evaluate, with conceptual scaffold.

Although students seemed to benefit from the grounded feedback, they did not understand it fully. One item on each test form measured this ability, asking students if a fraction addition problem was solved correctly (see Figure 2), with fraction bars as a conceptual scaffold. The sum fraction only lined up when the work was correct, making the task trivial for students with strong conceptual understanding. Student scores on this evaluation are low overall (Out of 1: C .66, .48, .58 and GF .47, .57, and .74 for pre, post, and delayed respectively). Process measures provide further evidence: GF students routinely indicated they were done when their fraction bars did not line up (mean: 1.48 times per problem). GF improved in their conceptual understanding from pre to post test and also from post to delayed, while C did not improve from pretest. Excluding students who were correct at pretest, a chi-squared test reveals that differences at delayed post-test are significant (p=.02).

Conclusion

This experiment found a tradeoff: although correctness feedback provided better immediate procedural gains, grounded feedback appeared to provide conceptual gains and preparation for future learning. Given that the grounded feedback group was not better at the delayed post-test and there was only one conceptual item, future work should seek stronger evidence to test its conceptual and future learning benefits. Initially we thought grounded feedback would only be successful when it was firmly within a student's prior knowledge. This experiment showed that students could still benefit from the grounded feedback even if they did not understand it fully. Future work should determine if grounded feedback is more beneficial once students have a stronger conceptual foundation. One way to test this hypothesis is to give students more conceptual practice with fraction bars before they are assigned to tutor conditions.

Acknowledgments. This research was supported in part by the Pittsburgh Science of Learning Center through NSF award SBE-0836012, and the Institute of Education Sciences, U.S. Department of Education, through Grant R305B090023 to Carnegie Mellon University.

References

Ainsworth, S., Bibby, P., & Wood, D. (2002). Examining the Effects of Different Multiple Representational Systems in Learning Primary Mathematics. *Journal of the Learning Sciences*, *11*(1), 25-61.

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*(2), 75–86.

Nathan, M. J. (1998). Knowledge and Situational Feedback in a Learning Environment for Algebra Story Problem Solving. *Interactive Learning Environments*, *5*, 135-159.

Stampfer, E., Long, Y., Aleven, V., & Koedinger, K. R. (2011). In G. Biswas et al, (Eds.): Eliciting Intelligent Novice Behaviors with Grounded Feedback in a Fraction Addition Tutor. *Proceedings of the 15th Annual Conference on Artificial Intelligence in Education,* (560-562) Auckland: Springer.