"What else can I do?" Examining the Impact of Community Data on Adaptation and Quality of Reflection in an Educational Game

Erica Kleinman
emkleinm@ucsc.edu
University of California, Santa Cruz
Santa Cruz, California, USA

Jennifer Villareale
jmv85@drexel.edu
Drexel University
Philadelphia, Pennsylvania, USA

Murtuza N. Shergadwala
mshergad@ucsc.edu
University of California, Santa Cruz
Santa Cruz, California, USA

Zhaoqing Teng
zhteng@ucsc.edu
University of California, Santa Cruz
Santa Cruz, California, USA

Andy Bryant
a@andymbryant.com
University of California, Santa Cruz
Santa Cruz, California, USA

Jichen Zhu
jichen.zhu@gmail.com
IT University of Copenhagen
Denmark

Magy Seif El-Nasr
mseifeln@ucsc.edu
University of California, Santa Cruz
Santa Cruz, California, USA

ABSTRACT

Adaptation, or ability and willingness to consider an alternative approach, is a critical component of learning through reflection, especially in educational games, where there are often multiple avenues to success. As a domain, educational games have shown increased interest in using retrospective visualizations to promote and support reflection. Such visualizations, which can facilitate comparison with peer data, may also have an impact on adaptation in educational games. This has, however, not been empirically examined within the domain. In this work, we examine how comparison with other players' data influenced adaptation, a part of reflection, in the context of a game that teaches parallel programming. Our results indicate that comparison with peers does significantly impact willingness to try a different approach, but suggest that there may also be other ways. We discuss what these results mean for future use of retrospective visualizations in educational games and present opportunities for future work.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in HCI**.

KEYWORDS

adaptation, reflection, learning, educational games, visualization, community data, retrospective visualization

ACM Reference Format:
Erica Kleinman, Jennifer Villareale, Murtuza N. Shergadwala, Zhaoqing Teng, Andy Bryant, Jichen Zhu, and Magy Seif El-Nasr. 2023. "What else can I do?" Examining the Impact of Community Data on Adaptation and Quality of Reflection in an Educational Game. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), April 23–28, 2023, Hamburg, Germany.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3544548.3580664

CHI '23, April 23–28, 2023, Hamburg, Germany

ACM ISBN 978-1-4503-9421-5/23/04.
https://doi.org/10.1145/3544548.3580664

1 INTRODUCTION

Educational games have become more prominent in recent years as empirical evaluations demonstrate their value as mechanisms for learning [10, 26, 27, 40, 54]. As their use grows, however, new questions emerge about how to ensure they effectively support learning and lead to knowledge gains. One element of learning that is of interest within the domain is adaptation and awareness of alternative approaches [66]. Adaptation refers specifically to making adjustments to one's approach. For example, a student playing a learning puzzle game may initially attempt a brute force approach, trying every option until they find the right one. Upon failure, they may then adapt their approach, instead examining available clues to determine the right solution. Learning literature has demonstrated how consideration of alternative solutions can help learners better understand a topic and achieve success in a learning environment [56, 72]. Games, including educational games, are dynamic environments where there are often multiple correct ways to overcome an obstacle, and thus, willingness to explore alternative solutions, to adapt, is inherently important to successful learning in educational games.

Retrospective visualization is a popular approach for eliciting reflection, and thus triggering adaptation, in digital learning environments and applications [50, 56]. This approach has also begun to gain traction in educational games [23, 66]. Essentially, retrospective visualizations, or visualizations of gameplay and performance data examined after the completion of a learning task, are used to prompt and support reflection, or "those intellectual and affective activities in which individuals engage to explore their experiences in order to lead to new understandings" [6]. Through these new understandings, students, and in the case of educational games, players, may become aware of actions they did not take or could have taken differently, triggering adaptation.

As such, there is value to advancing the use of retrospective visualizations to prompt adaptation and support reflection in educational games. However, while the retrospective visualization approach is common in other gaming domains, such as esports, [32, 67, 68], such studies are less common in the domain of educational games. In educational games, traditional prompts have been the primary means of eliciting reflection and adaptation has been less explored [15, 48, 55].

As retrospective visualizations are still relatively new in the educational games context, and largely understudied, there exist a number of open problems and unanswered questions [71]. One prominent question, which we address in this work, is the impact of peer data on adaptation. While traditional, written-response prompts are typically used to elicit reflection on a learner's own experience, visualization has the potential to present peer data to the learner. Further, it can present large amounts of community data, all at once. Existing work in the learning sciences has suggested that comparison with and examination of community, or peer, data can help students reflect on their own progress and consider alternative approaches to tasks based on what their peers are doing [9, 17, 56]. Similarly, work on retrospective visualization for esports, where the technique is common, has emphasized the presentation of peer data and the important role it plays in prompting adaptation of one's strategy [1, 68]. Other work has, however, suggested that comparison with others could result in players feeling inadequate compared to their peers, which could, in turn, result in poor quality reflection and discontinuation of play [14].

Thus, here, we examine the open question about using peer data in retrospective visualizations to elicit adaptation in educational games. Specifically, we aim to determine the impact of comparison with peers' data on reflection and adaptation. Given that exposure to community data may negatively impact reflection, it is valuable for designers of future retrospective visualizations for educational games to know if and how this data may benefit players, as no apparent benefit would suggest that this data may be better omitted. Further, this work informs researchers of whether or not there is value in further exploring this topic by providing a foundation upon which future work can explore the influence of additional factors. Specifically, we ask:

- How does comparison with peer data impact a player's willingness to consider a different approach?
- How does comparison with peer data impact a player's quality of reflection?

By answering these questions, we believe we can provide valuable insight into the impact of community data on educational game contexts where reflection is prompted post-play through retrospective visualization. Such insights, we argue, can not only help future designers make informed decisions, but also provide a flag for future research to further expand our understanding of adaptation in educational games, its relationship with visualization and reflection, and how we can prompt it. While there do exist other open problems within the domain, including concerns of safety, privacy, and fair play [33, 50, 71], we leave these for future work.

To answer our research question, we conducted a within-subjects study with 36 undergraduate computer science students using the educational game *Parallel*, which teaches parallel programming concepts [70]. Participants played the game and then performed two reflection tasks, one in which they compared their gameplay to their own gameplay in another attempt, and one in which they compared their gameplay to their peers' gameplay. Ordering of reflection condition was randomized. Qualitative data was collected and analyzed using Leijen et al.'s model of reflection quality [39]. McNemar-Bowker tests revealed that comparison with peers did make 1/3 of the players significantly more willing to consider a different approach if they were to play the game again. Additionally, there was no significant impact on the quality of reflection. Based on these results, we conclude that reflection on one's own performance in the context of peers' performance does have a significant influence on one's willingness to adapt. However, it is not an ultimate solution as it did not prompt all participants to adapt, suggesting that other factors may contribute. Based on these conclusions, we discuss considerations for the implementation of peer data in retrospective visualizations to prompt adaptation and present opportunities for future work to expand on these findings. We present these results as a first step towards developing more concrete guidelines for the design and implementation of retrospective community visualizations in educational games.

In summary, our contribution is as follows: first, we highlight the impact of comparison with peer data on willingness to consider a different approach and discuss the implications of this finding. Based on these results, we present implications for the future inclusion of comparison with peers in retrospective visualization systems in educational games. We additionally present five opportunities to build upon this work through future research.

2 RELATED WORK

2.1 Reflection, Adaptation, and Learning

Reflection has always been a central element of learning theories and is formalized in many ways [58, 59]. The significance of reflection within learning has led to specific theories of reflection and how it occurs, with many frameworks quantifying reflection across various levels [2, 16, 20, 39, 44, 51, 63, 69]. For example, Leijen et al. [39] quantify reflection across four levels (description, justification, critique, discussion), which quantify how the student is reflecting, and three foci (technical, practical, sensitizing), which quantify what the student is reflecting on.

In many cases, these framework of reflection present adaptation as a result of effective reflection. Using Leijen et al.'s framework as an example again, the highest level of reflection, according to the framework, is discussion, defined as "Moving beyond the evaluation and explanation of what is, and why they think that is, and pointed out what could be done to initiate changes, and why changes are needed in the first place" [39]. In other words, when one reflects at the highest level, they put cognitive effort into considering how to adapt their behavior. Many learning theories also discuss how high-quality reflection is a mechanism for change, which ultimately results in learning progress. For example, the theory of self-regulated learning [49, 73] explicitly includes adaptation and identification of different ways to move forward as a critical component of the self-reflection process [72].

Because reflection is so integral to learning, educational contexts often dedicate time and resources to eliciting reflection. One of the most common ways this is done is through prompts[64, 65]. For example, Rakovic et al, [52] had students respond to post-exam, reflective prompts in a biology course. They found that the quality of students' reflective evaluations were predictive of adaptations in their learning process, which were in turn predictive of increased frequencies of desirable learning behaviors and higher scores on future exams [52]. Intelligent tutoring systems and open learner models also prompt reflection in order to elicit change through negotiation with the system [12, 24, 62]. For example, NDLTutor [62] prompts the student to provide a self-assessment at set points which will prompt a negotiation if there is a discrepancy between the self-assessment and what the system believes. Negotiation is a means by which the student can either convince the system it is wrong, and prove themselves, or, through reflection, prompted by the negotiation, come to understand that the system is correct and, hopefully, adapt their learning strategies moving forward to account for gaps in their knowledge.

Similar to learning, prompting reflection can also help players learn and adapt in the context of games [30, 47]. Many educational games prompt players to reflect throughout gameplay in order to solidify educational concepts and encourage exploration of alternative solutions, and previous work demonstrates that doing so can improve student learning and performance. For example, O'Neil et al. [48] added a self-explanation prompt, which encouraged self-reflection processes, to an educational math game and found that students who responded to the prompts tended to have higher mean post-test scores than those who did not. Similarly, Fiorella et al. [15] found that adding prompts to a game that taught electrical circuits significantly increased student performance. In another example, Sabourin et al. [55] generated scores for students who played the educational game *Crystal Island* [54] based on their responses to a reflective prompt. They found that scores were significantly predictive of post-test learning gains and that high-scoring students appeared to make more use of the in-game curricular resources than low-scoring students and reported more immersion, interest, and enjoyment [55].

Reflecting on one's own data, however, can only get a learner so far with regards to adaptation, as there are often situations where one may not be able to perceive an alternative solution. This is especially true in circumstances where there are numerous correct solutions, which is often the case in games [60]. Within learning sciences, there are a number of theories that discuss the community aspects of reflection. For example, co and socially shared regulated learning both incorporate the input of others, through feedback or cognitive guidance, to aid in the adaptation of one's techniques through reflection [19]. Having others perform this role, however, is not always a viable option. In such circumstances, being able to view and reflect on community data may be a benefit to learners, prompting targeted research into the presentation, interpretation, and impact of community data in learning contexts.

2.2 Reflection, Adaptation, and Community

Existing work in the learning sciences has demonstrated benefits to having learners view the data of their peers while reflecting on their performance [5, 17, 22, 50]. Using community data to elicit reflection and adaptation is a focus of student-facing learning analytics dashboards (LADs), which are data-driven visual displays that "aggregate multiple visualizations of different indicators about learners, learning processes, and/or learning contexts" [4]. Student-facing LADs aim to help students make strategic decisions in learning environments related to resource (time and energy) management considerations such as what assignments to focus on, how long to study for an exam, or how often to interact with a course management system, such as Blackboard [50, 56]. In this context, being able to view the behavior of classmates and where they stand against their peers in terms of accomplishments is valuable in helping students adjust where they are spending their time and what they are prioritizing as the course progresses [9, 56, 57].

Certain gaming contexts have also adopted this approach, where community data is presented in retrospective visualizations to elicit reflection and motivate adaptation and more efficient learning. [7, 21, 41]. The benefits of learning from others in gameplay, especially esports, contexts are apparent in the work of Wallner et al., who conducted an interview and survey study examining how players use retrospective visualizations in esports [68]. Their results include themes focused entirely on what information players want or need about their opponents and illustrate how players use retrospective visualizations to learn from others. The emphasis on others' data, and the benefits of it, has resulted in many studies focuses explicitly on how information about others gameplay can be extracted from data [32, 67] and a surge in visualizations to support spectators during live gameplay [8, 34]. Many systems that have been developed also feature elements that allow players to look up other players in order to study their gameplay, compare it to their own, and adopt their strategies [1, 3, 37, 45, 61].

While the work described above is informative, it does primarily focus on esports rather than educational games, where community data is less commonly used as a way to elicit reflection and promote adaptation. There is, however, one work that does discuss this. Villareale et al.'s [66] review of reflection in educational games used existing frameworks to conduct a review of 12 programming games and identify four features used to elicit reflection. Among these is "social discourse" or "a space in the interface for community-based discussion where students can examine multiple perspectives and receive feedback on their process that can then be used for reflection" [66]. The authors go on to discuss how, through social discourse, players in programming games are exposed to different perspectives on the same problem and become more aware of alternative approaches, encouraging adaptation.

The use of community data to elicit reflection and adaptation in educational games is, however, under-studied and, to the authors' knowledge, there is no work that explicitly demonstrates that exposure to community data does result in willingness to try a different approach. Further, work on games at large has uncovered some drawbacks surrounding the use of community data. In one example, Esteves et al [14] found that social comparison in games could lead to disengagement if the player felt that they were under-performing compared to their peers. Even if disengagement does not occur, feelings of inadequacy, such as those observed by Esteves et al. [14], could result in lower quality reflection, which may impact, not only adaptation, but learning as a whole.

Ultimately, this highlights an open question regarding the impact of community data as a reflective tool for educational games. While

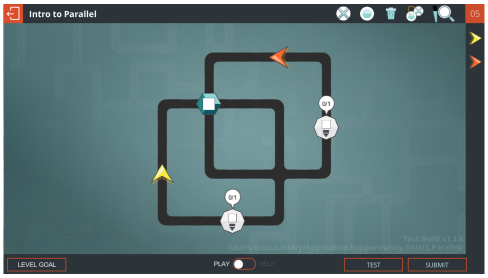

Figure 1: A screenshot of the game *Parallel*. Players need to place semaphores and signals to direct arrows, which carry packages and move along pre-defined tracks, to the designated delivery points. The player must coordinate threads executing in parallel. The level pictured was the subject of the retrospective interviews.

exposure to community data may help encourage adaptation, it has yet to be demonstrated that it does so more than exposure to only one's own data. Further, there are observed drawbacks to exposure to community data, especially in gaming contexts, raising concerns that using community data to attempt to elicit change may have a negative impact on reflection as a process. In this work, we examine these questions, specifically looking at willingness to adapt and the quality of reflections and exploring how comparison with peers' data impacts these factors. This examination will provide valuable insights into the benefits of including peer data in retrospective visualizations for educational games, which can, in turn, guide future design and development decisions.

3 METHODOLOGY

To answer our research questions, we conducted a within-subjects, repeated measures study where participants played an educational game and were then prompted to reflect on their gameplay twice, once in comparison to their own gameplay from another attempt and once in comparison to other players' gameplay. We chose to use a within-subjects design to control for individual differences among participants including individual reflection skills, learning differences, or personality traits. In this section, we describe our methods in detail.

3.1 Parallel - An Educational Game

For this study, we used *Parallel*, a single-player, 2D puzzle game designed to teach parallel programming concepts to computer science students using visual metaphors [29, 70]. The goal in *Parallel* is to coordinate arrows, representing programming threads, to pick up packages and deliver them to designated locations. The threads, however, move at random. Thus, to accomplish their goal, players must place signals and semaphores (the circle and "X" icons in the upper right corner in Figure 1) in appropriate locations to control the threads' movement and access to parts of the track. When a player thinks they have a correct solution they can test, to check the solution with one possible speed of arrow movement, or submit, to test all possible speeds. We chose *Parallel* for this study as each level has multiple possible solutions, providing players with a variety of approaches to compare to their own during reflection.

For this study, *Parallel* was hosted on a web domain (playparallel.com). The game was instrumented on the back-end such that it collected information about all of a player's actions in log files. Players were informed of and agreed to the data collection and use policies when creating accounts on the web site. Log files were only collected for Level 7, because it was a complex enough level to be a reasonable challenge to participants and could be solved in several ways of varying correctness, thus warranting some manner of reflection and adaptation, but not so much of a challenge that participants may become overwhelmed or fail to complete the level.

3.2 Recruitment

36 undergraduate computer science students were recruited from programs in the United States. Undergraduate computer science students were targeted as they are the intended user group for *Parallel*. Participants were additionally required to be 18 years of age or older, located in the United States, able to communicate in written and spoken English, and able to access the playparallel website. Participants were not required to have prior experience with parallel programming, however, 17 did. After completing the study, participants were sent an optional demographic survey. 17 participants responded, 13 of whom identified as male, 3 as female, and 1 as non-binary. 14 of the 17 were in their senior year, 2 were juniors, and 1 was a sophomore.

Race, gender, nationality, and similar demographic information was not collected to avoid biasing results as it was unrelated to the research question.

3.3 Protocol

Upon giving signed consent, participants proceeded through the following steps:

- **Account Creation and Tutorial:** Participants were provided with instructions for how to access playparallel.com, create an account, and complete the *Parallel* tutorial. They were given five days to complete this step.
- **Level 7 (First Playthrough):** The day after the deadline to complete the tutorial, participants were sent instructions to play level 7 of *Parallel*. They were given three days to complete this step.
- **Level 7 (Second Playthrough):** The day after the deadline to play Level 7 the first time, participants were asked to play level 7 a second time. They were given three days to complete this step.
- **Reflection 1:** The day after the deadline to play Level 7 the second time, participants were randomly assigned to a reflection condition and provided with a visualization setup, either self or peer (see below), and responded to a set of reflection prompts (see below). They were given three days to complete this step.
- **Reflection 2:** The day after the deadline to complete the first reflection step, participants were provided with a second visualization setup for whichever reflection condition they did not do the first time (see below), and responded to a set

of reflection prompts (see below). They were given three days to complete this step.

Work on post-game visualization is motivated on the assumption that presenting visualized data to players support reflection [31, 37, 67]. Research in educational domains has also demonstrated this [50, 56]. The goal of this work was to examine the impact of peer data on one element of reflection, adaptation, which much of the existing work additionally assumes to be an asset of visualization. Evaluating whether or not visualization itself impacts reflection is beyond the scope of this work and, therefore, a condition where reflection occurred in the absence of a visualization was not included.

Due to the COVID19 pandemic, this protocol was run remotely. Participants were compensated twice during the protocol. They received 20$ after completing *Reflection 1* and 30$ after completing *Reflection 2*. University IRB reviewed and approved the protocol.

3.4 Reflection

3.4.1 Visualization Setup. For the reflection steps, we used a variation of the network graph from the visualization system *Glyph* [46] to visualize players' gameplay as a sequence of actions. *Glyph's* network graph, seen in Figure 2, represents gameplay as a process using a node-link diagram. Each node represents an action taken and each link represents a transition from one action to another. We hereby refer to these node-link diagrams as "playtraces".

We chose this particular approach to visualization due to the potential for generalizability. The network graph presents actions taken, which represent the player's problem solving strategy, and participants were prompted to use this visualization to reflect on those strategies. This approach to reflection, on one's own actions rather than on one's understanding of the educational content, generalizes to other educational games as well as to digital learning environments, where learners' actions can be similarly logged and presented.

For the purposes of this study, we identified 15 key actions from *Parallel* gameplay that are indicative of a player's problem solving and learning processes within the game environment. Each playtrace is comprised of a subset of these actions, represented by the nodes. This list, seen in Table 1, was reviewed and endorsed by the game's lead designer. Additionally, each node, in the node title, also indicated where, on the game board, the respective action was taken. For this, the game board was split into sections based on the shape of the track and location of pick up and drop off spots, as seen in Figure 3. Each node would contain, after the action name, an indicator of which section of board the action occurred in along with exact coordinates, which could be used to differentiate between same elements in same sections (i.e. two semaphores in section g could be differentiated by their coordinates). For example, a node may say "place semaphore on H:[9,7]." Additionally, each node label would also display which board sections had elements in this already. For example, a node may say "F:[sem2,sig1]" indicating that the second semaphore placed and first signal placed are in section F. How these labels looked in the visualization can be seen in Figure 2. To better convey location and board state information, each playtrace was augmented with images of the board at every action taken, as seen in Figure 4. Visualizations also included the image seen in Figure 3 so that players knew what the sections of the board were.

3.4.2 Reflection Prompts. Our repeated measures study featured two reflection steps, which we refer to as "self" and "peer" reflection. During **self reflection**, a participant was shown their playtrace for their first playthrough of level 7 alongside their playtrace for their second playthrough. An example of this visualization setup can be seen in Figure 5. During **peer reflection**, a participant was shown their playtrace for their first playthrough of level 7 alongside two other players' playtraces, one that was similar to theirs and one that was different. An example of this visualization setup can be seen in Figure 6.

Similarity was determined using *Glyph*'s sequence graph feature [46]. The decision to depict two players, one similar and one different, was to inform the design of an open player modeling system that was in development at the time of this study and that was exploring the value of such an arrangement. The nature of a player's playtrace did not explicitly reflect their competence as players of the game or learners of parallel programming. As Parallel has multiple ways to approach a solution, the length and contents of a playtrace alone do not reflect competency. Exploring the impact of the competency, or perceived competency, of other players' as a part of peer reflection was beyond the scope of this research, which specifically sought to examine the impact of peer reflection itself. We leave exploration of these questions to future work. All participants were competent enough to complete the level.

Participants were told which trace was similar or different. The other players' traces could have been a first or second playthrough, but participants were not informed of this. Visualization setups were unique for each participant and created manually by the researchers using Miro. We acknowledge here that the inclusion of peer data meant that the peer condition exposed players to additional information. We reiterate that the goal of the study was specifically to examine the impact of that information on the reflection and adaptation processes.

During a reflection step, a participant was directed to their respective Miro board and provided with a short video on how to interpret the visualization. They were then asked to respond to a set of questions in a google form. For the peer reflection step these were as follows:

- Please look at your gameplay sequence for your first attempt. Based on your sequence, can you describe how you approached the level?
- Please look at P1's gameplay sequence. Based on their sequence, can you describe how they approached the level?
- Please look at P2's gameplay sequence. Based on their sequence, can you describe how they approached the level?
- Compared to the other players, what went well in your playthrough and why?
- Compared to the other players, what went poorly in your playthrough and why?
- If you were to play this level again, would you do anything differently?

For the self-reflection step these were as follows:

Action	*Definition*
Place Semaphore	The player places a semaphore on the board
Place Signal	The player places a signal on the board
Link Signal and Semaphore	The player links a signal and a semaphore
Test Passed	The player runs a test and it passes
Test Failed	The player runs a test and it fails
Stop Test	The player stops a test simulation before it completes
Stop Submission	The player stops a submission simulation before it completes
Toggle Semaphore	The player locks or unlocks a semaphore
Move Semaphore	The player moves a semaphore to another spot
Move Signal	The player moves a signal to another spot
Destroyed Semaphore	The player destroys a semaphore
Destroyed Signal	The player destroys a signal
Submission Passed	The player submits a solution and it passes
Submission Failed	The player submits a solution and it fails
View Help	The player views the help menu

Table 1: This table showcases all 15 in-game actions used to analyze the players' gameplay.

Figure 2: A playtrace depicted in *Glyph*'s network graph. For readability, we have enlarged the text in the labels.

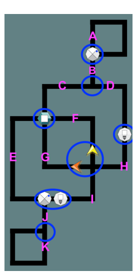

Figure 3: The division of the level 7 board into sections.

- Please look at your gameplay sequence for your first attempt. Based on your sequence, can you describe how you approached the level?
- Please look at your gameplay sequence for your second attempt. Based on your sequence, can you describe how you approached the level?
- Compared to your second attempt, what went well in your first playthrough and why?
- Compared to your second attempt, what went poorly in your first playthrough and why?
- If you were to play this level again, would you do anything differently?

In both conditions, only the last three questions were meant to elicit reflection. The first batch of questions in both conditions existed to ensure that the participant looked at all the playtraces in enough detail to reflect. These reflection questions were derived based on similar questions and prompts used in previous work [13, 18, 36, 43]. All participants completed both reflection steps but

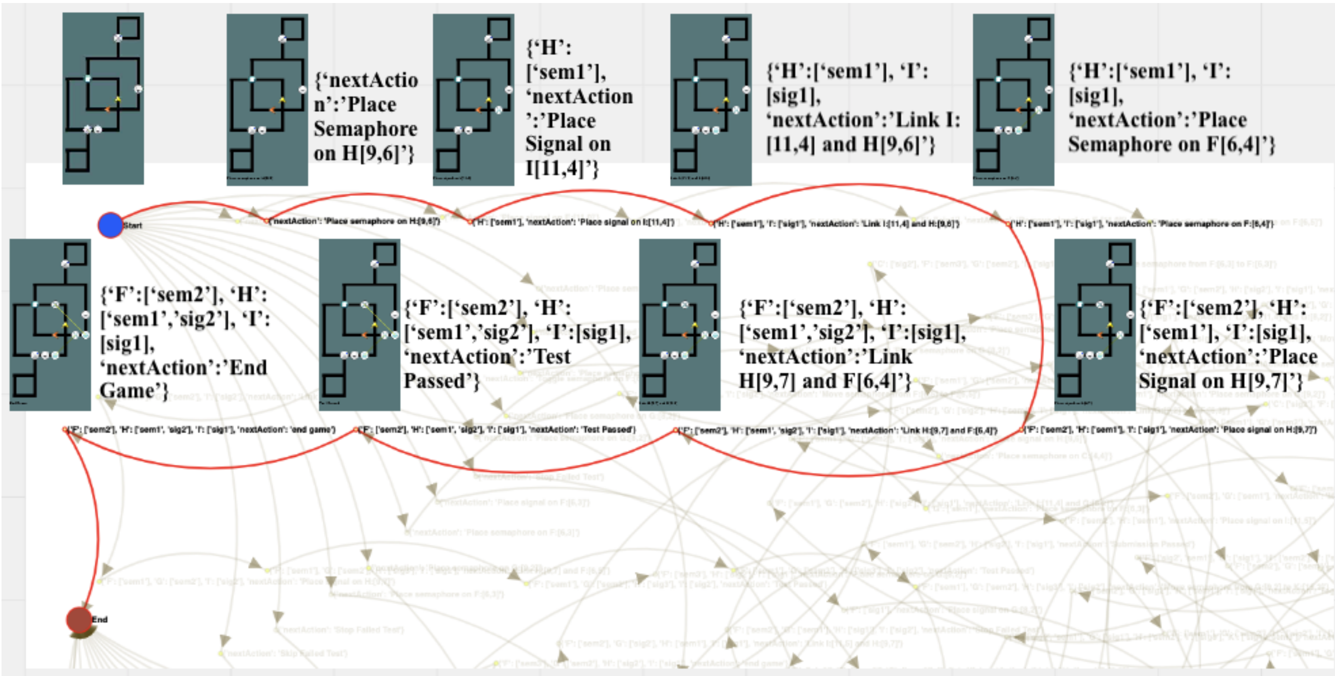

Figure 4: A playtrace with both the node link diagram and the images depicting board state at every action. For readability, we have enlarged the text in the labels.

the order was randomized with half the group doing peer reflection first and the other half doing self reflection first. The last question was answered with a yes or no while all others were open response.

3.5 Data Analysis

Open answer responses to the "what went well?" and "what went poorly?" prompts were analyzed qualitatively using Leijen et al.'s reflection model to measure the quality of reflection [39]. The model quantifies reflection based on focus and level as seen in Table 2. Two researchers worked together to define what these concepts mean within the context of the *Parallel* educational game and concluded on the definitions seen in Table 3.

The same two researchers then, separately, applied the codes to half of the data set. The unit of analysis was a single response and one code for focus and one code for level was applied to each unit. Inter-rater reliability was measured using Cohen's kappa [11] and an agreement of .85 was reached for focus and .88 for level, both indicating very strong agreement [38]. One researcher then coded the entire dataset.

Once the codes were applied, McNemar-Bowker tests [35] were used to determine if there were any significant differences in how often each level and focus was applied to responses in each condition (peer and self). McNemar-Bowker tests were also applied to the "yes/no" responses to the last question, which was used to determine if exposure to community data and reflection through comparison with that data impact willingness to consider a different approach. McNemar-Bowker tests were chosen for this analysis as they are the standard repeated-measures variant of the Chi-Square test for categorical data.

4 RESULTS

The counts for how many participants said they would consider a different approach next time after completing each reflection step (and changes from one step to the next) can be seen in table 4. As can be seen, there is a notable increase in the number of participants who indicated that they would try something else if they played again after reflecting on peer data.

McNemar-Bowker tests revealed that this difference was significant ($p=.004$), with players being significantly more willing to consider an alternate approach next time when reflecting on their own data in the context of others'. Effect size, calculated using Cramer's V, resulted in an effect size value of .5, indicating a large effect.

The counts for the focus and level codes in both conditions can be seen in Tables 5 and 6. As can be seen from the tables, technically focused reflection was most common by a large margin in both conditions and sensitizing reflection was the least common. Similarly, description and justification were far more common levels of reflection across conditions than critique or discussion. The tables also indicate that there was a slight increase in higher quality reflections (those of the sensitizing focus or discussion level) during peer reflection when players were asked to reflect on what went poorly. McNemar-Bowker tests, however, revealed no significant changes between conditions (all $p > .05$). Participants' playtraces were examined to check for any connections between how they played and their willingness to adapt, but no patterns emerged.

Label	***Definition***
Focus	
Technical	Concerned with the efficiency of means for reaching certain goals
Practical	Involves an open examination, not only of means but also of goals, the assumptions goals are based on and the actual outcomes
Sensitizing	Concerned with social, moral, ethical, or political aspects
Level	
Description	Mere descriptions of actions and thoughts
Justification	A rationale or logic for an action or viewpoint
Critique	An evaluation for an aspect and explained why this explanation was given
Discussion	Moving beyond the evaluation and explanation of what is, and why they think that is, and pointed out what could be done to initiate changes, and why changes are needed in the first place

Table 2: Leijen et al.'s [39] model for measuring the quality of reflections.

Label	***Definition***
Focus	
Technical	Discussing efficiency in terms of what the player did, does not include discussion of goals, but may include statement of a goal.
Practical	Discussion of goals, what they were, how they changed, if they were good or bad, etc...
Sensitizing	Discussion of more than just goals and actions taken, thoughts about the player's status as a learner or a player, etc...
Level	
Description	Simply describing what the player did or were thinking
Justification	Providing some kind of explanation or defense or justification for why the player did what they did
Critique	Discussions of how well the player did or any kind of evaluation of their process
Discussion	Any discussion of doing things differently, next steps, what would be done if the step was repeated or done differently next time

Table 3: The definitions we derived for how Leijen et al.'s model applies in the context of Parallel.

Condition	***Yes***	***No***
Would you do anything differently?		
Self	2	34
Peer	12	24

Table 4: Differences in willingness to do something different next time between self and peer reflection.

Condition	***Technical***	***Practical***	***Sensitizing***
What went well?			
Self	27	5	4
Peer	29	5	2
What went poorly?			
Self	28	7	1
Peer	24	9	3

Table 5: Differences in focus of reflection between self and peer.

5 DISCUSSION

While previous work in the learning sciences has demonstrated the value of reflecting on community data as a way to encourage adaptation in learning [24, 56], community data and retrospective visualizations are under-explored in educational games. As a result, the impact of community data on adaptation within the domain has not been empirically examined and, further, previous work in other game genres has suggested that exposing players to peer data may have a negative impact [14]. Based on this, our goal was to determine whether or not comparison with community data in an educational game context had a significant impact on players' willingness to consider alternative approaches and, alongside this, whether or not it had a significant impact on the quality of reflection. We pursued this goal in order to provide researchers with foundational knowledge upon which additional research may be conducted and to arm developers with insights that would allow them to make informed decisions with regards to when to leverage peer data.

Our results indicate that reflection on one's own data in the context of peers' data does, at least partially, motivate adaptation. This is highlighted by 1/3rd of the participants in our study demonstrating a significantly higher willingness to consider a different

Condition	*Description*	*Justification*	*Critique*	*Discussion*
What went well?				
Self	20	11	4	1
Peer	21	10	3	2
What went poorly?				
Self	18	12	4	2
Peer	16	11	2	7

Table 6: Differences in level of reflection between self and peer.

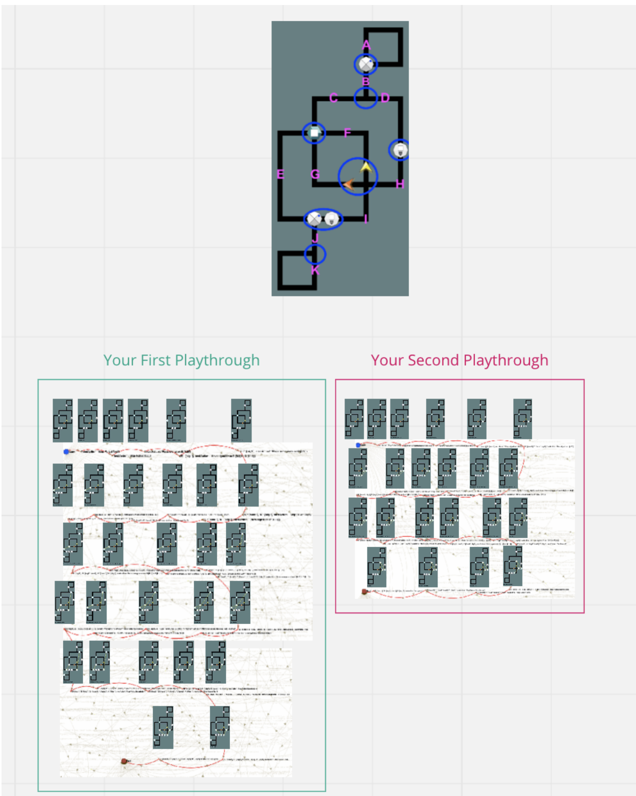

Figure 5: The visualization setup for the self reflection condition depicted the playtrace for the player's first and second playthroughs of level 7.

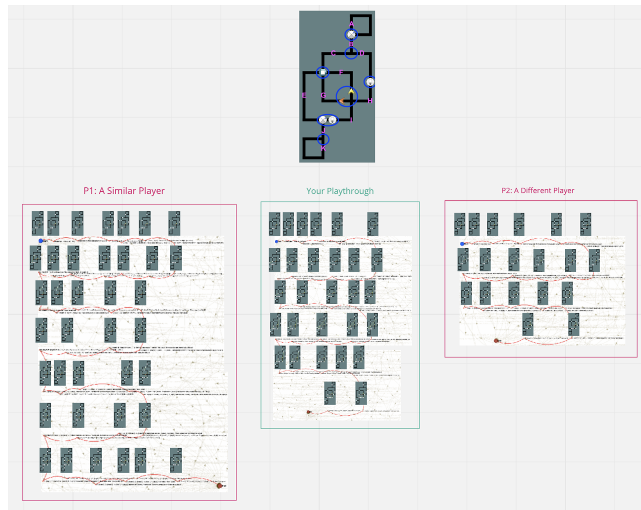

Figure 6: The visualization setup for the peer reflection condition depicted the playtrace for the player's first playthrough of level 7 alongside two other playtraces, one similar and one different.

approach if they were to play the level again after peer reflection. This is an important finding, as previous work in Learning Analytics Dashboards has alluded to willingness to try a different approach as a benefit of community data [17, 50, 57]. This has not, however, been explicitly examined in the context of educational games. It is further valuable to identify this relationship in the context of the concerns surrounding the use of community data in games [14]. This provides empirical evidence that there is a benefit to presenting such data to players, especially in the context of educational games. It can help players, especially those who may be struggling, entertain alternative approaches they may not otherwise consider or perceive and more efficiently arrive at the correct solution. This is especially beneficial in educational game contexts where there may be multiple correct solutions.

Notably, however, only about 1/3rd of participants said they would do things differently. This suggests that, while the presence of peer data has a significant effect, it is not an all-encompassing way to prompt adaptation. It is likely that there are other ways to prompt adaptation both in the context of retrospective visualization and as a part of reflection in general that may result in a greater number of people deciding to try a different approach. Further, there may be individual characteristics, such as confidence or stubbornness, that may impact an individual's willingness to adapt based on comparison to peer data alone, similar to how competitive preferences impact self-monitoring and learning in Riemer and Schrader's work [53]. Based on this interpretation, we present these results as a foundational understanding of the impact of peer data on adaptation. Specifically, we find that peer data encourages adaptation, but not for everyone, and may be one of many possible techniques for prompting change among learners, and may be impacted by individual characteristics. **As such, we suggest that developers of retrospective visualizations for educational games consider leveraging comparison with peer data as a**

way to motivate adaptation, but remain open to alternative techniques as well, as peer comparison alone will not prompt the entire population. Additionally, **we recognize an opportunity for future work to explore the alternative ways that adaptation can be prompted, either as an alternative to peer comparison, or in combination with it, and the ways in which player characteristics influence willingness to change.**

Given these implications, it is important to note that we did not observe a significant change in the focus or level of reflection when peer data was introduced. Overall, reflection for both groups favored the technical focus and the description and justification levels. This reflects the findings of previous work that found that reflection in games often does not rise to the highest levels [42]. This, however, does not conclusively mean that there is no effect of peer data on the quality of reflection. While we controlled for individual differences through a within-subjects design, it is possible that with a larger sample size or different context (perhaps if the players played the game as a part of a class environment) could lead to a significant change. Given the inherent risks of peer data impairing reflection or learning quality, such as by prompting a player to merely copy what they saw without thinking about it or, as demonstrated by previous work [14], the further exploration of this question remains relevant. As such, **we recognize these findings as motivation for the further exploration of the topic in future work.**

Given these implications, we also recognize additional questions about when to expose the community data in order to elicit such change. Schon [58] describes two types of reflection in terms of when they occur: reflection *in action* and *on action*, with the former referring to reflection occurring during an event and the latter occurring after the fact. In our study, we specifically examined reflection on-action, as our interest is in retrospective visualization. It may be that reflection in action produces different results. For example, more participants may be willing to consider an alternative strategy if they have not yet completed the task. It is also possible that quality of reflection could be impacted by peer data when the reflection occurs in the midst of the activity. Such a suggestion also aligns with LAD work that found that students liked seeing other students' approaches so that they could evaluate their standing and adjust their approaches before completion of a course [57]. Reflection in action, however, involves short cycles of thinking and doing, and there may not be enough time in such a structure for a player in an educational game to make meaning of community data, which requires them to consider context and make connections between unfamiliar data points. It may also prompt the above mentioned negative behavior of copying without thinking. As such, **we recognize an opportunity for future work to explore the open question of when, in relation to the progress and completion of a task within an educational game, reflection on community data should be prompted to, not only elicit change, but lead to success.**

Previous work has also found that exposing users to the best approach, in situations where there is a single best approach, may result in conformity among a population [28]. In other words, making the players of an educational game aware of what the best solution is could result in all players following that same solution. In some circumstances, this may be ideal, such as in the contexts of Learning Analytics Dashboards or Open Learner Models where students are encouraged to "follow in the footprints" of other more successful students [25] or when the goal of the game is to help students arrive at and understand a single correct solution. In circumstances where there are multiple correct pathways, or no "right" answer, however, sharing community data that could lead to conformity and may not benefit the students as much as it could reduce creativity or variety. This illustrates open questions regarding how to best present community data to players of an educational game, especially in contexts where there is not a single correct answer or conformity is otherwise not desired. Thus, **we recognize an opportunity for future work to explore how to present community data such that the "best" or "correct" solution is not exposed in such a way that inhibits players' ability to explore and learn naturally. Further, we see opportunities for future work to explore how this presentation should differ depending on the academic context and design of the game.**

Finally, we acknowledge that there exist a number of additional open problems surrounding fair use and fair play with community data. For example, Zhu and El-Nasr [71] discuss how public player data raises concerns of privacy and ethics in open player models. In another example, Kleinman et al. [33], who conducted interviews with esports players, found that the participants in their study were concerned that publicly available data could result in toxic behavior or foul play. While they specifically looked at esports, there findings mirror what was found by Park et al. when they conducted a requirements analysis before designing their learning analytics dashboard [50], suggesting generalizable concerns. While our work here does not address these questions of privacy, fair use, and fair play, we argue that, by demonstrating the value of community data within the domain of educational games, this work motivates the exploration, and hopefully resolution, of these open problems. Thus, **we recognize an opportunity for future work to use these results as motivation to explore open problems surrounding the social and safety concerns inherent in the use of community data.**

6 LIMITATIONS

We acknowledge several limitations of this work. First, we did not examine learning, only the quality of reflection, which is demonstrated by previous literature to contribute to learning. Thus, we recognize the possibility that our findings regarding reflection and adaptation may not be indicative of whether or not the players would actually learn more. We hope to examine this question in future work. Second, our study looked specifically at undergraduate students at American universities. We recognize the possibility that demographic is a variable in the impact of peer data on quality of reflection and that the results may differ for students in other age groups or other regions of the world. Again, we hope to examine this further in future work.

7 CONCLUSION

In this work we examined the impact of comparison with peer data on reflection and adaptation in an educational game in order to inform the design of retrospective visualizations for educational games. We found that comparison with peers led to significantly more willingness to try a different approach, but only in 1/3rd of participants, suggesting that it may be only one of many ways to motivate change. We also did not see a significant impact on quality of reflection, but recognize that these results alone are not conclusive. Based on these findings, we suggest that future retrospective visualizations for educational games can include community data if the goal of the visualization is to elicit change from players, but remain open to alternative approaches as peer comparison alone may not prompt the entire community. We, further identify five opportunities for future work: the exploration of other ways to elicit adaptation in place of or in combination with peer comparison, the confirmation of the impact of peer comparison on quality of reflection, the timing of the interaction with the community data, the presentation of the community data, and the exploration of additional concerns of fairness and privacy surrounding the user of community data. We present these findings as a valuable first step to generating a more comprehensive understanding of how to design retrospective visualizations to enhance reflection, and in turn learning, in educational games.

ACKNOWLEDGMENTS

This work is supported by the National Science Foundation (NSF) under Grant #1917855. The authors would like to thank all past and current members of the project. The authors would also like to thank Tyler Sorensen for his help with recruitment.

REFERENCES

[1] Ana Paula Afonso, Maria Beatriz Carmo, Tiago Gonçalves, and Pedro Vieira. 2019. VisuaLeague: Player performance analysis using spatial-temporal data. *Multimedia Tools and Applications* 78, 23 (2019), 33069–33090.
[2] Sarah L Ash and Patti H Clayton. 2004. The articulated learning: An approach to guided reflection and assessment. *Innovative Higher Education* 29, 2 (2004), 137–154.
[3] Blitz.gg. 2021. Blitz App. https://blitz.gg/. (2021). Online; accessed April 15th 2021.
[4] Robert Bodily, Judy Kay, Vincent Aleven, Ioana Jivet, Dan Davis, Franceska Xhakaj, and Katrien Verbert. 2018. Open learner models and learning analytics dashboards: a systematic review. In *Proceedings of the 8th international conference on learning analytics and knowledge.* 41–50.
[5] Robert Bodily and Katrien Verbert. 2017. Review of research on student-facing learning analytics dashboards and educational recommender systems. *IEEE Transactions on Learning Technologies* 10, 4 (2017), 405–418.
[6] D Boud, R Keogh, and D Walker. 1985. Reflection: Turning Experience into Learning.
[7] Brian Bowman, Niklas Elmqvist, and TJ Jankun-Kelly. 2012. Toward visualization for games: Theory, design space, and patterns. *IEEE transactions on visualization and computer graphics* 18, 11 (2012), 1956–1968.
[8] Sven Charleer, Kathrin Gerling, Francisco Gutiérrez, Hans Cauwenbergh, Bram Luycx, and Katrien Verbert. 2018. Real-time dashboards to support esports spectating. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play.* 59–71.
[9] Sven Charleer, Joris Klerkx, Jose Luis Santos, and Erik Duval. 2013. Improving awareness and reflection through collaborative, interactive visualizations of badges. *ARTEL@ EC-TEL* 1103 (2013), 69–81.
[10] Zhi-Hong Chen, Howard Hao-Jan Chen, and Wan-Jhen Dai. 2018. Using narrative-based contextual games to enhance language learning: A case study. *Journal of Educational Technology & Society* 21, 3 (2018), 186–198.
[11] Jacob Cohen. 1960. A coefficient of agreement for nominal scales. *Educational and psychological measurement* 20, 1 (1960), 37–46.
[12] Vania Dimitrova. 2003. STyLE-OLM: Interactive open learner modelling. *International Journal of Artificial Intelligence in Education* 13, 1 (2003), 35–78.
[13] Magy Seif El-Nasr, Shree Durga, Mariya Shiyko, and Carmen Sceppa. 2015. Data-driven retrospective interviewing (DDRI): a proposed methodology for formative evaluation of pervasive games. *Entertainment Computing* 11 (2015), 1–19.
[14] Jose Esteves, Konstantina Valogianni, and Anita Greenhill. 2021. Online social games: The effect of social comparison elements on continuance behaviour. *Information & Management* 58, 4 (2021), 103452.
[15] Logan Fiorella and Richard E Mayer. 2012. Paper-Based aids for learning with a computer-based game. *Journal of Educational Psychology* 104, 4 (2012), 1074.
[16] Rowanne Fleck and Geraldine Fitzpatrick. 2010. Reflecting on reflection: framing a design landscape. In *Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction.* 216–223.
[17] John Fritz. 2011. Classroom walls that talk: Using online course activity data of successful students to raise self-awareness of underperforming peers. *The Internet and Higher Education* 14, 2 (2011), 89–97.
[18] Gowri Betrabet Gulwadi. 2009. Using reflective journals in a sustainable design studio. *International Journal of Sustainability in Higher Education* (2009).
[19] Allyson Fiona Hadwin, Sanna Järvelä, and Mariel Miller. 2011. Self-regulated, co-regulated, and socially shared regulation of learning. *Handbook of self-regulation of learning and performance* 30 (2011), 65–84.
[20] Neville Hatton and David Smith. 1995. Reflection in teacher education: Towards definition and implementation. *Teaching and teacher education* 11, 1 (1995), 33–49.
[21] Erik Hazzard. 2014. *Data Visualization in Games.* http://erikhazzard.com/blog/game-development/data-visualization-in-games
[22] Caitlin Holman, Stephen Aguilar, and Barry Fishman. 2013. GradeCraft: What can we learn from a game-inspired learning management system?. In *Proceedings of the third international conference on learning analytics and knowledge.* 260–264.
[23] Danial Hooshyar, Emanuele Bardone, Nour El Mawas, and Yeongwook Yang. 2020. Transparent Player Model: Adaptive Visualization of Learner Model in Educational Games. In *International Conference on Innovative Technologies and Learning.* Springer, 349–357.
[24] Danial Hooshyar, Margus Pedaste, Katrin Saks, Äli Leijen, Emanuele Bardone, and Minhong Wang. 2020. Open learner models in supporting self-regulated learning in higher education: A systematic literature review. *Computers & Education* 154 (2020), 103878.
[25] I-Han Hsiao and Peter Brusilovsky. 2017. Guiding and motivating students through open social student modeling: Lessons learned. *Teachers College Record* 119, 3 (2017).
[26] Gwo-Jen Hwang, Han-Yu Sung, Chun-Ming Hung, Iwen Huang, and Chin-Chung Tsai. 2012. Development of a personalized educational computer game based on students' learning styles. *Educational Technology Research and Development* 60, 4 (2012), 623–638.
[27] Chaima Jemmali. 2016. May's Journey: A serious game to teach middle and high school girls programming. (2016).
[28] Jimoon Kang, June Seop Yoon, and Byungjoo Lee. 2022. How AI-Based Training Affected the Performance of Professional Go Players. In *CHI Conference on Human Factors in Computing Systems.* 1–12.
[29] Pavan Kantharaju, Katelyn Alderfer, Jichen Zhu, Bruce Char, Brian Smith, and Santiago Ontanón. 2019. Tracing player knowledge in a parallel programming educational game. *arXiv preprint arXiv:1908.05632* (2019).
[30] Kristian Kiili, Harri Ketamo, and Timo Lainema. 2011. Reflective Thinking in Games: Triggers and Constraints. *Leading Issues in Games Based Learning* (2011), 178.
[31] Erica Kleinman and Magy Seif El-Nasr. 2021. Using Data to" Git Gud": A Push for a Player-Centric approach tothe Use of Data in Esports. (2021).
[32] Erica Kleinman, Nikitha Preetham, Zhaoqing Teng, Andy Bryant, and Magy Seif El-Nasr. 2021. " What Happened Here!?" A Taxonomy for User Interaction with Spatio-Temporal Game Data Visualization. *Proceedings of the ACM on Human-Computer Interaction* 5, CHI PLAY (2021), 1–27.
[33] Erica Kleinman, Murtuza N Shergadwala, and Magy Seif El-Nasr. 2022. Kills, Deaths, and (Computational) Assists: Identifying Opportunities for Computational Support in Esport Learning. In *CHI Conference on Human Factors in Computing Systems.* 1–13.
[34] Athanasios Vasileios Kokkinakis, Simon Demediuk, Isabelle Nölle, Oluseyi Olarewaju, Sagarika Patra, Justus Robertson, Peter York, Alan Pedrassoli Pedrassoli Chitayat, Alistair Coates, Daniel Slawson, et al. 2020. DAX: Data-Driven Audience Experiences in Esports. In *ACM International Conference on Interactive Media Experiences.* 94–105.
[35] Anne Krampe and Sonja Kuhnt. 2007. Bowker's test for symmetry and modifications within the algebraic framework. *Computational statistics & data analysis* 51, 9 (2007), 4124–4142.
[36] Ulrike-Marie Krause and Robin Stark. 2010. Reflection in example-and problem-based learning: Effects of reflection prompts, feedback and cooperative learning. *Evaluation & research in education* 23, 4 (2010), 255–272.
[37] Yen-Ting Kuan, Yu-Shuen Wang, and Jung-Hong Chuang. 2017. Visualizing real-time strategy games: The example of starcraft ii. In *2017 IEEE Conference on Visual Analytics Science and Technology (VAST).* IEEE, 71–80.
[38] J Richard Landis and Gary G Koch. 1977. The measurement of observer agreement for categorical data. *biometrics* (1977), 159–174.
[39] Äli Leijen, Kai Valtna, Djuddah AJ Leijen, and Margus Pedaste. 2012. How to determine the quality of students' reflections? *Studies in Higher Education* 37, 2 (2012), 203–217.
[40] James C Lester, Hiller A Spires, John L Nietfeld, James Minogue, Bradford W Mott, and Eleni V Lobene. 2014. Designing game-based learning environments for elementary science education: A narrative-centered learning perspective. *Information Sciences* 264 (2014), 4–18.

[41] Ben Medler. 2011. Player dossiers: Analyzing gameplay data as a reward. *Game Studies* 11, 1 (2011).

[42] Elisa D Mekler, Ioanna Iacovides, and Julia Ayumi Bopp. 2018. " A Game that Makes You Question..." Exploring the Role of Reflection for the Player Experience. In *Proceedings of the 2018 annual symposium on computer-human interaction in play*. 315–327.

[43] Muhsin Menekse, Saira Anwar, and Zeynep Gonca Akdemir. 2022. How do different reflection prompts affect engineering students' academic performance and engagement? *The Journal of Experimental Education* 90, 2 (2022), 261–279.

[44] Jack Mezirow. 1981. A critical theory of adult learning and education. *Adult education* 32, 1 (1981), 3–24.

[45] mobalytics. 2021. Mobalytics. https://app.mobalytics.gg/. (2021). Online; accessed 25 January 2021.

[46] Truong-Huy D Nguyen, Magy Seif El-Nasr, and Alessandro Canossa. 2015. Glyph: Visualization Tool for Understanding Problem Solving Strategies in Puzzle Games.. In *FDG*.

[47] John L Nietfeld. 2017. The role of self-regulated learning in digital games. In *Handbook of self-regulation of learning and performance*. Routledge, 271–284.

[48] Harold F O'Neil, Gregory KWK Chung, Deirdre Kerr, Terry P Vendlinski, Rebecca E Buschang, and Richard E Mayer. 2014. Adding self-explanation prompts to an educational computer game. *Computers in Human Behavior* 30 (2014), 23–28.

[49] Ernesto Panadero. 2017. A review of self-regulated learning: Six models and four directions for research. *Frontiers in psychology* 8 (2017), 422.

[50] Yeonjeong Park and I-H Jo. 2015. Development of the learning analytics dashboard to support students' learning performance. *Journal of Universal Computer Science* 21, 1 (2015), 110.

[51] Jean H Powell. 1989. The reflective practitioner in nursing. *Journal of Advanced Nursing* 14, 10 (1989), 824–832.

[52] Mladen Raković, Matthew L Bernacki, Jeffrey A Greene, Robert D Plumley, Kelly A Hogan, Kathleen M Gates, and Abigail T Panter. 2022. Examining the critical role of evaluation and adaptation in self-regulated learning. *Contemporary Educational Psychology* 68 (2022), 102027.

[53] Valentin Riemer and Claudia Schrader. 2020. Playing to learn or to win? The role of students' competition preference on self-monitoring and learning outcome when learning with a serious game. *Interactive Learning Environments* (2020), 1–13.

[54] Jonathan Rowe, Bradford Mott, Scott McQuiggan, Jennifer Robison, Sunyoung Lee, and James Lester. 2009. Crystal island: A narrative-centered learning environment for eighth grade microbiology. In *workshop on intelligent educational games at the 14th international conference on artificial intelligence in education, Brighton, UK*. 11–20.

[55] Jennifer L Sabourin, Lucy R Shores, Bradford W Mott, and James C Lester. 2013. Understanding and predicting student self-regulated learning strategies in game-based learning environments. *International Journal of Artificial Intelligence in Education* 23, 1 (2013), 94–114.

[56] Jose Luis Santos, Katrien Verbert, and Erik Duval. 2012. Empowering students to reflect on their activity with StepUp!: Two case studies with engineering students.. In *ARTEL@ EC-TEL*. Citeseer, 73–86.

[57] Jose Luis Santos, Katrien Verbert, Sten Govaerts, and Erik Duval. 2013. Addressing learner issues with StepUp! an evaluation. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge*. 14–22.

[58] Donald A Schon. 1984. *The reflective practitioner: How professionals think in action*. Vol. 5126. Basic books.

[59] Dale H Schunk. 2012. *Learning theories an educational perspective sixth edition*. Pearson.

[60] Magy Seif El-Nasr and Erica Kleinman. 2020. Data-Driven Game Development: Ethical Considerations. In *International Conference on the Foundations of Digital Games*. 1–10.

[61] SENPAI.gg. 2021. SENPAI.gg. https://senpai.gg/. (2021). Online; accessed April 15th 2021.

[62] Raja M Suleman, Riichiro Mizoguchi, and Mitsuru Ikeda. 2016. A new perspective of negotiation-based dialog to enhance metacognitive skills in the context of open learner models. *International Journal of Artificial Intelligence in Education* 26, 4 (2016), 1069–1115.

[63] Linda Valli. 1997. Listening to other voices: A description of teacher reflection in the United States. *Peabody journal of Education* 72, 1 (1997), 67–88.

[64] Gerard Van den Boom, Fred Paas, and Jeroen JG Van Merrienboer. 2007. Effects of elicited reflections combined with tutor or peer feedback on self-regulated learning and learning outcomes. *Learning and Instruction* 17, 5 (2007), 532–548.

[65] Gerard Van den Boom, Fred Paas, Jeroen JG Van Merrienboer, and Tamara Van Gog. 2004. Reflection prompts and tutor feedback in a web-based learning environment: Effects on students' self-regulated learning competence. *Computers in Human Behavior* 20, 4 (2004), 551–567.

[66] Jennifer Villareale, Colan F. Biemer, Magy Seif El-Nasr, and Jichen Zhu. 2020. Reflection in Game-Based Learning: A Survey of Programming Games. In *International Conference on the Foundations of Digital Games*. 1–9.

[67] Guenter Wallner and Simone Kriglstein. 2016. Visualizations for retrospective analysis of battles in team-based combat games: A user study. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play*. 22–32.

[68] Günter Wallner, Marnix Van Wijland, Regina Bernhaupt, and Simone Kriglstein. 2021. What Players Want: Information Needs of Players on Post-Game Visualizations. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–13.

[69] Renee M Williams, Jean Wessel, Maryan Gemus, and Elaine Foster-Seargeant. 2002. Journal writing to promote reflection by physical therapy students during clinical placements. *Physiotherapy Theory and Practice* 18, 1 (2002), 5–15.

[70] Jichen Zhu, Katelyn Alderfer, Anushay Furqan, Jessica Nebolsky, Bruce Char, Brian Smith, Jennifer Villareale, and Santiago Ontañón. 2019. Programming in game space: how to represent parallel programming concepts in an educational game. In *Proceedings of the 14th International Conference on the Foundations of Digital Games*. 1–10.

[71] Jichen Zhu and Magy Seif El-Nasr. 2021. Open Player Modeling: Empowering Players through Data Transparency. *arXiv preprint arXiv:2110.05810* (2021).

[72] Barry J Zimmerman and Adam R Moylan. 2009. Self-regulation: Where metacognition and motivation intersect. (2009).

[73] Barry J Zimmerman and Manuel Martinez Pons. 1986. Development of a structured interview for assessing student use of self-regulated learning strategies. *American educational research journal* 23, 4 (1986), 614–628.