17. Who Played the Game Correctly?

Data Signatures of Interaction in Playful Assessment

ANTHONY PELLICONE, NATHAN HOLBERT, BETSY DISALVO, VISHESH KUMAR, AND MATTHEW BERLAND

Abstract: In this paper we examine a formative playful assessment called *Beats Empire*, in which learners use data analysis skills to take on the role of a music studio manager. We look at log data of players from classroom implementation of the game, finding 3 signatures: low activity, high activity with low data usage, and data-informed gameplay. Drawing from field notes and play-aloud interviews, we contextualize these data signatures in player conceptions of the game. We find that our data-informed player employed a number of contexts in interacting with the game, such as domain knowledge, game strategies, and perceptions of popular music. Furthermore, we find that money earned within the game, and achieving a win state (while both intuitively good metrics for assessment), are not as informative as gameplay signatures. We conclude by emphasizing that all 3 approaches to play are valid, and that each represents its own design challenge in terms of tuning the gameplay as well as creating supporting material to help players provide formative assessment to various types of players.

Introduction

Beats Empire (*Beats*) is a game designed to be a formative assessment tool to assist teachers in evaluating students' data-science knowledge by simulating a music studio (see Figure 1). Data and analysis skills are measured by how well players sign and develop artistic talent and guide their team of musicians toward creating songs that match the preferred moods, topics, and genres of their audience. The goal is to create a personally meaningful and culturally responsive method of formative assessment in which a player's studio represents an externalization of his or her understanding of the data that are uniquely generated for each play session. Drawing from theoretical foundations of formative assessment (Sadler, 1989), culturally relevant pedagogy (Au & Jordan, 1982), and constructionist games (Weintrop, Holbert, Horn, & Wilensky, 2016), *Beats* is situated in a nascent field known as Playful Assessment (PA) (Kim, 2018). PA refers to the use of games and play to provide formative assessments that are more meaningful for learners as well as providing richer data for instructors to practice formative assessment.

This paper examines *Beats* within the PA framework, using patterns derived from player log data as well as a deeper analysis using qualitative data from player surveys and interviews. We describe three data signatures from our early implementation data: low activity, high activity with low data usage, and data-informed gameplay. Through qualitative analysis of a data-informed player, we observe the various strategies that she employed in playing the game, including using data to make strategic choices, as well as creating in-game artifacts that reflected her personal tastes in music. We find that two metrics that would typically be associated with traditional assessment (money earned through play and achieving a win state) are not necessarily as informative as looking at patterns of interaction within the game. We conclude by discussing how all three approaches are valid, and how each suggests its own approach to configuring gameplay, as well as creating teacher resources to assist in formative assessment.

Theoretical Framework

Beats is designed as a constructionist assessment environment. Constructionism is an approach to pedagogy that engages learners in constructing a personally meaningful artifact that requires knowledge of a domain in order to function. Constructionist approaches provide rapid feedback, allowing for tinkering and gradual refinement of strategies; social feedback from within a community; and authentic problem spaces for construction (Harel & Papert, 1991). Constructionist games are useful for assessing student understanding because they are designed to be culturally relevant to students, necessarily involve a deep understanding of the domain that is being assessed, and present content in an active and engaging fashion (Berland, Baker, & Blikstein, 2014). Furthermore, when designed to provide rich information about how players interact and create in the constructionist game space, such games also serve as a powerful means of formative assessment. Thus, we can think of playful and game assessments as providing "objects-to-think-with," where students are able to put concepts from the target domain into practice through playful tinkering and experimentation (Berland, 2016; Holbert & Wilensky, 2019; Papert, 1980).

We draw the following analysis from an early implementation of the game in a classroom environment. Through discussion among field researchers we noticed a common pattern, in which some players were able to exercise what we termed a "meta-game" approach to *Beats*. Here, *meta-gaming* refers to applying knowledge of common game mechanics, and how those mechanics can be maximized or subverted against the intents of the design (Consalvo, 2009). This led us to a guiding question to approach preliminary data from our classroom implementations, as well as a design question that follows from that.

- *Research Question (RQ)*: How do players' strategies differ in a formative, playful assessment? Specifically, what gameplay patterns can we observe that indicate how players understand the role of data in *Beats*?
- *Design Question (DQ)*: How might these patterns inform the design of future playful assessments?

Design of *Beats Empire*

As described in our theoretical framework, we saw cultural relevance (Ladson-Billings, 1995) as being a primary design consideration for our game, and thus we employed design methodologies to understand not only student interests, but also values and conceptions of the domain (DiSalvo & Bruckman, 2011). Other research has identified music as an important social and emotional tool for youth (North, Hargreaves, & O'Neill, 2000; Tarrant, North, & Hargreaves, 2000). Through our previous experience as educational game designers, and a series of focus groups with our target population, we came to understand music as being a generally resonant theme that allowed for individual expression, but that also carried shared understandings in terms of our salient domain concepts (e.g., the recognition of the music industry using data to inform the creation and marketing of popular songs), and thus made it ideal as an aesthetic for a formative assessment centered on data analysis.

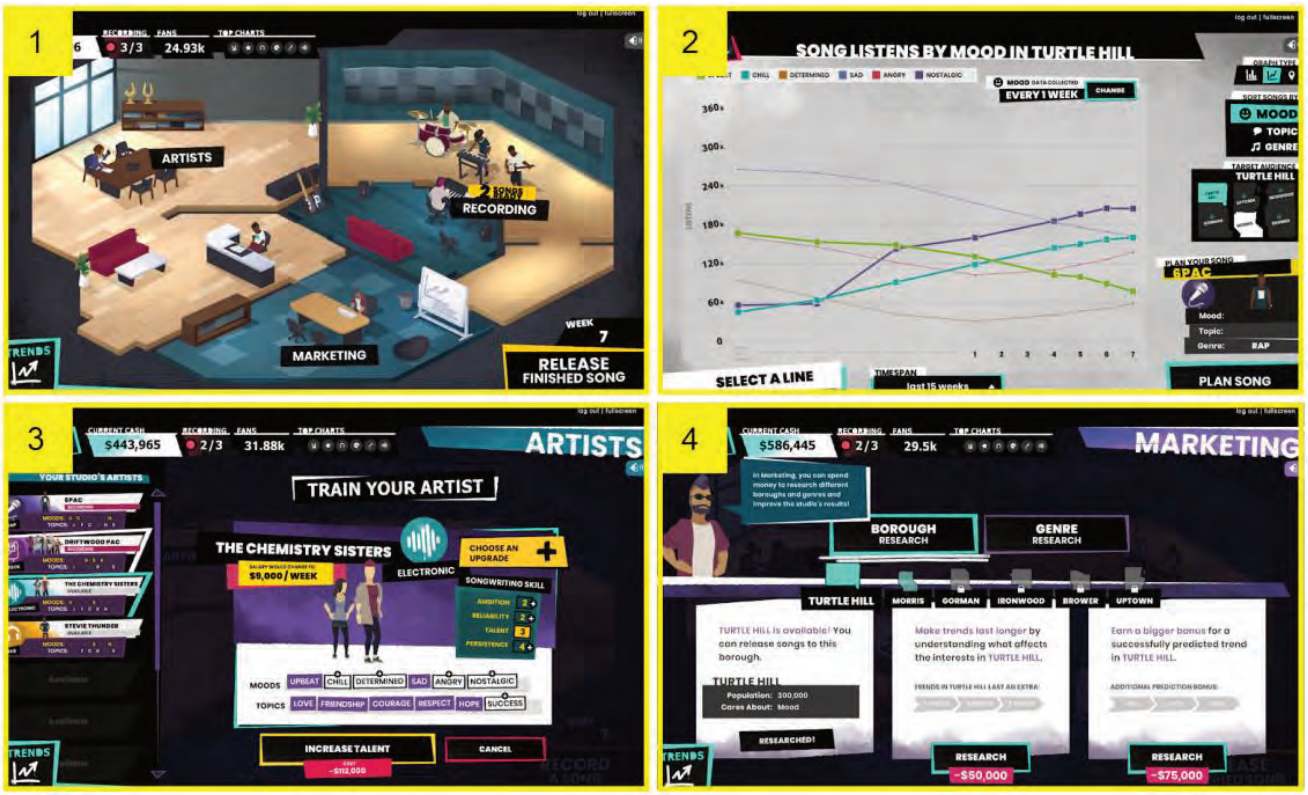

Figure 1. Beats Empire is modeled after management-type games. From the studio (screen 1) players follow trends in moods, topics, and genres (screen 2) and use that information to sign and develop artists who are assigned creating songs (screen 3); players can upgrade various aspects of their studio to increase the reach and impact of their music (screen 4).

Data Analysis

For this paper we are reporting on data collected from the second implementation of *Beats Empire*, which took place in winter of 2019 at a New York City public school with a group of 10 seventh-grade participants. All names used in this paper are pseudonyms. We collected data through automatic log files of gameplay, pre- and postsurveys with students, and one-on-one "play-aloud" sessions with field researchers, in which students narrated their reasoning behind choices in the game, as well as a final focus group with the entire class. Important for this analysis, we used the automatic log data to investigate our guiding question and to understand general patterns in how our participants approached the game. Those signatures then allowed us to explore the qualitative data with individual participants in order to understand their perceptions of gameplay. As shown in Figure 2, we focused specifically on three constructs from the log data: advancing the releasing songs, referencing the data tools, and advancing to the next turn (which in-game uses the metaphor of advancing to the next week).

Findings

Gameplay Signatures

We observed three general play styles that we have here termed as *low-activity play*, *high-activity/low-data play*, and *data-informed play*. In this subsection, we present participant interaction visualizations (see Figure 2) illustrative of each of these styles. Each graph represents the same duration of time (approximately 30 minutes of gameplay on the second day of data collection). On each tick mark, a dot indicates if a player: accessed data visualizations (insightScreens in Figure 2), released a song (songRelease), or advanced the game state to the next week (turns). These actions indicate interactions with core game mechanics and/or target content. Furthermore, an overabundance or an absence of one of the actions suggests quite different play styles, such as a player moving quickly through in-game time in order to build large banks of money (high count of *songRelease* and *turns*), or a player engaging deeply with data visualizations to make informed decisions (high count of *insightScreens*). Additionally, we examine total amount of cash on hand as we have observed players using this metric as a personal measure of success.

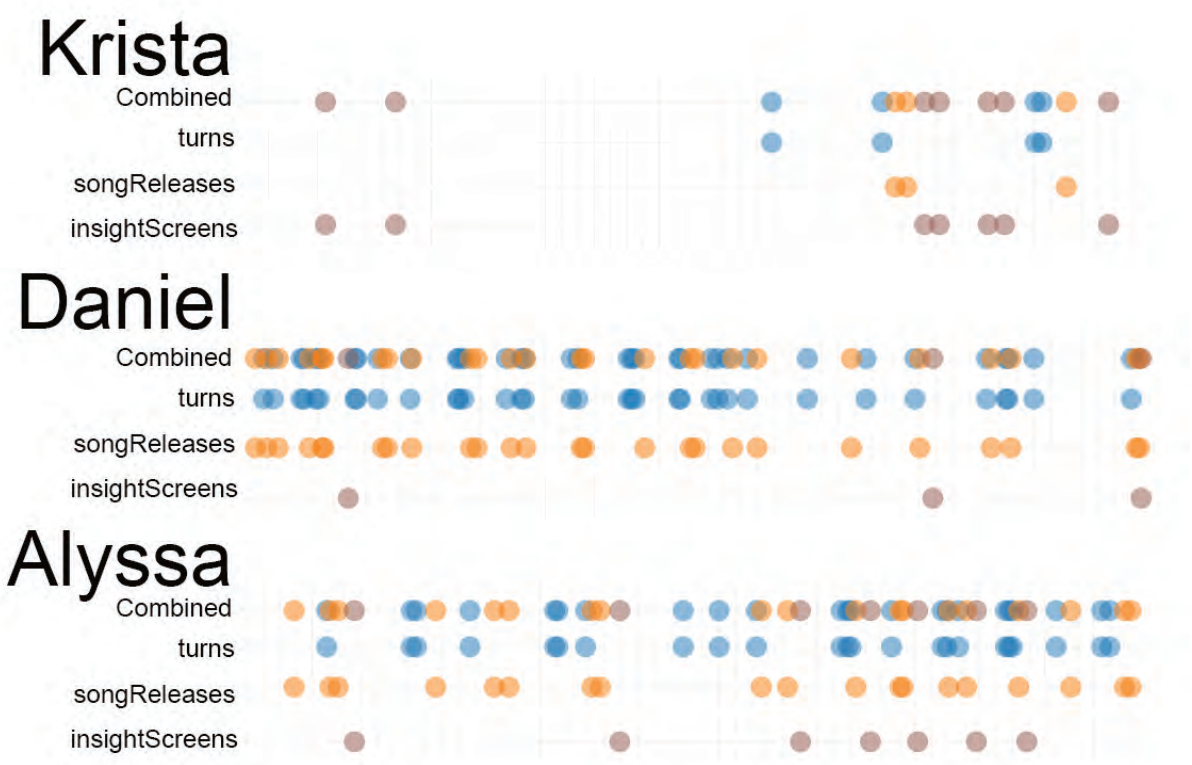

Figure 2. Visualizations derived from player log files. Blue dots on the top line of each player's visualization represent advancing the in-game calendar (turns), orange dots on the middle line represent releasing a song that generates revenue (songReleases), and brown dots on the bottom line represent use of data tools within the game (insightScreens). Below each line is a combined signature, showing a player's total interaction over the course of the 30-minute gameplay session.

Krista, low-activity play. Krista does make use of the data visualization tools, and during her play-aloud interview she justifies several activities with information garnered from the data visualizations of the game. However, after the play-aloud session, Krista also says that she was not very engaged with the game and did not find it appealing. This is reflected in her chart in Figure 2, which shows large gaps in real-world time between actions. From the play-aloud

interview, we know that Krista spent much of the time in that first gap methodically exploring the interface of the game, and then she devoted a few minutes per song picking one of the randomly generated titles. While Krista did draw on some knowledge of data analysis as a domain, she was also not engaged with the mechanics of the game and made very little progress toward either of the game's win conditions.

Daniel, high-activity and low-data play. Daniel amassed a great deal of in-game money during play and is one of the few players to beat the game within the three-day time span. On a survey asking him how often he plays games, Daniel did not report many hours per week of gaming, claiming to play approximately six hours per week. However, Daniel's gameplay was similar to that of gamers who play substantially more frequently. We have found these players often employ a strategy that relies on releasing a large number of songs, and moving quickly through in-game time, in order to earn a large amount of money. In the chart in Figure 2, Daniel looks at the data screens only three times, but he moves through 35 weeks—compared to Krista's looking at data seven times and advancing through four weeks.

Alyssa, data-informed gameplay. Finally, Alyssa represents a type of gameplay that relies heavily on using data visualizations provided in-game to make informed choices about which artists to sign, what type of song to record, and to gradually improve song quality through time.

Alyssa's Empire

In the presurvey Alyssa reported that she played about 20 hours of games per week, later claiming that this was mostly *Fortnite* (Epic Games, 2017) with her friends. When she was asked about her goals in *Beats*, Alyssa described them as "to build my empire." When the interviewer asks her to clarify if she means having higher popularity, the most money, or the best music, Alyssa replies, "All of the above." Alyssa uses data frequently while choosing what kind of song to record. When asked by the researcher why she has chosen the current set of parameters for one such song, Alyssa gives an answer derived from the data, "[I checked] the trends, and um, then I'll see that 'Sad' [a mood] is very popular and also 'Determined' [another mood]." Alyssa also recognizes a mechanical aspect of the game, which is that certain boroughs care only about one element of a song (in this case mood), saying, "Well, the genre doesn't matter, and the topic doesn't matter, the only thing that matters is the mood."

However, Alyssa doesn't use only data in crafting songs. In addition to data, Alyssa also prefers to record music from her favorite genre and often chooses the topic "Love" because that is what she says that she personally enjoys in the music that she listens to. Despite these non-data–driven choices, Alyssa is still able to create a successful music studio and make substantial progress toward winning the game. Her being able to succeed in the game, even when not using data for every decision, indicates that *Beats* allows players to call on their personal understandings of music while still advancing along a path toward victory that the curriculum-inspired mechanics have laid out for them.

It's Not All About the Money

An alternative source of insight for gameplay patterns is a player's total cash on hand. While our game design considers cash to be a resource for players—to invest in opening up new boroughs for song releases, artist and studio upgrades, or signing new talent, and so forth, data from observations and focus groups suggest players also viewed cash as an indicator of success. Examining how players accumulated and used money provides an alternate indicator of how students engaged with game mechanics. Figure 3 visualizes the cash-on-hand totals of our three example players over the same day of play.

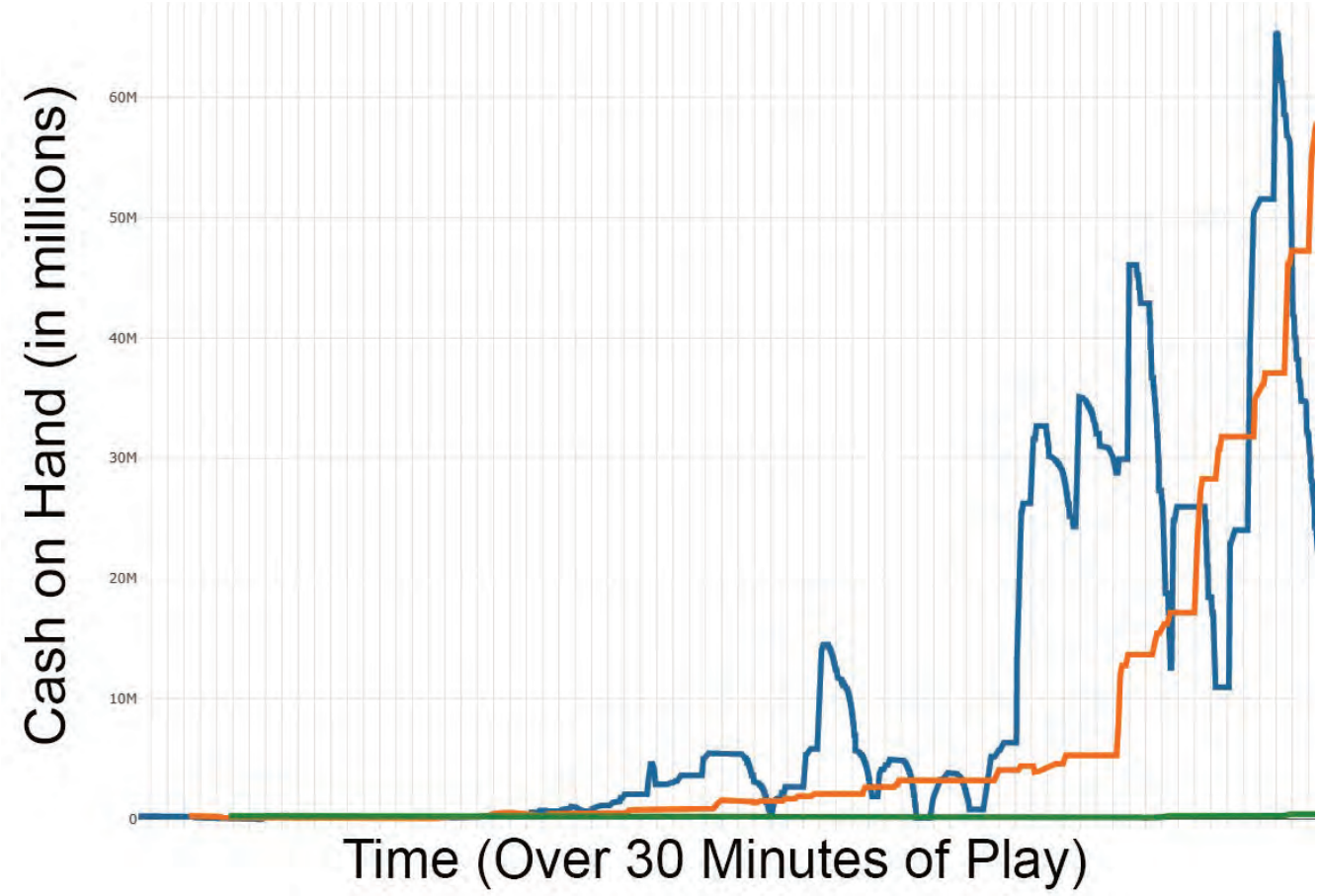

Figure 3. The spikier blue line represents Daniel, the more gradual orange line represents Alyssa, and the green line (which at this scale is close to the X axis) represents Krista.

Daniel often has a great deal of money invested in terms of artist salaries while also recording songs and advancing time at a high velocity, leading his overall bank to both grow and shrink rapidly (the dip at the end of Day 2 is an example of Daniel's having purchased new artists immediately before the end of the session). In the third day, after we have explicitly called out the concept of upgrading to players, Daniel is able to use his large bank, and exponentially expand his earning capabilities, resulting in a victory condition. By comparison, over this same period, Alyssa's studio's growth is steady and always increasing. Alyssa signs fewer artists, releases fewer songs, and does not spend money on upgrades. Despite this more constrained gameplay, because Alyssa uses data to make strategic recording/releasing decisions, she is able to produce roughly equivalent amounts of in-game cash while advancing through fewer in-game weeks.

Discussion and Conclusions

RQ: How Do Players' Strategies Differ in a Formative, Playful Assessment?

Each of the three players described above brings his or her own set of existing identities and literacies as data scientists, gamers, and music fans to *Beats Empire*. Alyssa's case gives a good example of what this integration looks like in

practice—in her play-aloud data we can see her bridging across her experience with digital games, her own cultural understandings of popular music, and her knowledge of data analysis. We have designed *Beats* to accommodate a wide variety of conceptions within these domains. For example, Alyssa has experience with games, and takes quickly to the mechanics, but she still has some gaps in understanding about how to use varying types of data visualization. She interacts with game mechanics strategically, with a stated goal of improving her "empire," but she also allows her personal preferences about music to drive aspects of her play. Our desire to create a culturally relevant constructionist formative assessment game led us to design *Beats* with this particular dynamic in mind. Offering multiple win conditions and including boroughs that have varying interests in moods, topics, or genres, ensures that a player such as Alyssa (who enjoys the genres of rap and R&B) can succeed while creating a studio that specializes in genres just as well as a student who takes a more generalist approach. However, both forms of gameplay enable players to create something that is personally meaningful and situated in a practical application of domain knowledge (Berland, 2016; DesPortes, Spells, & DiSalvo, 2016; Holbert & Wilensky, 2019).

Daniel's case presents us with a signature of a player who is taking an approach to *Beats* that we might think of as "meta-gaming," or using optimized strategies that are broadly applicable across individual games—in this case, quickly amassing a lot of resources while ignoring other elements of the game (Boluk & LeMieux, 2017; Consalvo, 2009). While Daniel spends less time with data tools central to our goals for the game, he is still exhibiting complex, strategic thinking in how he has created his music studio. If we think of *Beats* gameplay as a traditional assessment (in which there are right and wrong ways to play the game) we might conclude Daniel has either failed or in some ways "gamed the system." However, in recognizing digital gameplay as a social practice, we can instead ask, first, how can game mechanics be modified so that students who play this way are *also* encouraged to use data, and second, how can we prepare instructors to recognize the gameplay signature so that they can provide feedback and support to help students to leverage their gaming literacy for also better understanding the target domain (Holbert & Wilensky, 2019; Sadler, 1989).

Krista's case offers a real challenge for the design of playful assessments. Krista does employ data in making decisions, but she also is less engaged with *Beats* as a game experience. Not every player will enjoy every game, and it is helpful to recognize that other forms of assessment may be necessary to complement a playful assessment. It is useful to note that the field researcher administering Krista's play-aloud interview was using less prompting than the researcher administering Alyssa's. However, we can think of Alyssa's nurturing coaching in similar terms to theories of formative assessment, in which feedback helps students to perceive gaps between a desired goal and their current understandings and assists them in closing that gap (Black & William, 1998). The sort of subtle coaching that is present in Alyssa's gameplay session, but absent in Krista's, points toward the importance of data informing supportive, tailored feedback from an instructor (Sadler, 1989).

DQ: How Might These Patterns Inform the Design of Future Playful Assessments?

As our title suggests, a question that one might have in reviewing just the scatterplots in our findings is: What represents the *correct* way to play *Beats*? However, we reject this binary choice and argue that instead there are multiple valid approaches to gameplay. Although Daniel is not using data to the same degree that Alyssa is, he is still engaging with the mechanics of the game in a way that make sense given the mix of conceptions that he is bringing to the game involving music, data, and games themselves. Krista, who is not fully engaged with the game, might be guided toward an alternative form of assessment or work with an instructor in order to clear up questions she might have about game mechanics. In a formative assessment, it is the job of designers and educators to not only provide culturally relevant content that students can relate to but to also develop culturally responsive instructional tools (Richards, Brown, & Forde, 2007) that are adaptable in how students can approach play to meet their diverse cultural backgrounds, play practices, and motivations. In this way we anticipate students will be more engaged than through typical assessment, and thus provide better feedback to teachers on which to build a stronger knowledge of data analysis. Through our

data visualizations, however, we can see fuller pictures of player interaction and imagine various productive avenues of teacher guidance. Thus, we offer two primary design considerations for playful assessments:

1. Understand that your players will be drawing across multiple literacies and identities. The sort of feedback and guidance that Alyssa would require is not the same as for Krista or Daniel. Structuring teacher materials and resources for playful assessment should account for those differences.
2. Provide rich visualizations for instructors to help them recognize and adequately support all types of learners. A playful assessment is unique in the level of fine-grained data that it can deliver to an instructor, as seen in our Findings section above. Playful assessments need to present this information to teachers in a way that is both quickly obtained and grasped.

References

Au, K. H., & Jordan, C. (1982). Teaching reading to Hawaiian children: Finding a culturally appropriate solution. In E. T. Trueba, G. P. Guthrie, & K. H. Au (Eds.), *Culture and the bilingual classroom: Studies in classroom ethnography* (pp. 139–152). Rowley, MA: Newbury House.

Berland, M. (2016). Making, tinkering, and computational literacy. In K. A. Peppler, E. Halverson, & Y. B. Kafai (Eds.), *Makeology* (pp. 196–205). New York, NY: Routledge.

Berland, M., Baker, R. S., & Blikstein, P. (2014). Educational data mining and learning analytics: Applications to constructionist research. *Technology, Knowledge and Learning*, 19(1), 205–220. https://doi.org/10.1007/s10758-014-9223-7

Black, P., & William, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles Policy and Practice*, 5(1), 7–73.

Boluk, S., & LeMieux, P. (2017). *Metagaming: Playing, competing, spectating, cheating, trading, making, and breaking videogames*. Minneapolis: University of Minnesota Press.

Consalvo, M. (2009). *Cheating: Gaining advantage in videogames*. Cambridge, MA: The MIT Press.

DesPortes, K., Spells, M., & DiSalvo, B. (2016). The MoveLab: Developing congruence between students' self-concepts and computing. In *Proceedings of the 47th ACM Technical Symposium on Computing Science Education – SIGCSE '16* (pp. 267–272). Memphis, TN: ACM. https://doi.org/10.1145/2839509.2844586

DiSalvo, B., & Bruckman, A. (2011). From interests to values. *Communications of the ACM*, 54(8), 27. https://doi.org/10.1145/1978542.1978552

Epic Games. (2017). *Fortnite* [Video game]. Epic Games.

Harel, I., & Papert, S. (1991). *Constructionism: Research reports and essays, 1985-1990*. Norwood, NJ: Ablex.

Holbert, N., & Wilensky, U. (2019). Designing educational video games to be objects-to-think-with. *Journal of the Learning Sciences*, 28(1), 32–72. https://doi.org/10.1080/10508406.2018.1487302

Kim, Y. J. (2018). Game-based assessment. In B. B. Frey (Ed.), *The SAGE encyclopedia of educational research, measurement, and evaluation*. Thousand Oaks, CA: Sage.

Ladson-Billings, G. (1995). Toward a theory of culturally relevant pedagogy. *American Educational Research Journal*, 32(3), 465–491.

North, A. C., Hargreaves, D. J., & O'Neill, S. A. (2000). The importance of music to adolescents. *British Journal of Educational Psychology*, 70(2), 255–272. https://doi.org/10.1348/000709900158083

Papert, S. (1980). *Mindstorms: Children, computers, and powerful ideas*. New York, NY: Basic Books.

Richards, H. V., Brown, A. F., & Forde, T. B. (2007). Addressing diversity in schools: Culturally responsive pedagogy. *Teaching Exceptional Children*, 39(3), 64–68.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18(2), 119–144. https://doi.org/10.1007/BF00117714

Tarrant, M., North, A. C., & Hargreaves, D. J. (2000). English and American adolescents' reasons for listening to music. *Psychology of Music*, 28(2), 166–173.

Weintrop, D., Holbert, N., Horn, M. S., & Wilensky, U. (2016). Computational thinking in constructionist video games. *International Journal of Game-Based Learning*, 6(1), 1–17. https://doi.org/10.4018/IJGBL.2016010101

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant numbers 1742011 and 1741956.