Iyer, S. et al. (Eds.) (2022). Proceedings of the 30th International Conference on Computers in Education. Asia-Pacific Society for Computers in Education

What do Teachers Want to Know About Game-Based Learning Analytics: Cross-Case Study

Yoon Jeon KIM, Jennifer SCIANNA

University of Wisconsin-Madison, U.S.A
yj.kim@wisc.edu

Abstract: Game-based pedagogy requires innovation in learning analytics, so teachers can make sense of the complex data generated in the game in relation to evidence for learning. In this paper, we advocate for co-design as a process to elucidate the types of data that teachers want and need to effectively implement games in the classroom. We examine two cases to discuss affordances of co-design activities and present initial findings that built towards two prototypical dashboards.

Keywords: game-based learning analytics, teacher dashboard, co-design

1. Introduction

Educational games, a specific form of games designed to support and assess students' learning of educationally valuable knowledge and skills, provide a context to expand what learning means and what types of evidence educators can use to better understand what and how students are learning. Learning analytics (LA) and educational data mining (EDM) offer a unique opportunity to translate these insights, that are often hard to manually process, to the teachers who are using games as a learning tool in classrooms. There are many examples of the successful applications of LA and EDM techniques in game-based environments: sequence mining of problem solving processes (Gomez et al., 2021), prediction of how likely a player is to quit (Karumbaiah et al., 2018), prediction of prosocial behaviors (Alonso-Fernández et al., 2020), conceptual understanding of academic domains (e.g. Kim & Shute, 2015), identification of different patterns of engagement (Ruiperez-Valiente et al., 2020), exploration (Ruipérez-Valiente et al., 2019), and even affective states (Kai, 2015).

Despite these successful applications, little is known about how data visualizations of game analytics should be designed to fully actualize the unique pedagogical affordances of games (e.g., Plass et al., 2015). Particularly, the ecological validity of LA/EDM driven metrics in classrooms (i.e., how do teachers perceive and understand these measurements in relation to their own practices) is largely unknown. This paper presents preliminary findings from a cross-case study of two game-based learning projects where two independent research teams engaged classroom teachers as co-designers to create data visualizations for their respective, different educational games. The overarching research question that guided this cross-case study is: "*What did teachers want to know about students' learning, and how did the co-design activities elicit those insights for the design team*?" In this paper, we report (1) what co-design activities the teams used for what purposes and goals and (2) discuss findings related to what kinds/types of analytics that teachers requested for what goals.

2. Relevant Literature

Well-designed games can work as effective assessment (Shaffer & Gee, 2012). The researchers who advocate for game-based learning have recognized that the link between learning analytics for games and teachers' use of them has been largely disconnected (Kim & Ifenthaler, 2019). Because of this need, teacher dashboards have proliferated in the past decade, exploring a wide variety of approaches. For example, the MiGen project consisted of a suite of tools designed around student tracking, classroom dynamics, and goal achievements (Mavrikis, Gutierrez-Santos, & Poulovassilis, 2016). However, critics argue that many of the teacher dashboards for games still miss the mark on balancing complexity and simplicity in a way teachers can use (Duncan, Hoxie, & Miklasz, 2016). One potential reason for this challenge is that researchers' questions might not represent teachers' needs; thus, we might be developing tools that teachers do not really want, need, or understand how to utilize. Indeed, a systematic literature review by Jivet et al. (2018) found that very few dashboards use pedagogical decisions as the input for their design.

In summary, to support teachers' use of analytics from educational games in the classroom, the first step should include research on understanding what kinds of analytics teachers need and for what pedagogical and classroom management purposes. This step is particularly of importance in game-based learning because students learn and demonstrate their learning differently than more traditional modes of learning in schools (e.g., complete a quiz). Thus, the design teams highlighted in this case study sought to address this gap in the research by working directly with teachers in a co-design process to identify their needs as relevant to the unique affordances and pedagogical utility of each game.

3. Cases

Team A: Shadowspect is an online puzzle game where players solve geometry puzzles by selecting, placing, and rotating 3D geometric primitives such as cubes, cylinders, spheres, etc. based on orthogonal views of a figure. In 2020, the research team implemented a 12-month co-design process with 8 practicing teachers (called design fellows) to develop additional analytics and data visualizations while iterating some existing models (since the game was initially developed as an assessment tool).

Team B: Lakeland is a web-based strategic, city-builder game intended for use in secondary science classrooms (grades 7-10) to teach students about nutrient flow through natural and human systems. The game has students manage a lakeside town by producing enough food and ensuring fields stay fertilized while avoiding algae blooms. The design team used early analytics as a starting point to engage the teachers through a 2-day co-design workshop in Spring 2020 where they collaborated to design "dream dashboards." The team then built functional prototypes for testing over several months.

4. Method

Data sources for this paper include co-design sessions transcriptions, artifacts generated by the teachers (e.g. definitions for metrics, hand-drawn diagrams for data visualizations), and research teams analytics memos. Two researchers, one from each team, met after the co-design workshops ended to conduct thematic analysis (Braun & Clarke, 2012) by reviewing all the co-design activities, memos, and transcriptions. The guiding question throughout the process was, "*What did the teachers want to know about students' learning, and how did the co-design activities elicit those insights for the design team*?"

5. Findings

5.1. Similarities and differences in co-design activities

Both teams had three goals for codesign activities that eventually led to data visualizations using game analytics: (1) to elicit what teachers value and wish to know from gameplay, (2) to operationalize these values, constructs, and design metrics/indicators and (3) to provide insights for data visualizations. We discuss two examples, one from each team, to illustrate how co-design activities were used.

Team A utilized a variety of generative design tools/methods adapted from general design tools/methods (e.g. User Personas) to elicit deep values and desires related to using games in classrooms (i.e. what kinds of skills and attitudes do teachers wish to better promote by using games?). For example, Team A used the scenarios and storytelling method (Kumar, 2012) to ask the design fellows to identify meaningful sequences from the gameplay data using a sequence mapping tool (Figure 1). Each icon (e.g., eyes, cubes) represents a game event (i.e. Eyes mean "the player used the perspective tool) that the team visualized in advance to be able to reconstruct game play. The label, "Take another look and another, and another," as well as the text below the sequence is generated by the teachers.

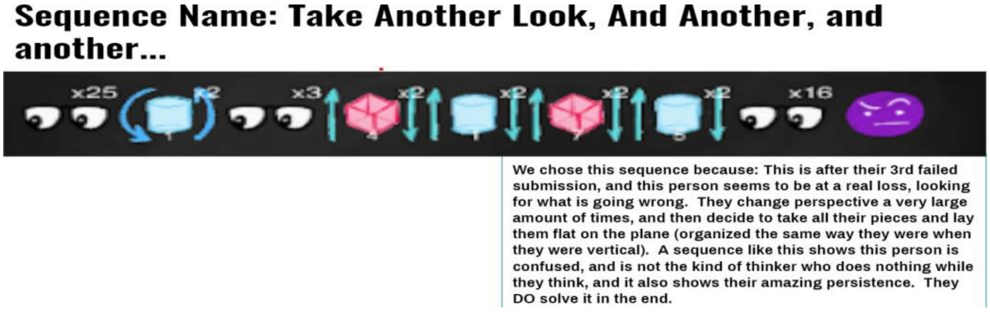

Figure 1. A sequence label created by teachers.

Team B began by having teachers observe one another playing Lakeland and recording moments of interest: any actions that evoked curiosity. These observations scaffolded teacher thinking about what they might want to "see" in their dashboards. Teachers were then provided with an activity (Figure 2) to expand on how the moments could become useful metrics for their dashboard. Each participant finally combined their metrics into a "Dream Dash" to explain the pedagogical utility of the elements.

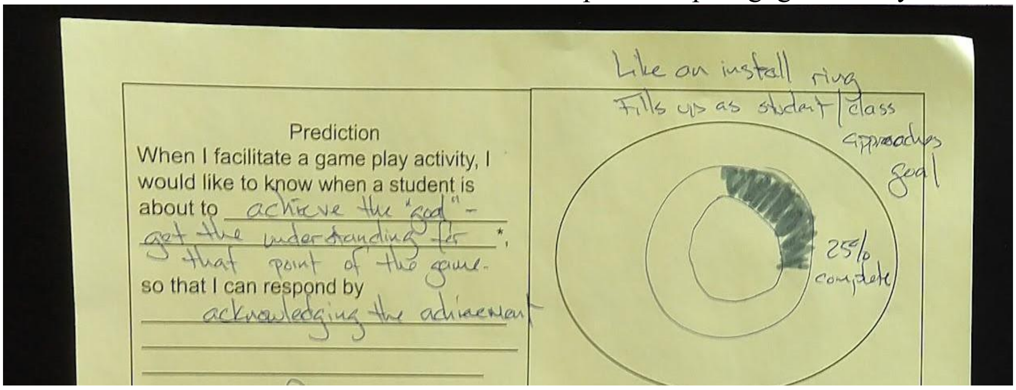

Figure 2. Scaffolding Sheet used by Team B.

5.2. Types of Teacher Needs

The following section describes four themes that emerged from comparing teachers' expressed needs across the two cases and the pedagogical implications associated with each theme.

Theme 1: How do I know how the student is progressing?

The design fellows for Team A discussed ways of thinking about progress. They expressed a strong need for diversifying achievements beyond completing levels: knowing what the overall pattern of progression is for different levels--classroom (e.g. on average, how many puzzles have the class worked on so far?), small groups based on filters (e.g. students who haven't logged into the game in last 2 days), and individual students (e.g. How many puzzles has Marisol tried and completed so far?). Transparency of these progress metrics to the students was also valued by the fellows to keep students motivated.

The fellows also recognized that there are many nuances about progress; therefore, they had a hard time defining what counts as evidence for progress. For example, fellows expressed concerns about students who are quickly solving the puzzles and "flying through" because this might mean that they are not learning anything new from the playing game. A design fellow said, *"So, I look at it from the perspective of, if they're just getting it right, and they're flying through, then they're not really learning, they already know it.* […] *That sweet spot of learning is where you struggle, learn from your mistake, and then now you've learned something, and now you can progress."*

Related to this notion that you can't really learn in the game unless you've made mistakes, the teachers expressed the need to know if/how students "revisit" problems even if they didn't solve them right away. A fellow from Team A said, "*Yeah, I almost wonder if, also, like ... So, say a student is on Puzzle B, they go on for five minutes, get stuck, you know, leave, but then go back into Puzzle B. And seeing if you compare the time and the actions of that first experience from the second and third. And so, like, if there's any progress."* They further elaborated a need to understand the characteristics of puzzles (e.g., difficulty) to make meaning with progress. For example, if a student is only solving easy puzzles, then the teachers can gently nudge them to try more challenging ones.

One of the challenges Lakeland presented for teachers is that student progression appeared to be nonlinear, allowing students to move towards any one of four defined (i.e. growing their town vs. growing their food production) and numerous undefined goals. One teacher reported that their student had kept their farm relatively small, so they could see how long it would sustain itself. Traditional progress metrics would likely flag this student as not progressing, but their ability to balance elements of their town demonstrated their mastery of the underlying system. Several teachers articulated a desire to capitalize on their own intuition, knowledge of their students, and ability to understand the game interface/actions by requesting a way to "see" what students had added to their town; visualizations such as a mini-map for each student or a linear counter to show buying activity allowed teachers make their own inferences about student growth.

In summary, both groups of teachers wanted to have analytics and data visualizations that help them to (1) flexibly define what progress means and what types of actions are being considered as evidence, (2) visualize how students are progressing, both linearly and nonlinearly, (3) visualize progress based on different types of achievement, so the students can see them.

Theme 2: What should we celebrate even if it wasn't successful?

Teachers who want to use games in the classroom are more likely to value and understand that there are multiple ways of solving problems. Teachers desire to better understand the intricacies of student thinking, so they can notice and celebrate it, even if the student is not "successful" (i.e., completed all tasks and puzzles). This need arises because games allow students multiple points of entry to demonstrate learning and as such motivate those who are not always academically successful, ultimately having a positive impact on learning (Karakoç et al., 2020).

One of Team A's design fellows considered the meaning behind achievements and success saying, *"Okay. So, let's see. So we picked ... So, if we just call this Achievements, just to give it, like, a measurement kind of thing, like measuring your achievements, like measuring your badges, I guess. Or we could call them Successes, like we talked about before, how it's not just your success of, like, answering the puzzle, but all these different things."* They were struggling with the tension of wanting to recognize more than one narrowly defined term could capture. Fellows also openly expressed the need to see how their less well-performing students shine through the game, *"...So I really don't necessarily care about kids who are doing well. I do like your use case, where if there is someone who has a really creative strategy, I would love for them to be able to share out with the class, as a great example of another way to approach it."*

Similarly, the teachers emphasized the need to recognize and celebrate "efforts" and "persistence": "*Revisiting something that you struggled with and trying to figure out what your misconceptions were, or where you were incorrect. And I think revisiting is definitely something that could be interesting in terms of a data collection, as well as just showing student or a teacher that this student bailed on this puzzle. But at the same time, two days later, they returned to it, and you know, that's good. We should give a shoutout or something to that.*" When it came to visualizing persistence, however, the teachers were less interested in knowing who's most persistent, instead valuing the description of different groups, i.e., persisting productively vs non-productively, so they can intervene.

Like Team A teachers, Team B teachers were quick to acknowledge that there were many creative ways to play the game. Teachers talked about "showcasing student solutions to the class" and "shouting out to the whole class" as actions that they would want to take in response to students achieving milestones, demonstrating proficiency, or even having a "lightbulb moment". In their Dream Dash, some teachers included displays for the SmartBoard to allow students to see others be recognized by the system. The integration of the dash into actionable information for both teacher and student indicates teachers thought of metrics as being useful to informing pedagogical decisions.

One question teachers came back to several times is "Who should be talking to who?" As Team B teachers considered whether metrics could detect and explain a student's strategy, they noticed potential to facilitate partnerships for collaborative discussion. If they could be alerted that two students were demonstrating opposing strategies, the teachers felt they could facilitate a unique instructional moment if those two students discussed progress. Similarly, when a student identifies a particularly effective strategy, potentially indicated by overall student success, other students who are struggling or frustrated may benefit from pairing temporarily with them to hear how they developed their strategies.

Furthermore, there is one moment that was important to teacher practice that moved away from the focus on individual students and instead centered around the entire class. Team A teachers wanted to know when the majority of students were confused, so they could stop instruction and reteach the directions or provide scaffolding. Similarly, one teacher from Team B wanted to add a "Whole Class" view to their Dream Dash, so they could have both the in-depth analysis of an individual student as well as the birds' eye view of the class.

Theme 3: How do I know if my student is actually learning something?

As teachers became engaged with thinking about metrics and dashboards, the most common request centered on progress as performance measures. Team B teachers wanted to know simple counts of how many things players had within their towns as a way of probing the complexity (or progress) of their system. Initial visualizations of these metrics included many different forms of progress bars, but one teacher conceptualized progress rings as being moving targets for varying parts of the game. As the player fills up this ring, the teacher could acknowledge the achievement, and the ring would reset for the next challenge.

Fellows from Team A also had ongoing questions and concerns about how they can differentiate "learning math" from "progressing" in the game. In contrast to their attitudes about other metrics, the fellows were consistent about how students' performance related to math standards should be viewed using the mastery of learning framework, therefore, wanted to know "how close the student is mastering a standard." For example, a fellow even suggested having a tier system where, *"...And one thing that I found interesting with the idea of like mastery, is there are other games I've played where, when you complete a level, it kind of like does a quick grade of how well you did it, and it's like one to three stars, right? And it's just interesting that, for Shadowspect, it's basically, if you get it right, they're just like, 'Cool, you've finished.'"*

When it comes to standards, the fellows were also more comfortable with more traditional visualizations, such as a bar graph. A fellow expressed, *"...like, puzzles have certain aspects of standards. And so, like, just completing a puzzle, you're making one step closer to mastering that Common Core Standard. And so, it's like, that's kind of more of like a completion, but it is still growth, right? Like, as they're going through and completing levels, they are getting closer and closer to that mastery bar."*

Theme 4: How do I know my students' common errors and misconceptions?

The design fellows in Team A expressed the need to understand what common errors the students are making, and how these errors are indicative of their underlying mathematical misconceptions. First, the fellows suggested that they might benefit from having a video replay function embedded in the dashboard to filter out these students and review what errors they are making. However, with a prototype that included the video replay functionality, the fellows realized that they might not need that much information at individual student level. A fellow said, *"Because of this puzzle like that information to have quickly, at my fingertips would be good. But I don't know that I need the detail of how many students made [errors] for each individual type. I just need to know for my instruction, which types I should review as a class, if it seems that cross-section errors on that puzzle are an issue."*

Team B teachers voiced a desire for action-based metrics that might indicate misconceptions. For example, one teacher requested a monitoring metric for knowing when students had multiple algae blooms, so they could respond by, "*asking is they understand the relationship of what is causing… If they have an idea how to prevent."* Similarly, another teacher wanted to detect when students "*placed fertilizer after a washout,"* so they can respond by, "*discussing how they think that cause & effect rela[tionship] will affect their town."* Teachers wanted to know specific moment-to-moment actions instead of a broader label like "Nutrient Flow Misconception".

6. General Discussion

Using a co-design process, both teams sought to identify what teachers believe to be meaningful, pedagogically relevant features for a data dashboard. From the cross-case examination, we uncovered four themes that can assist in informing others wishing to undertake similar tasks. We acknowledge that there is often tension in the co-design process between the end user (teachers) and the design/development team; creating dashboards and supporting analytics is no exception due to its technical nature. Similar efforts in the community documented similar challenges regarding eliciting meaningful input from teachers to create analytics and dashboards (Prieto-Alvarez, Martinez-Maldonado, & Anderson, 2018). However, by undertaking co-design as a sense-making and skill-building activity, the design teams were able to elicit useful feedback and insights into teaching practice that they otherwise would not have been privy to.

While not fully investigated in this paper, we speculate that this is a challenge fundamentally related to teachers' assessment literacy. Playful learning environments offer rich and continuous interactions with the environments, and in turn, complex data about not just what students know, but how they do it (Dicerbo, Shute, & Kim, 2017). Many teachers do not have experience using these forms of assessment or considering the volume of data that can be generated from them. Thus, the work of a design team is closely linked to thoroughly grasping teachers' motivations for data (e.g. "I want to know who is stuck or falling behind to penalize them vs. "I want to know who is stuck, so I can provide useful help."). The co-design process must present metrics/models in the context of teaching practice and should be closely tied to the actions that teachers want to take based on that information. Partnering with teachers to make meaning of their practice with respect to assessment more holistically, such as providing scaffolded prompts like Team 2's Fill in the Blank activity, can assist designers in better meeting the teacher's needs. Knowing what teachers want to know without knowing what they want to do with the information is not going to be helpful.

While preliminary, we could identify common themes, in terms of how different co-design activities are utilized by the teams, and what teachers need to know from analytics from educational games to fully amplify pedagogical affordances of games. In the future work, the team will continue to analyze the data to yield a set of generalizable design principles (Fu, Yang, & Wood, 2015) in terms of data visualization with specific attention to what actions and decisions that teachers should/need to make when they are implementing games in classrooms.

Acknowledgements

This material is based upon work supported by the National Science Foundation (STEM+C #1935450). Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

Alonso-Fernández, C., Martínez-Ortiz, I., Caballero, R., Freire, M., & Fernández-Manjón, B. (2020). Predicting students' knowledge after playing a serious game based on learning analytics data: A case study. *Journal of Computer Assisted Learning, 36*(3), 350-358.

DiCerbo, K., Shute, V., & Kim, Y. J. (2017). The future of assessment in technology rich environments: Psychometric considerations. Learning, design, and technology: An international compendium of theory, research, practice, and policy, 1-21.

Fu, K. K., Yang, M. C., & Wood, K. L. (2015, August). Design principles: The foundation of design. In International Design Engineering Technical Conferences and Computers and Information in Engineering Conference (Vol. 57175, p. V007T06A034). American Society of Mechanical Engineers.

Gomez, M. J., Ruipérez-Valiente, J. A., Martinez, P. A., & Kim, Y. J. (2020, October). Exploring the Affordances of Sequence Mining in Educational Games. In Eighth International Conference on Technological Ecosystems for Enhancing Multiculturality (pp. 648-654).

Kai, S., Paquette, L., Baker, R. S., Bosch, N., D'Mello, S., Ocumpaugh, J., ... & Ventura, M. (2015). A Comparison of Video-Based and Interaction-Based Affect Detectors in Physics Playground. *International Educational Data Mining Society.*

Karakoç, B., Eryılmaz, K., Özpolat, E. T., & Yıldırım, İ. (2020). The effect of game-based learning on student achievement: A meta-analysis study. *Technology, Knowledge and Learning,* 1-16

Karumbaiah, S., Baker, R. S., Barany, A., & Shute, V. (2019, October). *Using epistemic networks with automated codes to understand why players quit levels in a learning game. In International Conference on Quantitative Ethnography* (pp. 106-116). Springer, Cham.

Kim, Y. J., & Ifenthaler, D. (2019). Game-based assessment: The past ten years and moving forward. In *Game-Based Assessment Revisited* (pp. 3-11). Springer, Cham.

Kumar, V. (2012). *101 design methods: A structured approach for driving innovation in your organization*. John Wiley & Sons.

Mavrikis, M., Gutierrez-Santos, S., & Poulovassilis, A. (2016, April). Design and evaluation of teacher assistance tools for exploratory learning environments. In Proceedings of the Sixth International Conference on Learning Analytics & Knowledge (pp. 168-172).

Owen, V. E., & Baker, R. S. (2020). Fueling Prediction of Player Decisions: Foundations of Feature Engineering for Optimized Behavior Modeling in Serious Games. *Technology, Knowledge, and Learning, 25*(2), 225–250.

Plass, J. L., Homer, B. D., & Kinzer, C. K. (2015). Foundations of game-based learning. *Educational Psychologist, 50*(4), 258-283.

Prieto-Alvarez, C. G., Martinez-Maldonado, R., & Anderson, T. D. (2018). *Co-designing learning analytics tools with learners. In Learning Analytics in the Classroom* (pp. 93-110). Routledge.

Ruiperez-Valiente, J. A., Gaydos, M., Rosenheck, L., Kim, Y. J., & Klopfer, E. (2020). Patterns of engagement in an educational massively multiplayer online game: A multidimensional view. *IEEE Transactions on Learning Technologies, 13*(4), 648-661.

Shaffer, D. W., & Gee, J. P. (2012). The right kind of GATE: Computer games and the future of assessment. Technology-based assessments for 21st century skills: Theoretical and practical implications from modern research, 211-228.