International Society of the Learning Sciences

Capturing Students' Learning Strategies in Action Using Clickstream and Eye-tracking Data

Ana Saavedra, Kristen Pilner Blair, Rachel Wolf, Jordan Patrick Marx, and Doris B Chin
anamar@stanford.edu, kpilner@stanford.edu, rcwolf@stanford.edu, jpmarx@stanford.edu, dbchin@stanford.edu
Stanford University

Abstract: When interacting with open-ended environments, students are free to use different learning strategies. In this exploratory study, we focus particularly on college students (n=47) and their use of inquiry and feedback strategies as they explore a science simulation game. By capturing their behaviors through both clickstream and gaze data, we can more closely examine the effect of different strategy uses on their learning. Backwards regression analyses were used to examine students' early behaviors in the simulation and results indicate that learning outcomes can be best predicted by a combination of inquiry and feedback processing behaviors. Of particular interest: quality over quantity of inquiry was more predictive of learning; and both clickstream data of feedback *use* and eye-tracking data on *earlier stages* of feedback processing (e.g., noticing, predicted learning). These results have implications for the design of future open-ended environments that can better support student learning.

Introduction

Scenario*: Laura and Heidi are students in the same major with similar GPAs. For a class activity, they interact with an open-ended learning environment that requires inquiry skills and provides feedback. Afterwards, they complete a test to measure their knowledge. Laura receives a lower score than Heidi.* Assuming they started with the same prior knowledge, what did they do differently in the simulation that affected their learning outcomes? This study contributes to the identification of learning strategies used by students as they study a phenomenon and highlights the importance of using multiple types of measures to understand students' learning processes.

I. Learning strategies and technologies

Inquiry and simulations

Simulations offer scientifically authentic environments for students to learn about natural phenomena. In STEM fields, a boom in online simulation environments (e.g., PhET) has provided students with opportunities to put their inquiry skills into practice (Wieman et al., 2008). Here, we define scientific inquiry, broadly, as the ability to design experiments, then gather, test, and analyze evidence to reveal patterns in the data. Specifically, in the environment discussed here, we focus on the strategies of simplification and finding equivalence (Käser and Schwartz, 2020). Students' behaviors in science simulations, captured in the form of clickstreams, offer rich information about their science inquiry processes as they explore the phenomenon of interest. Going back to the example of Laura and Heidi, it may be the case that in Heidi's exploration of the system, she designed better experiments that provided more information about the model behind what she is observing in the data than Laura.

Feedback and learning technologies

The increase in accessibility of digital technologies provides a unique opportunity to capture real behaviors of students' engagement with feedback instead of using more limited measures such as self-reports. To expand our understanding, we shift from a general perspective to a study of specific, individual behaviors that students take while processing feedback; interactive, digital environments provide students with expansive opportunities to not only receive feedback, but to act on it (Segedy et al., 2013). To the best of our knowledge, there is limited work studying feedback processing (e.g., Tärning et al., 2020). Returning to the example of Laura and Heidi, it may be that Heidi scored higher because she processed the feedback in the system more thoroughly than Laura.

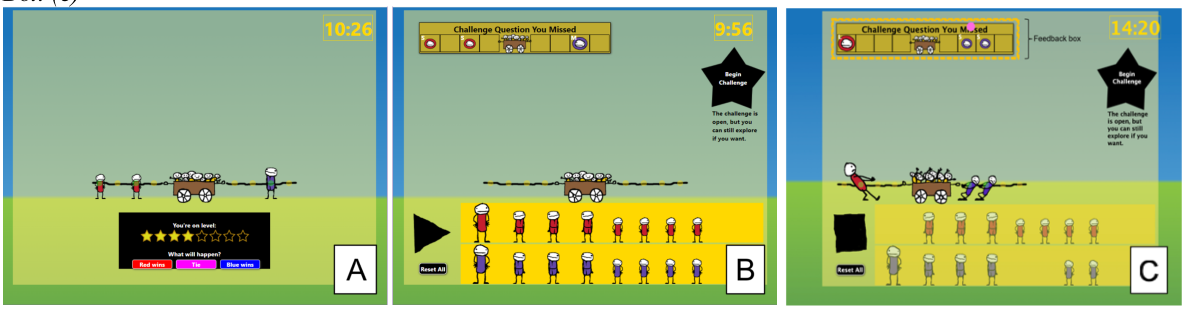
Different students' behaviors while interacting with a learning environment should be analyzed to better predict student learning and design the correct scaffolds to support their understanding. The study presented here uses an open-ended science environment that provides affordances for both learning strategies to be used and measured, with an end goal of understanding how these strategies might interact to support better student learning.

II. Pull!: An online environment for measuring learning strategies

Pull! is a physics game modeled on a PhET simulation, Forces and Motion Basics (Käser and Schwartz, 2020). Pull! is framed around a tug-of-war: teams of small, medium, and large characters on either side of a rope try to pull the other team across a center line. The object is to answer a series of eight questions in a Challenge mode and predict whether a given setup will go left, right, or end in a tie. In the simplified physics of the game, the winning side simply depends on the total "weight" of each team's characters; position does not affect the outcome. To succeed in the game, learners must understand the ratios between the characters. If learners miss a question in Challenge mode, they are sent to an Explore Room, where they have the opportunity (if they wish) to test any number of setups to learn about the simulation environment (see Figure 1). Questions progress in difficulty: three easy (e.g., a small character on either side), two medium level (e.g., a medium vs. 2 small), and three hard (e.g., 2 small, 1 large vs. 3 small, 1 medium); learners never see the same problem twice.

Figure 1

Pull! Screenshots Include an Example Medium-level Challenge (a), the Explore Room (b) and AOI Feedback Box (c)



The Explore Room is designed to provide feedback, as well as allow open-ended inquiry; upon entering the room, a player sees the question she missed in the top left corner as feedback and has unlimited choices to test setups among the provided characters by placing characters on the rope and pressing <play> to start the simulation. Some setups provide less information about the system, e.g., have identical characters on each side or have so many characters on each side that it would be difficult to discern relationships. Other setups are more informative, e.g., a medium vs. 2 small characters yields a tie. We hypothesize that certain behaviors in the Explore Room will predict better learning as measured by a posttest. It might be the case that more informative setups yield better learning, where more informative means students test setups which use principles of good inquiry, e.g., simplification. Another possibility is that feedback uptake yields better learning, where uptake is defined as students testing the same set-up as provided in the feedback box. In some cases, these factors will overlap, i.e., the feedback displayed shows an informative setup that is beneficial to students' learning.

Methods

Participants and Procedures

Community college students (n = 62) consented for the study and were compensated with a 0.5 class credit. Fifteen students with low-quality eye-tracking data or incomplete measures were removed, resulting in a final n = 47. Study procedures were approved by the Stanford Institutional Review Board. The study was conducted in 30-minute individual sessions, consisting of four stages: i) eye-tracking calibration, ii) interaction with the simulation, iii) learning measures, and iv) a debrief where experimenters provided a summary and answered questions.

Measures

For this study, we used several sources of data. The first two sources of data are measures of student behaviors while playing the game: 1) clickstreams extracted from students' interactions with Pull! and 2) gaze data extracted from the eye-tracking. The third source of data is the posttest administered immediately after students played the game and include two types of learning measures. More detailed information is provided below.

1. Clickstream data

All setups (experiments) run in the Explore Room were logged by the system, then characterized in two different ways. First, setups were categorized based on their simplicity and informativeness for discerning the ratio between character types. This included identifying tests that isolate one of the characters' sizes (one-sided), as well as tests that result in a tie (and thus are highly informative). Second, setups were categorized based on whether they likely resulted from feedback uptake or independent inquiry. They were classified as feedback uptake if after missing a question in Challenge mode, students test a setup with the same characters and same positions as the missed question (Exact Feedback) or the same characters in different positions (Inexact Feedback). Setups were classified as independent, self-driven inquiry behaviors if they were not related to the feedback. Metrics included the total number of setups tested, as well as counts of setups categorized by both informativeness and source (independent inquiry of feedback) as described above. Additionally, we coded for whether a player explicitly tested whether position had an effect by varying, in two consecutive tests, only the position of one (or more) characters.

<u>2. Gaze data and feedback noticing</u>

We used a screen-based eye-tracker (Tobii Pro Nano) to record students' eye movements as they played the game. Gaze data was cleaned using a smoothing algorithm (Hessels et al., 2016), then mapped onto the game coordinates and logged as a series of game events. For the purposes of this study, a question of particular interest was how the visual inspection of feedback affected student learning. Fixations were computed in the pre-defined area of interest (AOI), the feedback box in the top left corner of the Explore Room (see Figure 1C). The maximum time students spent looking at this AOI was calculated each time they entered the explore room.

<u>3. Learning measures</u>

Two types of learning measures were used in the posttest. 1) A performance measure focused on quantitative understanding of the ratio structure governing the system. It displayed a setup of characters on the left side of the screen. Students chose whether or not each of ten possible configurations on the right-hand side would result in a tie with the given left-hand configuration. Students received one point for every correct response and lost one point for each incorrect response. The maximum score was 10 and the minimum was -10. 2) In an open-ended response measure, students responded to the following statement: *Explain to someone what are the rules of the game you just played.* This prompt was included to measure how learners chose to verbally describe what they understood as important from the game and served as a measure of more nascent understanding of the system. Student responses were transcribed and coded using the following levels: 1=Qualitative response, 2=Mentioned arithmetic without including values, 3=Included general values, and 4=Included specific-detailed values.

Results

Overview of game play and learning measures

Individual students played until they either beat the game or timed out after 15 minutes. The average game play time was 8.8min (SD=4.4min). Most players beat the game (85%), with an average of 9.55 (SD= 8.35) Challenge attempts. The average on the learning measures was 5.23 (SD=4.22) on the performance measure and 2.72 (SD=1.31) on the open-ended response item.

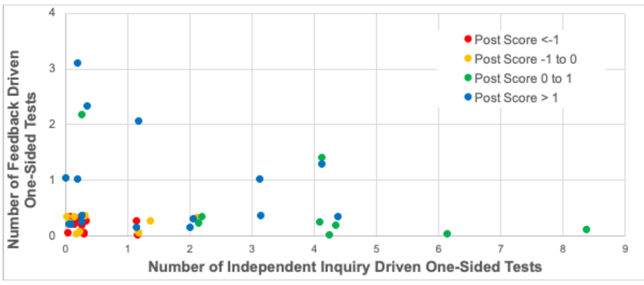
Which student behaviors in the online environment predict their learning?

We present exploratory analyses to determine which behavioral measures among students' clickstream and eye-gaze data best predict their learning as measured by the posttest. Due to space restrictions, we present only those analyses that: 1) used a dependent learning measure, Z-Post score, that is the sum of the z-scores of the posttest measures described above; 2) looked at behaviors early in the game, examining data up to the first five Challenge attempts to equalize for game play length across students and 3) focused specifically on behaviors after missing medium difficulty questions (levels 4 and 5) as these are the levels that require basic understanding of the underlying rules of the system. And thus, best captured the learning strategies that students were utilizing.

The behavioral measures described in the measures section, in addition to the number of medium difficulty questions missed, were entered as independent variables into a backwards regression model (SPSS v. 27, default criterion: probability of F-to-remove >=.100) to predict students' learning performance. The final model achieved ($F_{(4,42)} = 6.46$, $p = .000$; R-squared = .38) indicated that the Z-post score could be predicted by four variables: two feedback-related variables (Number of one-sided tests that reflected exact feedback setups and Maximum gaze time spent on feedback in a single Explore room visit; standardized **β**'s = .40 and .26 respectively) and two inquiry-related variables (Number of informative, self-driven one-sided tests and Check position score, a variable indicating if learners tested for position; standardized **β**'s = .24 and .22 respectively). Interestingly, neither the number of medium level questions missed within the first 5 challenge room attempts, nor the number of total tests conducted during subsequent exploration remained in the model, indicating that it is not the quantity but the quality of experiments that lead to better learning outcomes.

Figure 2

Scatterplot of One-Sided Tests Driven by Feedback vs. Independent Inquiry by Z-Combo Posttest Score (Note, points jittered up to .4 above whole number value)



We further examined students' use of one-sided tests (see Figure 2). Students with the lowest Z-post scores were least likely to test setups that were one sided (and thus highly informative) within their first five Explore room visits. Of students who engaged in one-sided tests, most focused on *either* feedback driven setups *or* independent inquiry tests. Only a handful (8.5%) conducted *both*. Together, these results suggest that both learning strategies, using feedback and conducting independent inquiry, can lead to better learning, but that students do not naturally use both, at least in the context of this learning environment.

Interestingly, among those students who did *not* explicitly engage in any feedback-related tests, there was still a correlation between time spent looking at feedback (maximum feedback gaze time) and the combined Z-post score, $r(38) = .36$, $p = .03$) indicating that *noticing* feedback (and likely decoding and making sense of it) may be a productive learning strategy, even if it is not directly *acted upon*.

Discussion and Conclusion

We examined students' use of feedback and inquiry strategies in a simulation-based, open-ended learning environment called Pull! Both clickstream and eye-tracking data were used to capture students' learning pathways as they explored the phenomenon of net forces. Results indicate that both feedback and inquiry-based strategies may lead to better learning, but that students may not naturally employ both. Notably, the feedback strategies include both active *use* of feedback, as well as earlier stages in processing, such as noticing and decoding. This has implications for the design of future learning environments and the affordances that could be built into these learning technologies to scaffold students such as Laura and Heidi. For example, the technology could draw students' attention to feedback features or help them employ inquiry strategies, like simplifying or isolating variables, that lead to more informative experimental outcomes.

The analyses presented here are exploratory and have some limitations, including the relatively small sample size and a deliberately constrained problem space. Future research and analyses will examine how affective variables may influence students' use of learning strategies and incorporate findings into improving Pull! and other new learning environments with more complex science content.

References

Hessels, R. S., Kemner, C., van den Boomen, C., & Hooge, I. T. (2016). The area-of-interest problem in eye tracking research: A noise-robust solution for face and sparse stimuli. *Behavior research methods*, *48*(4), 1694-1712.

Käser, T., & Schwartz, D. L. (2020). Modeling and analyzing inquiry strategies in open-ended learning environments. *International Journal of Artificial Intelligence in Education*, *30*(3), 504-535.

Segedy, J. R., Kinnebrew, J. S., & Biswas, G. (2013). The effect of contextualized conversational feedback in a complex open-ended learning environment. *Educational Technology Research and Development*, *61*(1), 71-89.

Tärning, B., Lee, Y.J., Andersson, R., Månsson, K., Gulz, A., & Haake, M. (2020). Assessing the black box of feedback neglect in a digital educational game for elementary school. *Journal of the Learning Sciences*, 1-39.

Wieman, C. E., Adams, W. K., & Perkins, K. K. (2008). PhET: Simulations that enhance learning. *Science*, *322*(5902), 682-683.

Acknowledgments

This research was funded by a grant from the Marianne and Marcus Wallenberg Foundation.