It's Good to Explore: Investigating Silver Pathways and the Role of Frustration During Game-Based Learning

Nidhi Nasiar[1](✉), Andres F. Zambrano[1], Jaclyn Ocumpaugh[1], Stephen Hutt[2], Alexis Goslen[3], Jonathan Rowe[3], James Lester[3], Nathan Henderson[3], Eric Wiebe[3], Kristy Boyer[4], and Bradford Mott[3]

[1] University of Pennsylvania, Philadelphia, USA
nasiar@upenn.edu
[2] University of Denver, Denver, USA
[3] North Carolina State University, Raleigh, USA
[4] University of Florida, Gainesville, USA

Abstract. Game-based learning offers rich learning opportunities, but open-ended games make it difficult to identify struggling students. Prior work compares student paths to a single expert's "golden path." This effort focuses on efficiency, but additional pathways may be required for learning. We examine data from middle schoolers who played Crystal Island, a learning game for microbiology. Results show higher learning gains for students with exploratory behaviors, with interactions between prior knowledge and frustration. Results have implications for designing adaptive scaffolding for learning and affective regulation.

Keywords: Game-based learning · pathways · frustration

1 Introduction

Game-based learning enhances engagement and learning, but its flexibility makes identifying struggling students challenging [16]. Researchers seeking to develop adaptive scaffolding for these environments have studied these issues using sequence mining, random walk analysis [15], and comparisons to an expert's most efficient solution—called a "golden path" [16]. However, efficient gameplay does not necessarily improve learning [14]. This study examines "silver pathways" likely to enhance learning by comparing [16]'s "golden path" to the paths taken by students with high and low learning gains in a game called Crystal Island (CI). We also examine how prior knowledge and frustration affect the relationship between these pathways and learning.

N. Wang et al. (Eds.): AIED 2023, CCIS 1831, pp. 497–503, 2023.
https://doi.org/10.1007/978-3-031-36336-8_77

2 Related Work

2.1 Inquiry Learning in Game-Based Environments

Inquiry-based learning, a foundational pedagogical approach, can be effective, but concerns about interest and cognitive load make it difficult to develop adaptive scaffolds. Prior work in [2] has modeled scientific inquiry skills in immersive virtual lab environments. This study is situated in [13]'s framework, which emphasizes the role of exploration and experimentation. [16] compares students' trajectories in CI to a golden path, hypothesizing that expert solutions are more efficient than novices, and acting as a proxy to determine whether a student's path reflects such expertise (Fig. 1).

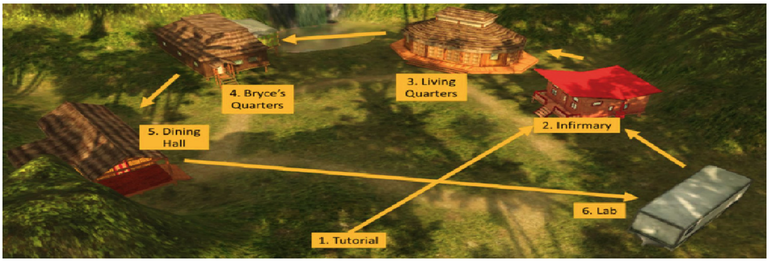

Fig. 1. Overview of Crystal Island with expert "golden path" as operationalized by [16].

2.2 Frustration and Learning

Frustration is known to be crucial to learning, and has been investigated using various tools [8]. The NASA TLX survey [6] (used in this work), was developed to explore astronauts' cognitive load, but has recently been used to measure negative emotions in multiple domains [7]. It assesses students' retrospective perception of their cognitive load, a measure affected by prior knowledge, time pressure, and mental effort (which can also impact frustration). Empirical studies using affect detectors show that the affective transitions in the most popular theoretical model [5] are not common, both the epistemic emotions themselves (e.g., frustration) and their hypothesized transitions (when present) appear to have strong relationships with learning [3, 10]. In general, frustration appears to have a 'Goldilocks Effect,' where either too much or too little leads to lower learning [5, 11]. Successful interventions have been designed [4], suggesting that understanding the relationship between frustration and learning in complex problem-solving can be used to scaffold learning further.

3 Methods

The study uses previously analyzed data from Crystal Island (CI) [12], a learning game for middle school microbiology that promotes inquiry-based learning by having students assume the role of a medical detective investigating an outbreak on a remote island. In this single-player game, students are tasked with identifying the disease and its source of transmission. To do so, they must travel to different locations on the island, collecting data and examining other information sources. The original data included 92 middle-school students enrolled in an urban public school in the southeastern US. Due to missing post-tests, 26 were excluded, leaving 66 students for analysis.

3.1 Measures of Learning and Frustration

Each student completed identical pre and post-tests (scaled from 0 to 13), which were used to calculate normalized learning gain using Eq. 1 [18].

$$\text{Norm\_gain} = \begin{cases} \frac{post-pre}{1-pre} post > pre \\ \frac{post-pre}{pre} post \leq pre \end{cases} \tag{1}$$

Students also completed surveys of interest and engagement, which included an adapted version of the NASA-TLX. This study examines the NASA-TLX frustration scale, which asks students to self-report a range of negative emotions on a scale of 0–100. We dichotomize these measures (i.e., high vs. low) by splitting on the median: learning gains (*0.15*), pretest (*8*), and frustration (*31.5*).

3.2 Operationalizing and Comparing Student Learning Pathways

In order to exclude brief erratic movements between locations, we only consider pathways to locations where students stayed for at least 20 s. We also consider self-loops, defined as locations where the student stays for at least 10 min (a threshold chosen by dividing the mean of total gameplay time (65 min) by the 6 locations). Finally, we identify paths followed by at least half of the students. To compare the paths of different student groups, we use two metrics: *similarity* and *density*. Similarity measures the number of common transitions between two graphs over the total number of possible transitions [17]. It ranges from 0–1, where 1 indicates identical graphs. Density also ranges from 0–1, and measures student exploration by dividing the number of transitions by the total possible transitions [19].

4 Analysis and Results

Our goal is to identify silver pathways—those that are less efficient but improve learning. We first compare students with high and low learning gains, but learning gains are impractical for triggering real-time interventions as students must complete the game for it to be calculated. Therefore, we compare pathways between students with high and low prior knowledge (i.e., pre-test scores), as moderated by frustration.

4.1 Pathways Differences in Students with High and Low Learning Gains

Figure 2 shows the common paths for students with high and low learning gains, respectively. We note that the full golden path appears for those with high learning gains, but 5/6 golden path transitions are also found among students with low learning gains, suggesting this path is not sufficient for improving learning. Both graphs in Fig. 2 have high similarity to the golden path (*0.71 vs. 0.80* for high vs. low learners). Instead, the main difference between these groups is in density scores (*0.47 vs. 0.42* for high vs. low learners). Low learners also have more self-transitions.

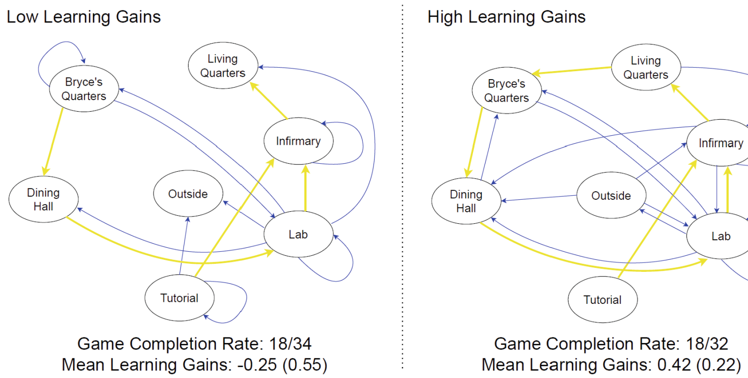

Fig. 2. Differences in pathways based on learning gains. Arrows denote paths taken by at least 50% of the students in that group; those that mirror [16]'s golden path are in yellow.

4.2 Pathway Differences Based on Prior Knowledge and Frustration

We next considered interactions with prior knowledge and frustration. Table 1 shows the similarity and density scores for different groups, while Fig. 3 shows their common learning pathways. Students are divided by both prior knowledge, (i.e. pre-test) and frustration. Specifically, Table 1 shows the similarity of different groups' learning pathways to the golden path and high learning gains path. These results show that students with low frustration and high prior knowledge have greater similarity to the golden path. Students with high prior knowledge also show greater similarity to the high learning graph. Likewise, Fig. 3 shows that high prior knowledge learners complete most of the golden path, but also other paths. Notably, the group with the highest learning gains (low pretest and low frustration) completed only 4/6 golden path transitions—but not those from the infirmary to the living quarters or from Bryce's quarters to the dining hall. This group also shows a self-loop (i.e., $>$10 min) in the tutorial.

When controlling for frustration, high prior knowledge learners tend to have slightly higher density scores than the low prior knowledge learners (Table 1). Higher frustration seems to increase density (*0.39 vs. 0.42* for low and high frustration with low pretests, respectively; *0.44 vs. 0.56* for low and high frustration with high pretests). Apparent exploratory behaviors (i.e., more paths) do not necessarily optimize learning, as learning gains trend slightly higher among those with low frustration than high frustration (*0.10 vs.* $-$*0.15* learning gains for low and high frustration learners with high pre-tests; *0.24 vs. 0.13* learning gains for low and high frustration learners with low pre-tests.

Table 1. Density and similarity scores of pathway analyses by frustration and prior knowledge.

	Low Frustration		High Frustration	
	Low PK	High PK	Low PK	High PK
Density	0.39	0.44	0.42	0.56
Similarity (Golden Path)	0.69	0.78	0.67	0.69
Similarity (High Learning Path)	0.78	0.86	0.80	0.94

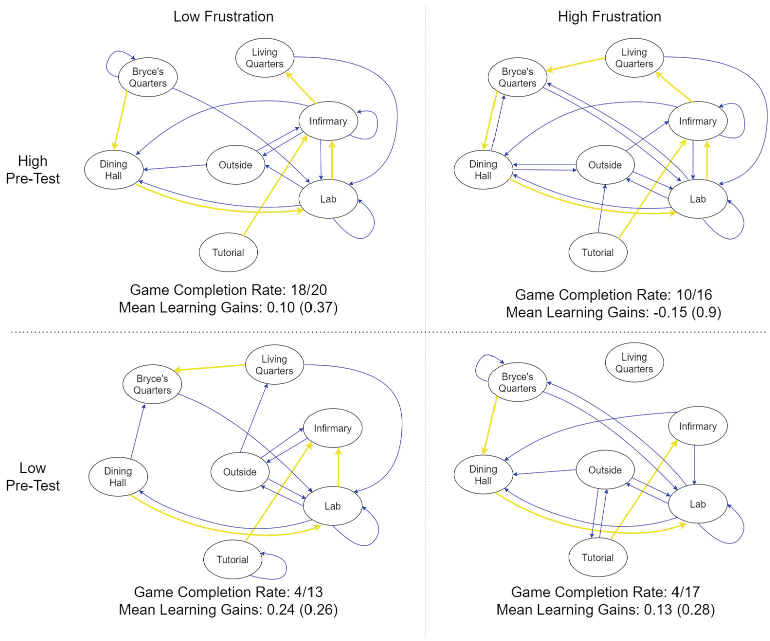

Fig. 3. Differences in pathways based on pre-test and frustration. Arrows denote paths taken by at least 50% of students in that group; those that mirror [16]'s golden path are in yellow.

5 Discussion and Conclusion

Deciding when to support student learning in a game-based environment is challenging. Prior work comparing student behaviors to experts [16] is important, but we should not expect novices to mirror expert behavior fully. This study builds on research showing that those with high prior knowledge replicate parts of [16]'s golden path through CI, but we show that exploration may be better for learning than efficiency. We also show that frustration interacts with learning gains by increasing what otherwise looks like exploratory behaviors. Future work should evaluate other frustration measures, including *in situ* measures used in a large body of work [9]. The single, retrospective measure administered in this study makes it difficult to evaluate the degree to which students' ultimate success (or failure) in the game has influenced their memory of this construct.

In situ frustration measures would also provide more nuance, allowing us to differentiate between intentional and haphazard exploratory behaviors. Such distinctions are important when developing adaptive scaffolds. Finally, *in situ,* measures would allow us to explore whether the Goldilocks effect (i.e., "just right" levels of frustration) varies across student groups. Given the range of variables that frustration interacts with, some learners may tolerate lower levels of frustration than others (e.g., [1]). This study has implications for improving interventions within gameplay based on frustration.

Acknowledgement. This study was supported by NSF under IIS grant Award #2016943, Award #2016993, and #1409639. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF.

References

1. Andres, J.M., Hutt, S., Ocumpaugh, J., Baker, R.S., Nasiar, N., Porter, C.: How anxiety affects affect: a quantitative ethnographic investigation using affect detectors and data-targeted interviews. In: International Conference on Quantitative Ethnography, pp. 268–283 (2021)
2. Baker, R., Clark-Midura, J., Ocumpaugh, J.: Towards general models of effective science inquiry in virtual performance assessments. J. Comput. Assist. Learn. **32**(3), 267–280 (2016)
3. Baker, R., et al.: Affect-targeted interviews for understanding student frustration. In: Proceedings of the International Conference on Artificial Intelligence & Education (2021)
4. DeFalco, J., et al.: Detecting and addressing frustration in a serious game for military training. Int. J. Artif. Intell. Educ. **28**(2), 152–193 (2018)
5. D'Mello, S., Graesser, A.: Dynamics of affective states during complex learning. Learn. Instr. **22**(2), 145–157 (2012)
6. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (task load index): results of empirical and theoretical research. Adv. Psychol. **52**, 139–183 (1988)
7. Hart, S.G.: NASA-task load index (NASA-TLX); 20 years later. In: Proceedings of theHuman Factors & Ergonomics Society Annual Meeting, vol. 50, no. 9, pp. 904–908 (2006)
8. Hutt, S., Grafsgaard, J., D'Mello, S.: Time to scale: Generalizable affect detection for tens of thousands of students across an entire school year. In: Proceedings of the 2019, CHI, pp. 1–14 (2019)
9. Jensen, E., Hutt, S., D'Mello, S.: Generalizability of sensor-free affect detection models in a longitudinal dataset of tens of thousands of students. In: International EDM (2019)
10. Karumbaiah, S., Baker, R.S., Ocumpaugh, J.: The case of self-transitions in affective dynamics. In: Proceedings of the 20th International Conference on Artificial Intelligence in Education, pp. 172–181 (2019)
11. Liu, Z., Pataranutaporn, V., Ocumpaugh, J., Baker, R.S.J.d.: Sequences of frustration and confusion, and learning. In: Proceedings of the 6th International Conference on Educational Data Mining, pp. 114–120 (2013)
12. Min, W., et al.: Multimodal goal recognition in open-world digital games. In: Thirteenth Artificial Intelligence and Interactive Digital Entertainment Conference (2017)
13. Pedaste, M., et al.: Phases of inquiry-based learning: definitions and the inquiry cycle. Educ. Res. Rev. **14**, 47–61 (2015)
14. Sabourin, J., Mott, B., Lester, J.: Discovering behavior patterns of self-regulated learners in an inquiry-based learning environment. In: International Conference on Artificial Intelligence in Education, pp. 209–218 (2013)

15. Snow, E., Likens, A., Jackson, T., McNamara, D.: Students' walk through tutoring: Using a random walk analysis to profile students. In: Proceedings of the International Conference on Educational Data Mining (2013)
16. Sawyer, R., Rowe, J., Azevedo, R., Lester, J.: Filtered time series analyses of student problem-solving behaviors in game-based learning. In: International EDM (2018)
17. Tversky, A.: Features of similarity. Psychol. Rev. **84**(4), 327 (1977)
18. Vail, A.K., Grafsgaard, J.F., Boyer, K.E., Wiebe, E.N., Lester, J.C.: Predicting learning from student affective response to tutor questions. In: Micarelli, A., Stamper, J., Panourgia, K. (eds.) ITS 2016. LNCS, vol. 9684, pp. 154–164. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-39583-8_15
19. Voloshin: Introduction to graph theory. Nova Science Publishers (2009)