International Society of the Learning Sciences

How reward- and error-based feedback systems create micro-failures to support learning strategies

Allison S. Liu, Kirk Vanacore, & Erin Ottmar
aliu2@wpi.edu, kpvanacore@wpi.edu, erottmar@wpi.edu
Worcester Polytechnic Institute

Abstract: Feedback in educational technologies can teach and engage students in math, but questions remain on how to present failure feedback that supports positive learning behaviors. We explore how error- and reward-based feedback influenced students' choices to replay completed problems in *From Here to There!*, a math game-based educational technology. We conducted a multilevel logistic regression in 1,031 seventh-grade students predicting whether a problem was replayed, using whether an error was made during the first problem attempt and how many clovers were earned after solving the problem (with more clovers earned for more efficient solutions). Preliminary results show that making an error or earning two clovers increased the likelihood of replays, whereas earning one or three clovers decreased the likelihood. Our findings suggest that clovers and errors may be "micro-failures" that encourage replays by supporting feelings of competency and/or autonomy, which has implications for game design elements for learning and engagement.

Introduction

Educational technologies are becoming increasingly popular as an engaging medium for teaching math. One major benefit of educational technologies is their ability to provide immediate feedback on students' performance (Moyer-Packenham et al., 2019), but developers must consider how to show students where they can improve without causing them to disengage entirely from learning. It is important to understand how different types of feedback systems motivate students to adopt effective learning behaviors in response to failure. Here, we explore whether two feedback systems—errors made and rewards earned—influence the likelihood that a student will replay completed problems within *From Here to There!* (FH2T), a game-based educational technology that teaches algebra.

Feedback systems and failure in game-based educational technologies

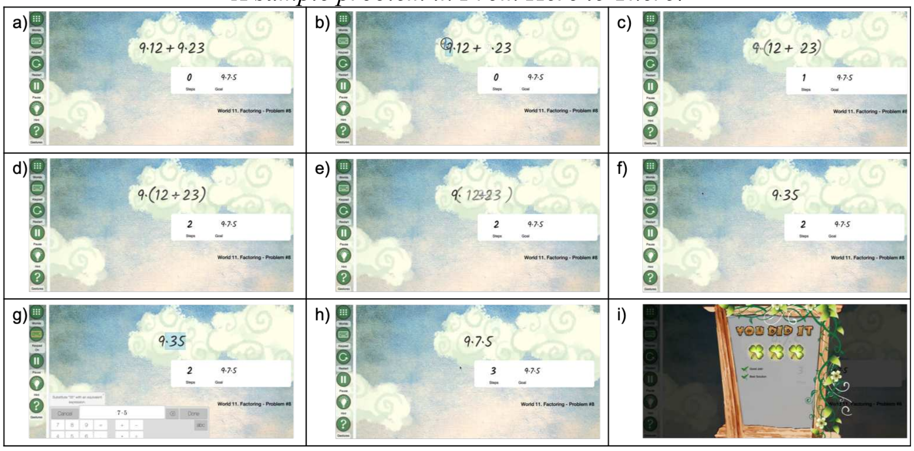
Feedback is often conceptualized in terms of Self-Determination Theory (Ryan & Deci, 2000), which defines three types of motivation: intrinsic (pursuing activities because they are inherently enjoyable), extrinsic (pursuing activities because of some external outcome), and amotivation (lacking any intention to pursue an activity). Supporting learners' innate needs of autonomy, competence, and relatedness can maintain intrinsic motivation or self-determined forms of extrinsic motivation, which have been linked to greater persistence, positive self-perceptions, and engagement (Ryan & Deci, 2000). Thus, feedback may promote these forms of motivation if the feedback makes learners feel competent and in control over the feedback they receive.

Game-based educational technologies overlap with both educational and gaming spaces, each of which brings their own expectations; as such, these tools must grapple with the challenge of evoking "game-like" failure, which can lead to greater engagement and persistence akin to non-educational games, as opposed to "classroom" failure, which can cause students to question their abilities and feel demotivated akin to failing in a school context (Williams-Pierce, 2019). An open learning design question is how to implement feedback systems that increase feelings of game-like failure instead of classroom failure. One answer may be to implement feedback systems that feel low-stakes by using "micro-failures," defined as quiet and gentle failures that encourage the player to seek feedback (Williams-Pierce, 2019). Using feedback to draw students' attention to micro-failures may encourage optional but positive learning behaviors, supporting autonomy and competence. The question then is what types of feedback systems students will perceive as micro-failures.

The current study: Feedback and replays in *From Here to There*

The current study explores error- and reward-based feedback as potential micro-failures, both of which are commonly implemented in game-based educational technologies. We use FH2T as our game context (https://graspablemath.com/projects/fh2t). Students are given a starting algebraic expression (Figure 1a) and must use learned gesture-actions (e.g., in 9·12+9·23, dragging the 9s together to factor them out; Figure 1b-1d) to transform the expression into a goal state (Figure 1h). After completing each problem, students can choose to replay the problem or move on. FH2T has been effective for improving elementary and middle school students' mathematical understanding (Chan et al., 2021; Hulse et al., 2019). A cluster analysis found that a sub-sample of FH2T users repeatedly replay completed problems, and these users show among the largest learning gains after using FH2T (Lee et al., under review). We investigate whether FH2T's error- and reward-based feedback systems incentivize replay strategies in students.

Figure 1
A sample problem in From Here to There!



In FH2T's error feedback system, if students make a mathematical error (e.g., attempting to add 12+9 in 9·12+9·23), then the expression shakes and does not change, providing implicit visual feedback that an error has been made. This feedback may help students to understand where they can correct their behaviors and motivate students to replay and fix these errors. On the other hand, students may feel that they have less control over error feedback (as they must fix the error to progress) or perceive errors as indicators of their lack of competence, reminiscent of classroom failure; such feelings may be less motivating to go back after completing problems. In FH2T's reward system, students earn one to three clovers after each problem based on the number of steps taken to solve the problem, with more clovers earned for fewer steps (see Figure 1i). Clovers are one of the primary ways students have to understand how well they are performing, though the rules that determine the number of clovers are never explicitly explained. These performance-contingent rewards could be seen as a gentle micro-failure because earning one or two clovers suggests that improvements can be made, but students are not penalized in any way. Because fewer clovers do not prevent progression, this optional aspect may feel game-like and promote feelings of autonomy, motivating students to replay problems to gain a higher number of clovers. However, failing to achieve three clovers may still be perceived as a form of "classroom" failure since clovers are tied to students' in-game performance, which may lead them to avoid replays. To begin exploring how these two types of feedback systems influence replay behaviors, we investigate:

RQ1: a) Does the number of clovers earned after the first problem attempt predict whether students replay problems? b) Does receiving one to two clovers encourage students to replay problems more than receiving the maximum three clovers?

RQ2: Do errors during the first problem attempt separately predict whether students replay problems?

Method

As part of a larger study comparing the impacts of three technology interventions, 4,092 seventh-grade students were recruited from 10 middle schools within a large suburban district in the Southeastern United States. The current study focuses only on students randomly assigned to use FH2T (N = 1,031). First, students completed a 40-45 minute pre-test of their algebraic knowledge (Star et al., 2015; α = .89) and math anxiety (Ganley & McGraw, 2016; α = .87). Next, students participated in nine 30-minute FH2T sessions across the school year. Because the study occurred between September 2020 and April 2021 during the COVID-19 pandemic, students chose their instructional modality (100% in-person or virtual) prior to the fall semester. Regardless of modality, all study assignments were administered online, and students worked individually at their own pace using individual devices. Students saw up to 252 unique problems; for each problem, FH2T recorded: 1) whether students made an error during their first problem attempt (error vs. no error); 2) the number of clovers earned after the first problem completion (three clovers for solving the problem in the optimal number of steps, two clovers for solving within two of the optimal number of steps, and one clover for all other solutions); and 3) whether students replayed the problem after completion (replayed vs. not replayed). Errors were not counted as solution steps and therefore did not influence the number of clovers earned.

Results

To investigate whether clovers and errors influenced the likelihood of replaying a problem, we estimated a multi-level logistic regression predicting whether a problem was replayed or not. The model was estimated using 74,359 problems in which students completed at least one attempt. The model contained two cross-classified levels—problem and student levels—because students attempted multiple problem sets and problems were attempted by multiple sets of students. Random intercepts were entered for both students and problems. Problem-level predictors were the number of clovers earned and whether the student made at least one error on the first problem attempt. Pre-test algebraic understanding and math anxiety scores were included as student-level predictors. The number of clovers earned on first attempt was dummy coded, such that two clovers and three clovers were both compared with a referent group of one clover; this split allowed us to compare whether receiving a specific number of clovers differentially influenced the likelihood of replaying a problem.

For **RQ1,** the number of clovers earned on the first problem attempt significantly predicted the likelihood of replay (Table 1). Additionally, there were differential associations between the likelihoods based on the number of clovers earned: students who earned two clovers were 1.45 times as likely to replay a problem compared to students who compared to students who earned one clover ($\beta_1 = 0.42$, $SE = 0.07$, $p < 0.001$), based on relative risk calculation of students with average pretest scores who did not make an error (Table 2). Those who earned one clover on their first attempt were less likely than chance to replay the problem ($\beta_0 = -2.38$, $SE = 0.11$, $p < 0.001$). As expected, students who earned three clovers had the lowest likelihood of replaying a problem compared to those who earned one or two clovers ($\beta_2 = -2.37$, $SE = 0.07$, $p < 0.001$).

Table 1
Logistic Regression Output for Model Predicting Replay

	β	SE	p
Intercept	-2.38	0.11	< 0.001
Two Clovers	0.42	0.07	< 0.001
Three Clovers	-2.37	0.07	< 0.001
Error Made	0.91	0.03	< 0.001
Pre-test Algebra	0.35	0.07	< 0.001
Pre-test Math Anxiety	0.05	0.06	0.46
Random Effects			
σ_2	3.29		
$\tau_{00\ \text{Student}}$	2.58		
$\tau_{00\ \text{Problem}}$	0.59		

$N_{\text{observations}} = 74{,}359$; $N_{\text{student}} = 1{,}031$; $N_{\text{problem}} = 252$

Table 2
Probability and Relative Risk of Problem Replay by Number of Clovers and Whether an Error Was Made in the Problem.

	Errors		
Clovers Earned	N	Y	Relative Risk
One Clover	.08	.19	2.38
Two Clovers	.12	.26	2.17
Three Clovers	.01	.02	2.00

Note: Probabilities were computed for students with average pre-test math and anxiety scores.

For **RQ2**, making at least one error on the first problem attempt significantly increased the likelihood of replay ($\beta_3 = 0.91$, $SE = 0.03$, $p < 0.001$). Table 2 presents the probability of replay based on number of clovers earned and whether they made errors, as well as relative risks calculations, which show how many more times students are likely to replay when they make at least one error compared to when they do not make an error. Students who made at least one error were at least two times as likely to replay compared with students who did not make an error regardless of the number of clovers earned (range: 2.00-2.38), with relative risk being highest for students who earned one clover and lowest for students who earned three clovers.

Discussion

In summary, both errors and clovers appear to incentivize students to replay problems. Earning two clovers on the first problem attempt may have been perceived as a "micro-failure", encouraging students to replay completed problems to improve their solutions and earn the maximum three clovers. Notably, the magnitude of the "micro-failure" appears to influence students' motivation to replay: earning two clovers was associated with a substantially greater likelihood of replaying than earning one clover. One potential interpretation is that students who feel closer to earning the maximum three clovers are more likely to exhibit internal motivation to replay, incited by their feelings of competence that come from being close to perfect. Alternatively, students who are further away from earning three clovers (i.e., earn only one) are less likely to display the internal motivation required to replay because of feelings of incompetence. Students who made more errors were also more likely to replay even though errors are not implicated in clover allocation. It is possible that students do not realize that clovers and errors are independent from one another and they return to the problem to complete it without making errors in the hopes of earning more clovers. The directness of error feedback may also build feelings of competence by helping students identify misunderstandings, encouraging them to replay.

Based on prior work suggesting that replays may lead to higher learning gains (Lee et al., under review), building on performance-contingent reward systems similar to FH2T's clovers may thus be an effective method for improving engagement and learning. In future work, we plan to delineate whether replays specifically are what lead to better learning in FH2T users, and to explore what mechanisms underlie this relationship. Further, because the percentage of students who choose to consistently replay problems on their own appears to be relatively low (especially being optional for progression), we will also explore methods that encourage more students to choose to replay (without taking away the feeling of autonomy from being untied from progression).

References

Burgers, C., Eden, A., van Engelenburg, M., & Buningh, S. (2015). How feedback boosts motivation and play in a brain-training game. *Computers in Human Behavior, 48*, 94-103. https://doi.org/10.1016/j.chb.2015.01.038

Chan, J. Y.-C., Lee, J.-E., Mason, C. A., Sawrey, K., & Ottmar, E. (2021). From Here to There! A dynamic algebraic notation system improves understanding of equivalence in middle-school students. *Journal of Educational Psychology*. https://doi.org/http://dx.doi.org/10.1037/edu0000596

Ganley, C. M., & McGraw, A. L. (2016). The development and validation of a revised version of the Math Anxiety Scale for Young Children. *Frontiers in Psychology, 7*, 1181. https://doi.org/10.3389/fpsyg.2016.01181

Hulse, T., Daigle, M., Manzo, D., Braith, L., Harrison, A., & Ottmar, E. (2019). From here to there! Elementary: a game-based approach to developing number sense and early algebraic understanding. *Educational Technology Research and Development*, *67*(2), 423–441. https://doi.org/10.1007/s11423-019-09653-8

Moyer-Packenham, P. S., Lommatsch, C. W., Litster, K., Ashby, J., Bullock, E. K., Roxburgh, A. L., Shumway, J. F., Speed, E., Covington, B., Hartmann, C., Clarke-Midura, J., Skaria, J., Westenskow, A., MacDonald, B., Symanzik, J., & Jordan, K. (2019). How design features in digital math games support learning and mathematics connections. *Computers in Human Behavior, 91*, 316-332. https://doi.org/10.1016/j.chb.2018.09.036

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology, 25*, 54-67. https://doi.org/10.1006/ceps.1999.1020

Star, J. R., Pollack, C., Durkin, K., Rittle-Johnson, B., Lynch, K., Newton, K., & Gogolen, C. (2015). Learning from comparison in algebra. *Contemporary Educational Psychology, 40*, 41-54. https://doi.org/10.1016/j.cedpsych.2014.05.005

Williams-Pierce, C. (2019). Designing for mathematical play: Failure and feedback. *Information and Learning Sciences, 120* (9/10), 589-610. https://doi.org/10.1108/ILS-03-2019-0027

Acknowledgments

The research reported in Paper # was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A180401 to Worcester Polytechnic Institute. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.