Kim Y.J., Lin G. and Ruipérez-Valiente J.A. (2021) Expanding Teacher Assessment Literacy with the Use of Data Visualizations in Game-Based Assessment. In: Sahin M., Ifenthaler D (eds) Visualizations and Dashboards for Learning Analytics. Advances in Analytics for Learning and Teaching. Springer, Cham

DOI: 10.1007/978-3-030-81222-5_18

EXPANDING TEACHER ASSESSMENT LITERACY WITH THE USE OF DATA VISUALIZATIONS IN GAME-BASED ASSESSMENT

Yoon Jeon, Kim, University of Wisconsin-Madison, yj.kim@wisc.edu
Grace Lin, MIT Education Arcade, gcl@mit.edu
José A. Ruipérez-Valiente, University of Murcia, jruiperez@um.es

Abstract: The use of learning analytics (LA) in educational technology has emerged as a key interest with the promise that this technology will help teachers and schools make data-informed decisions that were not feasible without big data and AI-driven algorithms. Despite its potential, LA has not yet effectively connected research and practice broadly, and it is yet to understand how research-based advances in LA can become accessible assets for teachers, and often LA tools are generally not aligned with teachers' needs. To see the real impact of LA in classrooms, the first step is to understand teachers' literacy for using sophisticated technology-enhanced learning systems that use algorithms and analytics. In this chapter, we present a framework that enables a collaborative design and development process for learning analytics and data visualizations, specifically using games developed for learning and assessment purposes. Using a 3D puzzle game, Shadowspect, the team has been exploring a balanced design of data visualization that considers teachers' needs and desires as well as their assessment literacy. In this chapter, we define what it means to be assessment literate in the context of game-based learning and assessment, and present a process of creating data visualizations with teachers as co-designers presenting several use cases. This chapter can contribute to establishing the foundations of how to design dashboard systems for learning games that can lead to broad use of game data in classrooms.

Keywords: Data visualization, game-based assessment, teacher co-design, learning analytics, data literacy

1. BACKGROUND

Why do teachers value games for classroom instruction? How do they want to use games in classrooms? In a 2014 national survey (Takeuchi & Vaala, 2014), the participating teachers reported that they value using games because students can be more engaged and motivated as well as games can support social emotional learning in addition to academic standards. Similarly, a report from the A-GAMES Project (Teachers Use Games as a Formative Assessment Tool) highlights that teachers often use games as formative assessment by looking at students' performance in the game or asking them questions based on their game play (Fishman et al., 2014). In both reports, teachers responded that the importance of selecting games that are aligned with academic standards while they recognized games can be useful to measure and support skills beyond that.

Because games have unique affordances as a learning and assessment tool, understanding teachers' beliefs, knowledge, and desires about the use of games in the classroom should be a priority. This is particularly of importance to create assessment models and visualizations of assessment data in games because teachers' assessment practices are closely connected with their pedagogical beliefs (Lim & Chai, 2008). Therefore, even though the teachers might not draw a direct connection between their practices and the literature on game-based learning, the assessment literacy in the context of game-based learning should account for teachers' ability to fully leverage affordances of games in terms of data and assessment.

Thus, what are these affordances? First, games implement rich and complex problems that require a lot of trial and errors and creative problem-solving (Gee, 2003; Shute et al., 2009). Therefore, games can be a great environment to elicit evidence for not just content knowledge but also related cognitive and reasoning skills in a multi-dimensional manner. Second, because of the very nature of games as an interactive environment, they capture the full process of learning and solving problems, instead of capturing evidence at one time point unlike how assessment is typically done at the end of unit or lesson. Therefore, teachers should understand that game environments provide evidence based on the process, not just based on something that students do at the end of the gameplay (Kim & Ifenthaler, 2019). Third, teachers should understand that specific actions and choices in the game can be linked to non-cognitive skills and dispositions, different strategies, different problem-solving styles, how they collaborate with other players in the game and how they are progressing in the game. For example, given the "pleasantly frustrating" nature of the game (Gee, 2004), games can encourage learners to persist through difficult problems, and persistence has been well documented as one of the skills that games can be good at supporting and measuring (DiCerbo, 2014; Ventura & Shute, 2013).

Fortunately, many of these affordances can be available in games environments via the rapid processing of click stream data thanks to the advancement of learning analytics techniques and applications of artificial intelligence. The application of data science techniques in educational games is becoming widespread in recent years. In a systematic literature review (Alonso-Fernandez, Calvo-Morata, Freire, Martinez-Ortiz, & Fernández-Manjón, 2019), authors reported that learning analytics and EDM techniques are used to predict performance or assess learning, to study in-game behaviors, to validate game design, and to produce student profiles, and these techniques include a wide variety of models including decision trees, regression models, correlation, and clustering. For example, sequence mining—a data mining method to discover sequences of actions—can be applied in the game environment to unveil for teachers how the learner has been interacting with the game (Gomez et al., 2020; Kim & Shute, 2015). Similarly, data-driven algorithms can be created to identify when students are not productively engaged in the game (Owen et al., 2019). These techniques, through classification models, can also be used to predict which students are struggling, therefore, more likely to quit (Karumbaiah, Baker, & Shute (2018). Moreover clustering techniques can be used to extract students' profiles based on their activity with the game, and provide formative feedback based on the findings. However, as previous authors have raised, game learning analytics is not informagics, and strong pedagogical foundations are required to avoid confounding learning behaviors with game behaviors that does not add value to the learning process.

Despite these affordances, however, there are only a dearth of game-based learning systems that are widely used in classrooms to teachers' assessment of students' learning beyond content standards. Given that teachers are not used to some of these metrics and constructs and that they often don't have access to these data, there is a disconnect between the potential affordances and the practical affordances of game-based learning systems or assessments. To support teachers to fully leverage rich affordances of games for assessment, one solution is providing these analytics coupled with visualization dashboards, which can make concepts teachers care about visible, raise their awareness, and allow them to make pedagogical decisions based on the visualized data (Martinez-Maldonado et al., 2020). Therefore, these visualizations in game-based environments can present a strong opportunity to support teaching, learning, and assessment (Ifenthaler & Erlandson, 2016).

One of the proposals from the community has been to make the end-user more central in the learning analytics design process, with approaches such as human-centered learning analytics (Buckingham Shum et al., 2019) or participatory design (Prieto-Alvarez et al., 2018). Moreover, while visualization dashboards represent an unprecedented opportunity to improve the learning process, they also require the teachers that will consume them to have certain assessment and data literacy capabilities that were previously not required. This shortage of guidance for developing data literacy among end-users has been depicted as one of the main challenges of learning analytics (Tsai & Gasevic, 2017). Additionally, to create learning analytics and visualizations for and with teachers, the field needs to re-imagine what assessment literacy is aiming to support. Unlike teachers' assessment literacy with conventional forms of assessment, game-based environments also require teachers' ability to critically evaluate how the system is processing the data.

This chapter reports a work that is situated at the intersection of these two problems—the limited use of games for learning in classrooms and creating learning analytics and supporting tools to enhance practices on the ground. While multiple studies used learning analytics techniques in games, for example to examine how students are collaborating with each other (Ruipérez-Valiente & Kim, 2020), to function as game-based assessment purposes (Kim & Ifenthaler, 2019), or to model learning behaviors within the game (Kang et al., 2017), teachers' implementation of games coupled with learning analytics in classrooms are still somewhat limited. One of the barriers is the lack of actionable assessment data, the fact that teachers often do not have a clear sense of how students are interacting with the game, and if the gameplay is leading to productive learning (Martınez et al., 2020).

2. ASSESSMENT LITERACY IN GAME-BASED LEARNING AND ASSESSMENT

The recent demand for classroom teachers' data literacy is driven by multiple factors such as pushing for data-driven decisions in schools and government policies that require data-driven decision-making, and this demand has been accelerating with the increasing availability of big data in education (Mandinach & Gummer, 2013). Data literacy can be broadly defined as the ability to understand and use data effectively to inform decisions (Mandinach & Gummer, 2013). It is composed of a specific skill set and knowledge base that enables educators to transform data into information, and ultimately into actionable knowledge (Mandinach et al., 2008) including (a) knowing how to identify, collect, organize, analyze, summarize, and prioritize data, (b) knowing how to develop hypotheses, identify problems, interpret the data, and (c) knowing how to determine, plan, implement, and monitor courses of action.

Teacher's assessment literacy, that can be viewed as a subset of data literacy (Mandinach & Gummer, 2013) where the primary source of data is assessment, incorporates teachers assessment knowledge base (e.g. different goals and types of assessment, pedagogical beliefs, reasoning and communication skills (Xu & Brown, 2016) while their practices are often continuous compromises between what they know and believe and the influence and needs of other stakeholders (e.g. school's priorities, parents). Assessment literacy includes also teachers' ability to interpret data using statistical models (DeLuca et al., 2016b) as well as evaluate the quality of assessment based on psychometric qualities (e.g. reliability).

We also should note that data literacy is often confused with or interchangeably used with assessment literacy. However, the distinction between data and assessment literacy in the context of technology-enabled data rich environments is blurry. That is, while many of these environments provide rich raw and descriptive data (e.g. when did the student log in last time? How long did the student play the game during the last log-in), these systems also use algorithms and artificial intelligence to process data into meaningful categorizations or predictions (e.g. which students are at risk of falling behind?). These sense-making has been viewed as part of teachers' assessment literacy in the conventional notions.

The meaningful use of data from technology-enhanced data-rich environments, such as digital games, in classrooms requires skills and mindsets beyond the conventional notion of assessment literacy skills. For example, one common element of the existing assessment literacy is the teachers' use and understanding of measurement theories and properties (i.e. psychometrics) (DeLuca et al., 2016a). It is very unlikely that teachers will handle scoring data obtained from game environments as well as evaluate psychometrics qualities of the measurement models (i.e. algorithms). Also, use of AIs in such technological environments require the teachers to understand and examine how data are being processed. Therefore, the field needs a better understanding of teacher assessment literacy that interacts with technology and big data to create data visualizations and algorithms that can foster evidence-informed teaching practices.

Moreover, because of the nascence of the learning analytics as a field and lack of emphasis on innovative assessment in pre-service teacher education, it is unrealistic to assume that classroom teachers would feel comfortable with the use of learning analytics coupled with rich technological environments. Even with the conventional assessment tools and data, many studies reported that teachers do not feel prepared to use data to inform their practice (Earl & Fullan, 2003; Ikemoto & Marsh, 2007), struggle with the use of data (Huguet et al., 2014), and lack a sound of understanding measurement models (Oláh et al., 2010). Similarly, simply providing teachers with data visualizations might not be sufficient to address these challenges. For example, Means and colleagues worked with 52 individual teachers and 70 small groups of school staff to investigate teachers' challenges with data-informed decision-making. While most teachers were capable of finding information on a graph, they experienced difficulties comprehending complex data visualizations and showed a limited understanding of key statistical concepts of test validity, score reliability, and measurement error, leading to invalid inferences (Means et al., 2011). In addition, teachers might have challenges in using students assessment data to improve their instruction (Goertz et al., 2009).

In summary, to fully leverage the affordances of digital games and rich data affordances of games in classrooms, the field needs to envision and test new design processes that can help develop learning analytics tools that can be used by the teachers while scaffolding assessment and data literacy skills to make an impact with them. In this chapter, we discuss the needs for re-examining what teacher assessment literacy in the era of big data and educational technology, especially in the context of game-based learning and assessment. In the following sections, we introduce a framework for research and development of learning analytics and visualizations to consider teacher assessment literacy. We situate our discussion within the Shadowspect project to illustrate how we considered different aspects of assessment literacy in addition to teachers' pedagogical goals and purposes to engage teachers in a collaborative design process.

3. CONTEXT: SHADOWSPECT DASHBOARD PROJECT

Shadowspect is a 3D geometry puzzle game where players construct a figure that matches various silhouettes with different geometric shapes (i.e., cube, sphere, pyramid, cylinder, cone, ramp). The silhouettes represent the cross-sections of the figure from different angles. In the game, players can scale and rotate the shapes, change the camera angle to view the figure they are constructing from different perspectives, and take snapshots of their figure that would produce a silhouette of their figure from the selected camera angle. Once players submit their solution, they will be able to see which (if any) of the silhouettes were matched. Figure 1 displays a sample screenshot of the game interface. There are nine tutorial basic level puzzles, nine intermediate puzzles, and 12 advanced-level puzzles, and players can jump to any puzzle they would like to try.

Figure 1. A Puzzle from Shadowspect

Figure 1. "Bird Fez" is a puzzle from the intermediate level. Thus, more hints and constraints are in place for the players, e.g., "You can add 4 more objects." The objective is to create a figure that would match all three of the silhouettes displayed on the top of the screen. The buttons for shape manipulations are laid out on the bottom of the screen. The top right cube lets the player select the camera angle. The current view is from a top/front angle. Once a shape has been inserted (picture on the left), players will have the opportunity to take a snapshot with the camera

button. When a player hits the "submit" button, they will receive feedback on which of the silhouettes (if any) are matched (picture on the right).

No teachers participated as co-designers during the development of the game itself, but the development team determined, with input from a few math teachers, a set of constructs that the team is embedding in the game for using the evidence-centered design framework (Kim et al., 2019). While developing and refining assessment models of the game, the team began thinking about ways to make data assessment output usable for teachers, and this led to the expansion of the project with the goal of creating a generalizable framework to develop data visualization tools for game-based learning. The data visualization in Shadowspect project involves eight middle school math teachers as co-designers who participate in a year-long co-design program where (1) the teachers inform different types of analytics and models that are useful in the context of using Shadowspect in classrooms, (2) teachers co-create and refine different functions and visualizations to match with their decision-making processes, and (3) engage in various participatory design activities to inform iterative prototyping. These teachers, whom we call "design fellows," were selected because they have high interest and ample experience with game-based learning and assessments. One fellow, for example, was involved in the development of ASSISTments—a math platform and tool for assigning and assessing homework.

4. DESIGN FRAMEWORK FOR DATA VISUALIZATIONS FOR TEACHERS

This chapter reports a framework that researchers and designers can consider to design learning analytics models or select modeling techniques accompanied with visualization tools to support pedagogical decisions in the context of game-based learning. We are constraining our scope specifically to the game-based learning, rather than technology-enhanced learning environments broadly, to acknowledge unique affordances of game environments for the kinds of learning, behaviors, and patterns that could be limited either in traditional assessment or less open-ended technology-enhanced environments (e.g. tutoring systems, learning management systems).

The overarching questions that drive our research that interconnects games for learning, learning analytics, data visualization, and teachers' assessment literacy are: *how can different types of learning analytics and algorithms be developed in collaboration with classroom teachers to inform instructional and assessment practices? How should these data be presented, so teachers can make sense of often hard-to-comprehend algorithms? How can we create visualizations to be aligned with teachers' desires while unveiling new insights about learners that might not be apparent to teachers?*

To guide this inquiry, we propose the following framework (Figure 1) to make decisions about the extent to which and the points at which we engage teachers through the development of data visualizations and data analytics models (or computational assessment models). To illustrate how this design framework can be used to guide a research team's efforts to (1) plan for research and development activities, (2) iterate learning analytics and visualization over time in relation to each dimension of the framework, and (3) develop a series of co-design activities, now we discuss this framework with examples from the Shadowspect project.

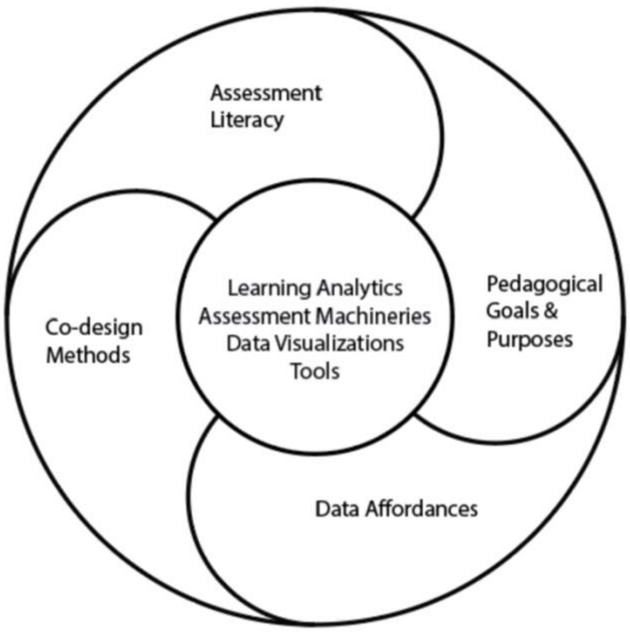

Figure 2. Four Dimensions of Designing Data Visualizations and Analytics Model in Game-Based Assessment

4.1 Assessment Literacy

To create a meaningful analytics model and data visualization, the researchers and designers first need to define what they mean by assessment literacy for game-based learning, which then can help them to define whom the target user is and clarify for or with whom the team develops these tools. In our case, we defined our assessment literacy in game-based learning as follows:

A teacher with assessment literacy in the context of educational games (1) value non-academic, nontraditional and process-oriented skills and attributes of learners that game environments can afford supporting, (2) understand what these constructs mean and identify possible evidence for those constructs based on students' gameplay, (3) critically and curiously investigate how the data was processed based on what rules and understand the role of computing and artificial intelligence and its limitations even if not fully understand how the algorithms are being built, (4) use data and visualization tools to identify strengths, weaknesses, growth, productive and unproductive struggles of learners beyond proficiency and strive to gain new and delightfully surprising insights about learners that they couldn't see with traditional forms of assessment, and finally (5) explore and dig the data at various levels (i.e. individual, subgroup, classroom, grade) and diverse goals (e.g. what's the puzzle that everybody is struggling with, so I can intervene?).

After establishing the definition, then we decided who our target users are. We identified our target user group as teachers who are already on board with the values of video games or open-ended learning environments such as simulations for learning and assessment, and have interests in alternative forms of assessment. These teachers might be using games in classrooms already and looking for opportunities to bring assessment that gets students' interests and creativity. This is a different target user group, for example, from the teachers who don't particularly value games, and not interested in using more data for her own practices.

This operating definition of assessment literacy can also guide the research team to determine which aspects of assessment literacy that the analytics models and visualization tools intend to foster. That is, without a clear vision for teachers' assessment practices that one can better support by creating visualizations, it is difficult to articulate specific functions and purposes of data analytics and visualizations. In our case, establishing the assessment literacy helped us to come up with an initial set of design principles as described in Table 2:

Table 1. Design principles based on our definition of assessment literacy.

The visualization should be easy to navigate and inviting for teachers to dig deeper and play with.
The visualization should foster curiosity of the teacher to explore the data.
The data that teachers saw on visualization should match with their desires and intentions for using games in classrooms.
The visualization should allow the teachers to see multiple aspects of a learner, which might be surprising and unexpected.
The visualization should allow the teachers to see the learners' growth over time.
The visualization should allow the teachers to identify and celebrate productive struggle.
The data visualization should allow the teacher to question how the model was created.

4.2 Pedagogical goals and purposes

To determine a process of developing which analytics models and algorithms and accompanied visualizations, the research team also needs to consider what pedagogical goals and purposes teachers have in mind, i.e. how do they want to use the data for what purposes? This helps to determine the scope and overall direction of the visualization tools. Also, depending on the goal, the qualities of analytics models and the scope of technical development (therefore, how to engage teachers in the process) will vary.

The literature in game-based learning suggests three different pedagogical goals and purposes are commonly observed in classrooms (Fishman et al., 2014). First, games can function as formative assessment. When the goal is formative assessment, teachers might need learning analytics models and visualizations that enable them to identify students who need support, i.e. where they are struggling, and what's the source of struggle. Therefore, for formative assessment goals, rather than providing highly processed decisions or predictions based on algorithms, providing descriptive and fine-grained analytics related to students' performance in the game might be more appropriate. Second, the teacher might choose to use the game as a motivational tool. For this goal, instead of assessing "how well" the student is performing in the game, learning analytics models and visualizations should focus on various types of achievements beyond numbers of completed puzzles and quests. Third, the teacher might want to have students play the game as a form of summative assessment.

In our case, the teachers expressed their desires to use the game as an enhancement tool as well as a formative assessment tool. In addition, given the teachers' intention to implement the game as part of regular math curriculum, they expressed the needs to know how student's performance in the game is related to the math standards and what potential misconceptions the student might hold. In our case, the team aims to develop visualizations and analytics models that can allow the teacher to monitor how productively or unproductively students are making progress in the game, and how their interactions with the game can inform how much students know about specific standards or how they might hold geometric misconceptions. Furthermore, the visualizations and analytics models need to provide actionable insights or information for the teachers to bring back to the classroom. For example, teachers would like to know the most common misconceptions students have made, view representative video playbacks of when these misconceptions occurred (see Figure X), and bring them back to the classroom to facilitate whole-class discussion, or "puzzle talks," as one of our co-design teachers would like to call them.

Figure X. A screenshot of the video playback on the Shadowspect Dashboard. In this example, student 424 has made a total of 32 misconceptions across 2 puzzles. The video is displayed when the user clicks on "Show Full Replay" on the right panel indicating the puzzle attempts.

Additionally, given that every teacher's context is unique, it is critical for the research team to invite teachers in the collaborative process early on to identify their values and priorities, particularly related to what to measure. For example, early on in our collaboration in the set of metric introduction exercises to gauge teacher priorities, we included metrics as abstract as persistence to ones that are concrete and specific, such as sequences of player actions within a puzzle. In midst of the widely varying metrics, the teachers identified persistence as one of their most highly valued metrics that they would like to explore and investigate in the game context.

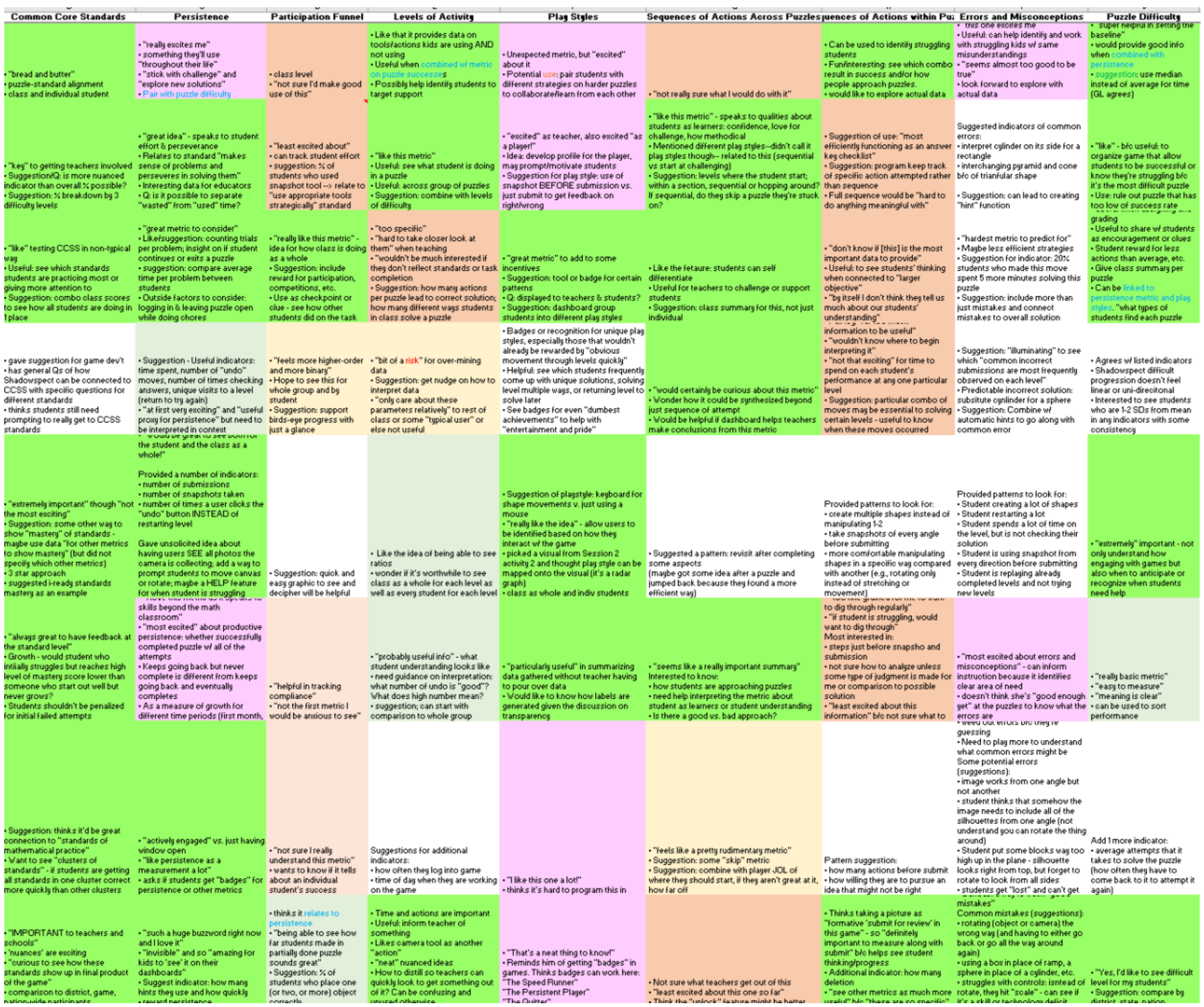

Figure X. A screenshot of one pass of coding through teachers' reflections on various metrics. Each column represents a metric, and the rows are separated by fellows. Green indicates a positive response, pink indicates that the teacher is excited about the metric, beige or yellow indicates uncertainty or slight reservations regarding the metric, and darker orange indicates more negative sentiments. As can be seen on this screenshot, "persistence" is the most well-liked metric with a couple of fellows being excited about it. In contrast, fellows express the most doubts regarding the usefulness of the "sequences within a puzzle" metric. As can be seen by the rows, some fellows are more critical (more oranges, beige, or neutral) while others are more overwhelmingly optimistic (mostly green or pink).

In our case, it was clear that the co-design teachers (as well as our target audience) can look beyond the most immediate, traditional "math scores" and value nontraditional, process-oriented skills that become assessable and accessible through the game-based learning context. As can be seen in Figure X, the co-design teachers' initial responses to the persistence metric is overwhelmingly positive. In fact, the co-design teachers believed persistence to be a great metric to consider because it is an "invisible" (co-design teacher 1) skill that students can use "throughout their life" (co-design teacher 2) and "beyond the math classroom" (co-design teacher 3). It is an important lifelong skill transferable beyond the game, and something that the fellows believe educators must coach students with. As one teacher puts it, "persistence in the face of challenge is what leads us to success" (co-design teacher 4). At the same time, though, the positive desire for insights into students' persistence is juxtaposed with a need for action. As mentioned and illustrated earlier, the teachers want recommendations and next steps to bring back to the classroom. Some teachers (e.g., co-design teacher 1) are apprehensive about how they can help students with "low persistence" as they acknowledge that there could be other life circumstances that prevent the students from playing the game consistently and, therefore, persistently. Therefore, they would like the dashboard to provide meta information such as when the students log in—if they do so at all—to have a more comprehensive picture of students' engagement and situation beyond a simple metric.

4.3 Data affordances

Depending on game mechanics, genres, single play vs. multi-player collaborative, cooperative vs. competitive, how the player can progress in the game environment (i.e. linear or nonlinear), how teachers might implement these games, and the kinds of data one can acquire from gameplay can vary (Groff, 2018). Therefore, data visualizations in the context of game-based learning and assessment should consider possible skills and outcomes that the game is better suited for as well as how the game elements affect the classroom implementation (thus data collection). For example, a single player puzzle game like Angry Birds (Rovio Entertainment) can be great at measuring physics understanding, persistence, and problem-solving, but inappropriate to measure one's collaboration skills by simply using the in-game telemetry data. Similarly, how the game intends to be implemented in the classroom should be considered. For example, a game like Food Fight (BrainPOP) is a turn-based game that has only single player log-in, and it is designed for a pair to share the monitor and use one mouse and take turns. For a game like Food Fight, actionable analytics might focus less on the individual players, but the overall qualities of the food web that was created at the end, which provides insights about the pair's collective understanding of a Savannah ecosystem.

Communicating affordances of the game in terms of what is feasible to measure early on is a key to create co-design activities. For example, in our case, the research team came up with a potential list of what is possible to model and measure, but a few teachers in our co-design cohort were aware that games can be a good context to further illustrate student effort beyond whether they complete the work or not. At the same time, the co-design process should

encourage the teachers to challenge and question what these constructs mean and what "evidence" will be considered to create learning analytics models and visualizations.
The process to communicate the affordances is often cyclical. After the potential list of what is possible to model and measure, we start out with early renditions of visualizations that we dubbed "tools to think with" to facilitate the exploration of the data, which allowed the co-design teachers to better grasp the kind of evidence available and constructs that can be created. Within those confines, the teachers are then able to illustrate the metrics they would most like to see presented on a dashboard.

For example, through the exploration of the early visualizations, it became clear that the co-design teachers had strong opinions on and understanding toward the construct of persistence and were enthusiastic and capable of finding evidence for various "flavors" of persistence from data. These flavors were informed by their increased understanding of the data affordances, as they became aware of what the game data could tell them about students' activity levels within a game, the types of activities in the game (e.g., submitting a solution, taking a snapshot of their constructed figures to check the silhouettes), and the active vs. passive time students spend in the game. As the co-design teachers investigated, they also became adept at navigating through multiple levels and perspectives of data. Figure X is an example illustrating a flavor of persistence—productive persistence to be exact—extracted by a co-design teacher, Melinda, as she explored a radar chart "tools to think with" at an individual student-level as the student's performance is compared to the class average. Typical with usage in the classroom, comparison to the class average or across multiple students may be helpful to identify students who may be struggling—or persisting—more. Along this line, fellows like Melinda utilized Radar Charts to identify students with outliers. As Melinda described it, "...it looks like this student didn't check their solution very often, wanting instead to make sure that they have evaluated the correctness of this solution in every possible way before submitting it. This is evidenced by this student rotating the view many, many times, but not really ever checking the solution. It looks like this student is spending a lot of time attending to the precision of the object."

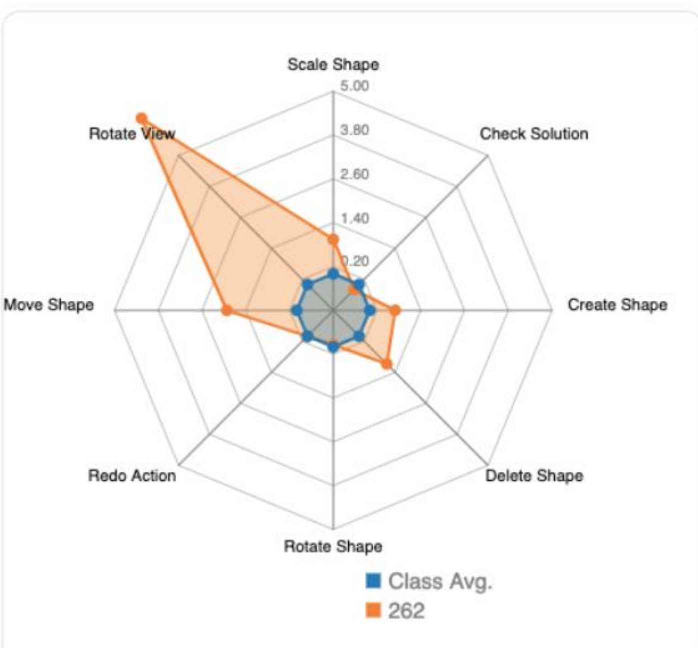

Figure X. An example of productive persistence by co-design teacher Melinda. Student 262's actions in the puzzle Bird Fez as compared with the class average. The Radar Chart display has been normalized.

On a different strain, the design fellows were also interested in students' personal, individual progressions. In one example, a fellow utilized the Radar Chart's Puzzle View function to investigate how a student progressed across puzzles of varying difficulty (see Figure X). In this case, the fellow noted that the student "seemed to complete Pi Henge with ease. Bird Fez was harder for him but he stuck with it with lots more manipulations and snapshots before completing it." This progression of putting more effort into solving a more complex puzzle was indicative of the students' *persistence relative to themselves*. Too often were students being compared to their peers or class average that if the fellows did not investigate the data on this other level, the insight would have been missed.

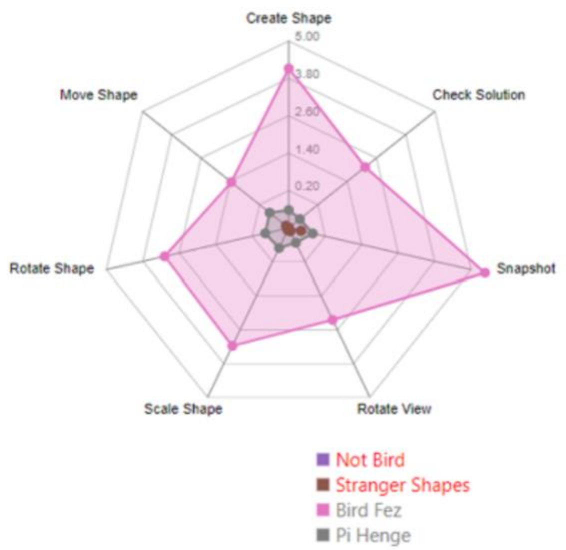

Figure X. In examining a single students' activities across multiple puzzles, Fellow Barbara identified a flavor of persistence that hinges on putting in more effort on more difficult tasks.

The fellows utilized different digital tools to think with, uncovered the affordances of the data, and identified patterns that they believe resemble different "flavors of persistence" they care about. The process resulted in five distinct patterns: (a) actions after failed submission, (b) checking solution or not, (c) precision and detail oriented: checking views, (d) more actions than others, and (e) others: unproductive persistence, lack of persistence, and miscellaneous flavors.

4.4 Co-design methods

While analytics models and data visualization tools can be developed without teachers actively participating in the process as collaborative partners, many argue that using participatory design methods with practitioners can increase the overall usefulness and usability of such tools in classrooms (Buckingham Shum et al., 2019). Co-designing analytics models and visualizations with teachers, however, require different levels of scaffolding mechanics, depending on the target audience's assessment literacy, as teachers often do not have technical skills that are required for model building as well as technology development. Considering specific aspects of assessment literacy and how competent the target audience is also can support the team's decision-making regarding how to structure co-design sessions.

For example, early co-design activities such as metric definition reflections and storytelling showed our co-design teachers had a good understanding of how to define persistence in a general sense, yet they also had both technically feasible and non-feasible ideas about how to identify evidence for persistence based on gameplay and the existing design of the game mechanics. Below is an excerpt from a storytelling activity where the research team asked the co-design teachers to come up with a story of how teachers and students might use Shadowpect in the classroom:

> *Norman is in Miss Greta's class. The class is playing Shadowspect. Miss Greta is trying to monitor the class and the student's progress. Norman completed 3 of the 4 puzzles that map onto the congruence standards. He's doing well. We can tell this because Norman achieved 2 out of 3 stars for the beginning puzzles on that standard. We see that they're completing that standard, but not in a very efficient manner, suggesting that there is some guess and check and exploration still happening. The teacher then encouraged Norman to get back to play the puzzle more in order to get the 3 stars to "full" mastery. This also fits in nicely about persistence because he spent 100 moves to get to 2 stars, but he would return to it later to get to 3 stars by solving in fewer than 15 moves*.

In the excerpt, co-design teachers Barbara, Josh, and Kim expressed the idea of students being able to receive full mastery of 3 stars and would treat students' return to a puzzle—despite fewer moves later on—as a sign of persistence. While it is feasible for the game to track if and when a student returns to a puzzle, Shadowspect does not have a star rating system built in; it is therefore unfeasible to disentangle the reason for which a student may return to a puzzle (e.g., to solve a puzzle more efficiently, to show a friend).
This potential disconnect between the focus of the teachers' design attention and the research team's intention led to a series of follow-up co-design activities that engaged the teachers with a few rounds of collaborative generations of indicators specific to Shadowspect. First, to allow for a fuller understanding of the indicators we can draw from the data, we had the teachers explore with "Tools to Think With." Other than the Radar Charts showcased in the previous section, these digital data explore tools that allow teachers to try out different configurations to unveil what is working and what is not working also include a Caterpillar chart (see Figure XX) that displays types of student activities on a given puzzle against a time scale.

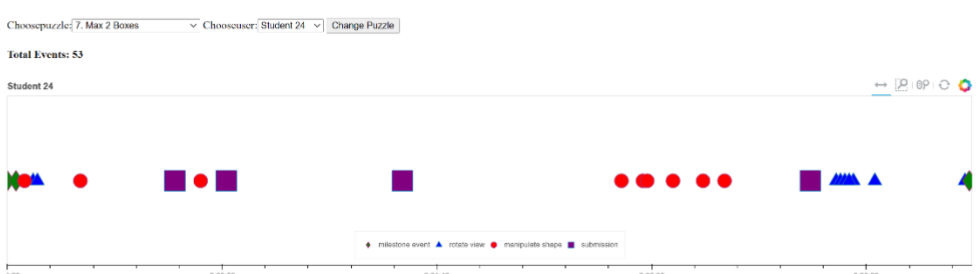

Figure X. An example of a Caterpillar chart. Co-design teacher Tara selected this student as she explored the tool because "[for] this particular puzzle, you can see the student persisted and try to solve the problem multiple times (4 submissions), you can also see the big gap between minute 01:40 and 02:39, probably to consider other modifications, and the many attempts to manipulate the figure showed on amount of red dots.”

Additionally, we introduced a “mad libs” prompt where the teachers re-clarify what their intended use of Shadowspect is and how they envision to use it in their classrooms (Figure X). Based on their response, the teachers work as a pair to specify what kinds of indicators they would consider as evidence for persistence and how they would use them, and this process involves a blend of both generation and “remix” of indicators the research team has extracted from literature and the teachers’ earlier inputs. Figure X displays virtual panel sorting task where the teachers rated the usefulness of various potential indicators as well as some of the teachers’ own remix of indicators. Allowing teachers to see existing indicators and remixing them to generate their own ideas appear to be a productive co-design method. As one teacher stated in a final reflection activity, “I really liked all the opportunities to ‘remix’ because it allowed us to be creative, while being grounded! It also meant that we could quickly iterate because the art/design was there for us to use!”

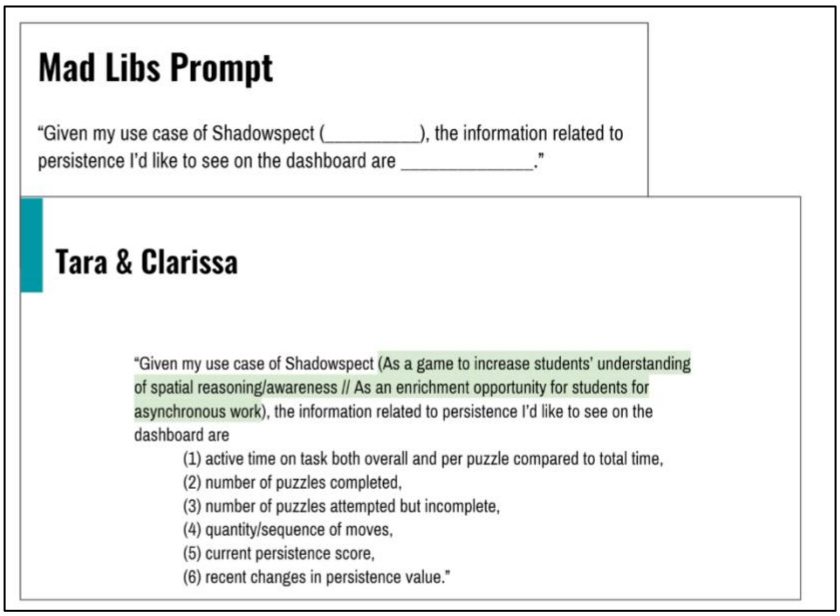

Figure 3.3a. (the picture on the top) is the prompt that was provided at the beginning of a co-design session for persistence, and Figure 3b (the picture on the bottom) is two teachers’ responses.

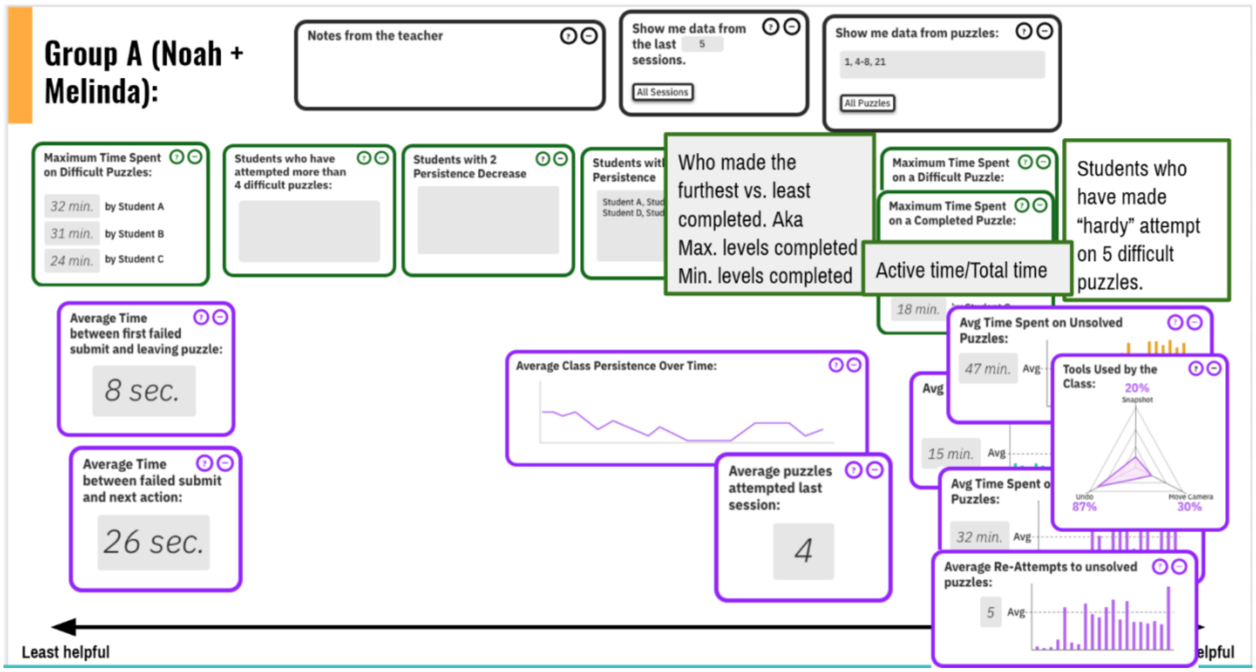

Figure X. The panel sorting activity by Noah and Melinda. The teachers added in a couple more panels for indicators that they believe would be useful in capturing persistence.

5. CONCLUSION

In this chapter, we discussed a framework that the research team developed to create learning analytics and visualizations tools that enable teachers to use gameplay data to support students learning in classrooms. The goal of this framework is to guide designers and researchers to consider four interconnected dimensions—assessment literacy, pedagogical goals and purposes, specific data affordances of the game, and co-design methods—to develop learning analytics models and visualization tools. Based on what we learned from using this framework to plan both research and development activities, we further illustrated how each dimension is connected to each other using examples from the Shadowspect project.

Our goal with this chapter is to encourage researchers to apply this framework and document their process, so the field can continue to grow this body of knowledge, by providing various use cases, that can either be buried in the process or not documented beyond the project. Particularly, we hope that this is a beginning of work where we can expand what teacher assessment literacy means in the era of Big Data and educational technology in the context of open-ended learning environments like games.

We foresee multiple directions of this work in the future. First, the current framework does not explicitly describe teacher learning, while how these models and tools can help teachers to reflect and modify their existing assessment practices should be considered to evaluate the effectiveness of these tools. Xu and Brown (2016 p.156) describe that "becoming assessment literate is fundamentally a transformative, consciousness-evoking one. However, teachers may be content to have conceptions and practices of assessment that are entirely consistent with external contexts without casting doubt on their own practices.” To what extent these visualizations allow teachers to reflect and challenge the current practices can be an additional element of this design framework. In addition, the process of engaging teachers in the process can be also a professional development opportunity to them to build their assessment literacy skills. Future works can investigate different applications of this framework across different contexts and different types of co-designers and target users. Second, in the Shadowspect project, we aimed to build visualization tools that are targeting the teachers who are already on board with the pedagogical affordances and values of the game. Future work should investigate how different definitions of assessment literacy can lead to different co-design activities as well as visualization tools and data models.

6. REFERENCES

Alonso-Fernandez, C., Calvo-Morata, A., Freire, M., Martinez-Ortiz, I., & Fernández-Manjón, B. (2019). Applications of data science to game learning analytics data: A systematic literature review. Computers & Education, 141, 103612.

Buckingham Shum, S., Ferguson, R., & Martinez-Maldonado, R. (2019). Human-centred learning analytics. Journal of Learning Analytics, 6(2), 1–9.

DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016a). Approaches to classroom assessment inventory: A new instrument to support teacher assessment literacy. Educational Assessment, 21(4), 248–266.

DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016b). Teacher assessment literacy: A review of international standards and measures. Educational Assessment, Evaluation and Accountability, 28(3), 251–272. https://doi.org/10.1007/s11092-015-9233-6

DiCerbo, K. E. (2014). Game-based assessment of persistence. Journal of Educational Technology & Society, 17(1), 17–28.

Earl, L., & Fullan, M. (2003). Using data in leadership for learning. Cambridge Journal of Education, 33(3), 383–394.

Fishman, B., Riconscente, M., Snider, R., Tsai, T., & Plass, J. (2014). Empowering educators: Supporting student progress in the classroom with digital games. Ann Arbor, MI: University of Michigan. Retrieved from Http://Gamesandlearning. Umich. Edu/Agames.

Gee, J. P. (2003). What video games have to teach us about learning and literacy. Computers in Entertainment (CIE), 1(1), 20–20.

Gee, J. P. (2004). Learning by design: Games as learning machines. Interactive Educational Multimedia: IEM, 15–23.

Goertz, M. E., Olah, L. N., & Riggan, M. (2009). Can interim assessments be used for instructional change.

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

Groff, J. S. (2018). The potentials of game-based environments for integrated, immersive learning data. European Journal of Education, 53(2), 188–201. https://doi.org/10.1111/ejed.12270

Howard, T. (2014). Journey mapping: A brief overview. Association for Computing Machinery. https://doi.org/10.1145/2644448.2644451

Huguet, A., Marsh, J. A., & Farrell, C. C. (2014). Building teachers’ data-use capacity: Insights from strong and developing coaches. Education Policy Analysis Archives/Archivos Analíticos de Políticas Educativas, 22, 1–28.

Ifenthaler, D., & Erlandson, B. E. (2016). Learning with Data: Visualization to Support Teaching, Learning, and Assessment. Technology, Knowledge and Learning, 21(1), 1–3. https://doi.org/10.1007/s10758-015-9273-5

Ikemoto, G. S., & Marsh, J. A. (2007). chapter 5 Cutting Through the “Data-Driven” Mantra: Different Conceptions of Data-Driven Decision Making. Yearbook of the National Society for the Study of Education, 106(1), 105–131.

Karumbaiah, S., Baker, R. S., & Shute, V. (2018). Predicting Quitting in Students Playing a Learning Game. International Educational Data Mining Society.

Kang, J., Liu, M., & Qu, W. (2017). Using gameplay data to examine learning behavior patterns in a serious game. Computers in Human Behavior, 72, 757–770. https://doi.org/10.1016/j.chb.2016.09.062

Kim, Y. J., & Ifenthaler, D. (2019). Game-Based Assessment: The Past Ten Years and Moving Forward. In D. Ifenthaler & Y. J. Kim (Eds.), Game-Based Assessment Revisited (pp. 3–11). Springer International Publishing. https://doi.org/10.1007/978-3-030-15569-8_1

Kim, Y. J., Ruipérez-Valiente, J. A., Philip, T., Louisa, R., & Klopfer, E. (2019). Towards a process to integrate learning analytics and evidence-centered design for game-based assessment. Companion Proceedings of the 9th International Learning Analytics and Knowledge Conference, 204–205.

Kim, Y. J., & Shute, V. J. (2015). The interplay of game elements with psychometric qualities, learning, and enjoyment in game-based assessment. Computers & Education, 87, 340–356.

Lim, C. P., & Chai, C. S. (2008). Teachers' pedagogical beliefs and their planning and conduct of computer-mediated classroom lessons. British Journal of Educational Technology, 39(5), 807–828.

Mandinach, E. B., & Gummer, E. S. (2013). A systemic view of implementing data literacy in educator preparation. Educational Researcher, 42(1), 30–37.

Mandinach, E. B., Honey, M., Light, D., & Brunner, C. (2008). A conceptual framework for data-driven decision making. Data-Driven School Improvement: Linking Data and Learning, 13–31.

Martinez-Maldonado, R., Elliott, D., Axisa, C., Power, T., Echeverria, V., & Buckingham Shum, S. (2020). Designing translucent learning analytics with teachers: An elicitation process. Interactive Learning Environments, 1–15.

Martınez, P. A., Gómez, M. J., Ruipérez-Valiente, J. A., Pérez, G. M., & Kim, Y. J. (2020). Visualizing Educational Game Data: A Case Study of Visualizations to Support Teachers.

Means, B., Chen, E., DeBarger, A., & Padilla, C. (2011). Teachers' Ability to Use Data to Inform Instruction: Challenges and Supports. Office of Planning, Evaluation and Policy Development, US Department of Education.

Oláh, L. N., Lawrence, N. R., & Riggan, M. (2010). Learning to learn from benchmark assessment data: How teachers analyze results. Peabody Journal of Education, 85(2), 226–245.

Owen, V. E., Roy, M.-H., Thai, K. P., Burnett, V., Jacobs, D., Keylor, E., & Baker, R. S. (2019). Detecting Wheel-Spinning and Productive Persistence in Educational Games. International Educational Data Mining Society.

Prieto-Alvarez, C. G., Martinez-Maldonado, R., & Anderson, T. (2018). Co-designing learning analytics tools with learners. Learning Analytics in the Classroom: Translating Learning Analytics Research for Teachers, 93–110.

Ruipérez-Valiente, J. A., & Kim, Y. J. (2020). Effects of solo vs. Collaborative play in a digital learning game on geometry: Results from a K12 experiment. Computers & Education, 159, 104008. https://doi.org/10.1016/j.compedu.2020.104008

Shute, V. J., Ventura, M., Bauer, M., & Zapata-Rivera, D. (2009). Melding the power of serious games and embedded assessment to monitor and foster learning. Serious Games: Mechanisms and Effects, 2, 295–321.

Takeuchi, L. M., & Vaala, S. (2014). Level up Learning: A National Survey on Teaching with Digital Games. Joan Ganz Cooney Center at Sesame Workshop.

Tsai, Y.-S., & Gasevic, D. (2017). Learning analytics in higher education—Challenges and policies: A review of eight learning analytics policies. Proceedings of the Seventh International Learning Analytics & Knowledge Conference, 233–242.

Ventura, M., & Shute, V. (2013). The validity of a game-based assessment of persistence. Computers in Human Behavior, 29(6), 2568–2572.

Xu, Y., & Brown, G. T. (2016). *Teacher assessment literacy in practice: A reconceptualization. Teaching and Teacher Education, 58*, 149-162.