International Journal of Gaming and Computer-Mediated Simulations
Volume 14 • Issue 1

Playing Aloud:
Leveraging Game Commentary Culture for Playtesting

Anthony Pellicone, University of Maryland, USA*
David Weintrop, University of Maryland, USA
Diane Jass Ketelhut, University of Maryland, USA
Ekta Shokeen, University of Maryland, USA
Michel Cukier, University of Maryland, USA
Jandelyn Dawn Plane, University of Maryland, USA
Firoozeh Rahimian, Department of Defense, USA

ABSTRACT

Think-alouds are a common method of collecting design data where a player describes their play for a facilitator. Games promote a feeling of immersivity and player presence, which is in tension with traditional think-aloud methods. This work introduces a new type of think-aloud protocol intended for game-based contexts that leverages the genres of video blogging and livestreaming in game culture. This new approach, called Play Aloud testing, has participants take on the role of a game streamer by expressing their thoughts, feelings, and experiences as they play – modeled after live streaming commentary. This paper demonstrates the potential of the Play Aloud approach using playtest data from a game called HEX of the Turtle Islands. The authors highlight how Play Aloud testing generated useful data providing insight into the experience of young players in a way that was authentic to the format of digital games and consistent with youth gaming practices.

KEYWORDS

Game Culture, Game Design, Game-Based Learning, Live Streaming, Playtesting

INTRODUCTION

The goal of a game designer is to structure mechanics, narratives, and aesthetics in a way that not only is pleasurable, but also challenging and engaging to the player (Schell, 2015). One key tool in a designer's toolbelt is the think-aloud approach (Boren & Ramey, 2000; Denning, 1990). In think-alouds, users talk through their interaction with a technical system aided by the prompting of an expert facilitator (Nielsen, 1992). The facilitator attends to both verbal and non-verbal (e.g., behaviors, body language) responses to the system to gain insight into the user experience. While think-alouds can provide invaluable insight, the methodology is not without its challenges, including being cognitively

DOI: 10.4018/IJGCMS.296705
*Corresponding Author

taxing, requiring the user to be skilled at verbalizing their thoughts, and the potential for bias introduced by the setting or the presence of the facilitator. These challenges are all felt more acutely when attempting to get feedback on newly designed video games for learning where immersivity and presence within the game are an important aspect of the designed experience.

One central practice for feedback in game development is the use of playtesting as a means to gather data on player experience. Playtesting think-alouds are a widely-used playtesting approach for gaining insight into the players' experience. While the same challenges exist for think-alouds in the context of game design, there are additional challenges a game designer faces in trying to get meaningful insight into the player experience. For example, a common goal of video games is to design for immersion, where the player feels "present" in the game (Dede, 2009). However, asking a player to think-aloud while playing pulls them out of the game, lowering presence. Additionally, gameplay is an increasingly social activity, either in the form of players playing synchronously with others or by connecting via social media while they play. Thus, isolating a player for the purposes of data collection and playtest is inauthentic relative to the actual experience of gameplay. In response to these methodological challenges, this paper introduces a methodological variation on the think-aloud protocol that draws on contemporary youth gaming practices that we call the *Play-Aloud methodology*. In this paper, we first review literature related to the potential utility of leveraging game commentary culture for playtesting. We then describe our implementation of game commentary culture through the Play-Aloud methodology. Next, we present a short example study using this methodology to gain insight on the design of an in-development game, and then present findings on what this method revealed about our game design. We conclude by drawing conclusions from our implementation and findings, pointing towards how this method might be used by other game designers and researchers, and future work to refine the method.

PREVIOUS WORK

Think-Aloud Protocols for Playtesting

Think-aloud methodologies ask users to interact with a particular technology or work through a given task and are asked to voice their thought processes as they do so (Nielsen, 1992). This method of user research has a long tradition in the design of technical systems (Boren and Ramey, 2000) and has been productively used in games-user research as a means of playtesting (Knoll, 2018).

Think-aloud protocols, while widely used, have several key methodological drawbacks that can negatively impact the data collected (Boren & Ramey, 2000). First, there is the risk of the facilitator influencing the participant which can happen when the facilitator is also the creator of the technology (Boren & Ramey, 2000; Knoll, 2018). This is compounded where the participant and the facilitator have a natural power imbalance – this is often the case when designing and studying technologies for children (Nielsen, 1992). Second, the presence of an observer can change the way the players approach the challenges of a game (Firely & Engl, 2010). Third, think-alouds add to the cognitive load for the user. Games are typically designed to be challenging, and the process of engaging in a think-aloud imposes additional cognitive load (Knoll, 2018), which can make both playing and narrating the play taxing (Blumberg & Randall, 2013; Knoll, 2018; Nielsen, 1992). Lastly, think-alouds can break player immersion with the act of gameplay breaking the player sense of presence, and change the context of play through the laboratory setting (Fireley & Engl, 2010; Knoll, 2018; Louvel, 2018).

Playtesting in Game Design

Game designers use a combination of rules (called mechanics), visuals, audio, and technical systems to provide players with certain experiences (Salen & Zimmerman, 2003). Unlike productivity software, which seeks to eliminate frustration and challenge, games are often designed explicitly around providing players with satisfyingly challenging scenarios (Juul, 2013; Pagaulayan et. al 2002). For

learning games, sometimes called serious games, this process is further complicated by requiring developers to successfully integrate additional content knowledge into the game experience (Back et al., 2017; Denham, 2016; Eckhardt & Robra-Bissantz, 2018). The most widely used methodology to understand player experience is the direct observation of players as they interact with the game, called playtesting (Schell, 2015, Fullerton, 2014).

Playtesting is the act of interpreting and understanding player experience, and then incorporating that data into the evolving design of the game (Choi et al., 2016; Fullerton, 2014). Fullerton (2014) phrases the importance of playtesting as such, "If you learn to listen to your playtesters and analyze what they are saying, you will be able to see the game mechanics for what they are, not what you want them to be or imagine they should be," (p. 304).

Successful game design involves the designer understanding the full array of ways that players will use and interpret aspects of the game (Aslam & Brown, 2020). Playtesting is a vital but difficult part of the game design process (Schell, 2015). Choi et al. (2016) found that novice game designers grasped the basic idea of playtesting, but often did not understand the concept of player experience, nor how to capture this data through playtesting. In addition, playtesting is resource intensive, which makes it difficult for small independent teams to perform (Mizra-Babaei et al., 2016), especially using constant iterative playtesting as recommended by Fullerton (2014).

Live and Pre-Recorded Commentary in Game and Youth Culture

To address limitations of think-aloud methods described above, we draw on practices from game commentary culture where players comment on games during game play. Game commentary has grown as a form of entertainment in youth culture due to the rise in popularity of gaming content on streaming platforms such as Twitch and gaming communities on YouTube (Berg, 2019). For instance, most youth report regularly playing digital games, and report YouTube as a primary social platform they use daily to socialize and get information (Anderson & Jiang, 2018; McRoberts et al., 2016; Yarosh et al., 2017).

Game commentary culture tends to take two the form of one of two genres: live streaming for an audience or creating pre-recorded video content. Live streaming refers to playing a digital game for an audience, typically with commentary (Taylor, 2018). Game video blogging focuses on game commentary delivered over pre-recorded videos (Postigo, 2014). Both forms place players into the dual role of performer (provide running commentary on their gameplay) and player (interacting with either a live or assumed audience) (Hamilton, 2013; Pellicone & Ahn, 2017; Postigo, 2014; Taylor, 2018; Walker, 2014). Although we recognize that these genres are distinct, they still have many commonalities; thus, in this work, we use the term 'game commentary' to capture both genres of gameplay media.

Game commentary excels in eliciting the sort of reactions that are typically sought after in user testing - namely an individual's moment-by-moment thought processes and emotional reactions to a game (Taylor, 2018). Game commentary genres provide a pre-existing model for those reactions since this form of media has become a popular form of youth entertainment (Yarosh & Jiang, 2018; Yarosh et al., 2017). Game commentary media provides two useful affordances for think-aloud playtesting research, which has inspired our use of it in our methodology. First, game commentary provides a context for feedback that is natural to the player, resolving the tension between playing the game and talking about playing the game that comes with conventional think-aloud playtesting. Second, game commentary provides a model for players to interact with the think-aloud protocol, alleviating facilitator bias.

INTRODUCING THE PLAY-ALOUD METHODOLOGY

The central premise of the Play Aloud methodology is for the participant to take on the role of a video game streamer while playtesting the game. In this role, playing the game becomes a performance

where gameplay is accompanied by verbal commentary, feedback, and other forms of interaction (e.g., facial expression, gestures). This type of gameplay both mimics key features of think-aloud protocols, while also aligning with a common form of gaming practice. In doing so, the Play Aloud methodology leverages the game commentary cultural practice as a model of how to verbally engage with a video game while remaining immersed in the experience.

The Play Aloud method has several features that can address some of the traditional challenges of think-alouds in the context of designing video games:

- **Providing a Model for Participant Feedback:** Game commentary media as a genre is inherently structured around consistent, ongoing descriptions of a player's action in a game for the benefit for an audience (Pellicone & Ahn, 2017; Taylor, 2018). By framing think-aloud data collection as game commentary in the Play Aloud methodology, the player has an existing model to provide feedback about the game.
- **Separating the Facilitator from the Feedback:** The implied audience of game commentary can be assumed to be friendly, receptive, and peer level (Hamilton et al., 2014; Taylor, 2018), thus providing more naturalistic feedback that reduces issues of the player trying to please an adult facilitator.
- **Allowing for Immersed Responses to Game Data:** Game commentary requires players to merge their experience of a game with a running commentary to an audience (Pellicone & Ahn, 2017). The Play Aloud methodology employs game commentary genres as a method of playtesting data collection that allows us to capture more natural form of immersion by player participants.
- **Rich Multimodal Data:** Playtesting often records multiple modes of data to capture player experience (Fullerton, 2014; Schell, 2015). These data must be combined and collated across sources, which makes analysis complex (Choi et al., 2017). Since multimodal media products are common across game commentary genres (Pellicone & Ahn, 2017; Taylor, 2018), there are a large number of well supported, free technical solutions for capturing gameplay relevant data from multiple sources. Thus, the Play Aloud methodology allows researchers to tap into widely available technical and social resources to employ the method.
- **Providing Rich Data at Scale:** Playtesting is both time consuming and expensive (Schell, 2015), requires training for a facilitator to elicit useful data (Choi et al., 2017), and is typically administered in a one-on-one fashion (Nielsen, 2000). The Play Aloud methodology allows for collecting quality data at scale by removing the facilitator from the protocol and using game commentary as a frame for prompting participant responses.

Given the feature above, the Play Aloud methodology has great potential for assisting developers in gather meaningful, honest feedback from participants. In the next section we present data from a study on an in-development game that used the Play Aloud methodology.

THE PLAY ALOUD METHODOLOGY IN ACTION

In this section, we present a series of vignettes from our study highlighting how the Play Aloud methodology provided useful insights into players' experiences. The focus of this section is to show what the Play Aloud methodology looks like in action, and the types of insights that it gleaned from using it during the design process. We first provide contextual and methodological details to help frame this section. We then present findings structured around the benefits of the method identified through our second round of coding, providing vignettes from our dataset that highlight those affordances in practice for the gameplay experience of *HEX of the Turtle Islands*.

Game Context: HEX of the Turtle Islands

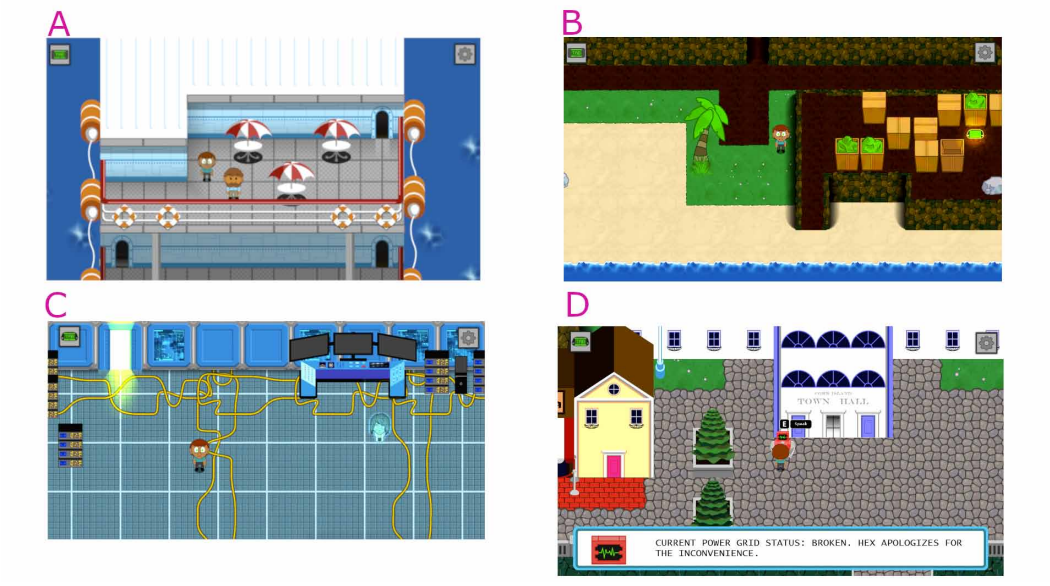
HEX of the Turtle Islands is an isometric 2-Dimensional adventure game, which places players into the role of an intern aboard a research vessel investigating recent ecological changes to a remote island chain called the Turtle Islands. The game is designed to graphically resemble *Pokemon* (Gamebrain, 1990) or The Legend of Zelda (Nintendo, 1989) in terms of movement and graphics. However gameplay draws heavily from adventure games in terms of narrative, puzzle, and scenario design. Figure 1 presents several locations of gameplay that players encounter while exploring the first island during the game commentary sessions.

The design goal of *HEX of the Turtle Islands* is to introduce players to central concepts and practices of cybersecurity and raise players' interest in the field. With this round of Play Aloud testing, we were interested in players' experiences with our puzzles (ciphers and computational logic), immersion in the game and the narrative, and self-efficacy building moments. We were also attuned to more generic playtesting goals, such as player experiences of fun and enjoyment, and identifying bugs and missteps in our game's design.

Study Context and Methods

The data collection for this study used three instruments: a parent pre-survey to collect demographic and background data about the participants, a Play Aloud gameplay session, which was the primary data collection instrument, and a brief post-game interview. These instruments were adapted from previous work in the space of cybersecurity gaming (see Coenraad et al., 2020). They were methodologically motivated to understand participants' former experiences with gaming, computer science, and cybersecurity. This paper focuses on gameplay data collected during the Play Aloud sessions while drawing on pre survey and post interview data to provide additional context. The method and gameplay prompts were refined over the course of 4 pilot studies that took place in the month prior to this session. We have included these instruments as an appendix.

Figure 1. Below are several screenshots from our game. We will refer to these by their labels in the findings section below. This represents the primary areas of the game in narrative order: the player starts on a research vessel (Label A), is shipwrecked on an island (Label B), discovers a scientist in a hidden lab (Label C) who gives direction for finding the kidnapped crew, and then arrives in a seaside town (Label D) where they must find boat passes to advance to the next island.



Our method proceeded through 3 steps. First the participant was introduced to the idea of game commentary as a model for game play, giving the example of either live-streaming on Twitch, or pre-recorded videos on YouTube. Next, we introduced the participant to the data capture instrument, allowing them to see what aspects of their play will be recorded. Finally, we hid the data capture instrument, and bring the game to full-screen and ask the player to start playing as their gameplay is recorded (see Figure 2).

To collect data during our Play Aloud session, we used the Open Broadcaster Software (OBS), a free, open-source program that supports collecting and presenting numerous data streams in parallel. Our primary data collection instrument was a 'scene' (meaning an array of inputs laid out in a graphical format) in OBS. Figure 2 presents a screenshot of an OBS video scene that we used for the data capture.

In the upper part of the screen, the game is captured as seen by the player. Below that is an input logger (courtesy of https://github.com/univrsal/input-overlay), which shows the mouse clicks and key-presses made by the player in real time. The lower right section of the screen shows the webcam capture of the player's face as they play the game (anonymized for the purposed of this paper). Captured alongside the video elements is the audio feed from the microphone of a gaming headset attached to each test laptop. All of these data streams (game video, player video and audio, and input logger) are synchronized and recorded in real time. Finally, this view of the data is only available to the researcher, and the player only sees the game (the top left portion of the screen), so their gameplay experience is not impacted by the data collection instrument during play.

Participant Selection and Data Analysis

We sought a diverse sample of players within our target demographic of youth. We advertised with several local libraries located in ethnically and economically diverse communities located near a mid-Atlantic metropolitan area in the United States. Participants ranged in age from 10 to 14. The vignettes below are drawn from a series of one-on-one Play Aloud sessions with 11 participants (2 girls and 9 boys). Participants included players of a wide array of ethnicities representative of

Figure 2. A screenshot of our instrument in OBS. Each data stream that is merged into the instrument is labeled. This is the view of the data collection instrument; however, the player only sees the 'Game Data Stream' window in full screen.

the neighborhood where the local library was located, and predominately self-reported as African American or Black, but included Asian, and Latino and Hispanic participants. We received both informed parental consent and child assent from all participants. All participant names have been anonymized with pseudonyms consistent with their stated gender. Participants were compensated for their time and effort with a $15 online gift card.

We engaged in two distinct rounds of analysis: the first was an initial round of open coding that was intended to identify and categorize major themes within player data, and to orient the analysts to both our participants and their individual experiences in gameplay (Charmaz, 2014). The second round of coding resulted in the data presented below, and focused on identifying the affordances of the method, and are represented by the headings for each sub-section.

Benefits of the Play Aloud Methodology

In this section, we present a series of vignettes from our study highlighting how the Play Aloud methodology provided useful insights into players' experiences. We have structured these findings around the benefits of the method identified above, providing vignettes from our dataset that highlight those affordances in practice for the gameplay experience of *HEX of the Turtle Islands*.

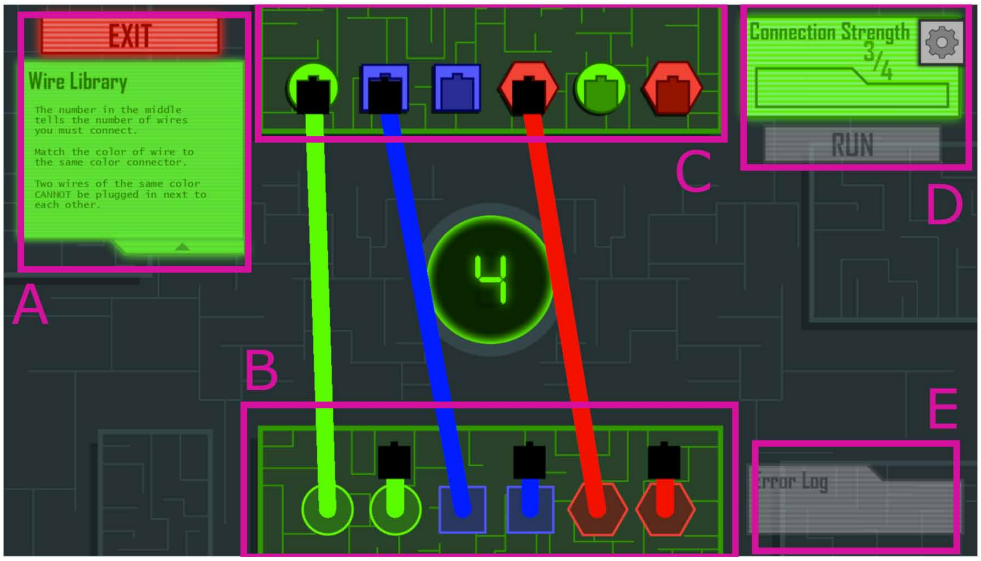
Providing a Model for Participant Feedback

Across our 11 participants, 8 players responded consistently with running commentaries of their gameplay, 1 participant was less consistent in verbal commentary whereas 2 players were not responsive in verbal commentary. These differences did not correspond to any larger pattern in our participant demographics. The majority of our players gave their own version of running commentary as they played.

A clear example of consistent running commentary comes from Aidan. As Aidan approaches the town (Figure 1, Label D), he is required to solve a wire puzzle (Figure 3). Aidan has a brief moment of frustration with the first wire puzzle and requires a facilitator's help to understand a core aspect of the wire puzzles – the concept of connection strength (Figure 3, Label D). Given this hint, Aidan is able to solve the puzzle and advance. Aidan proceeds through a brief maze section, commenting "*Guys, this game is really good,*" solves the second wire puzzle with little difficulty, and emerges in town. This section of the game is bookended by a text prompt, which Aidan answers in detail, talking about where he thinks the plot will go, what he has done so far, and for a suggestion how he wishes the wire puzzles were more challenging. After he closes the text prompt and continues moving, he has a second thought, saying that he wishes that the player could either move in cutscenes, or skip them altogether. He moves to a pier which is bordered by a fence, and quickly switches to an in-character response, saying, *"I'm about to find out who's behind all of this… oh, wait, can I jump over this (the railing)? [this is accompanied by hitting the jump button repeatedly]*" He then moves up to a robotic NPC (See Figure 1, Label D), which gives a clue about the role of HEX as the game's antagonist, and he excitedly says, "*Oh! Oh! Wait! This is from HEX! The corporation! HEX Corp, isn't that the person we're trying to figure out?*" Aidan then proceeds a bit further, running into the third wire puzzle, and solves it on his first attempt, saying "*I love the mechanics of the puzzle system, it's challenging, but they… they need to change it around,*" – an idea we were able to get Aidan to expand upon in his interview as wanting a wider variety of puzzles.

Across this snippet of data, Aidan has provided a continuous stream of data, from his utterances, to his body language, and to the keys he presses as he navigates the world. Each of his thoughts leads into the other and provides a running commentary that describes his experience with the game. In terms of design data, this is extremely useful and allows us to see the way that a player responds to our game in real time. An example of the implied third party viewer comes from Aidan, who (after he successfully completed a wire puzzle and celebrated) said, "*Make sure to subscribe to my videos,*" mirroring a common refrain in game commentary media that seeks to build an audience (Pellicone & Ahn, 2017; Postigo, 2014), and an example of a common practice in youth video production

Figure 3. This figure displays the wire puzzle, which is referenced in Aidan's vignette. Label A is the wire library, which lists the logical rules that guide the puzzle, Label B are the wires which are connected to corresponding ports (Label C), Label D is the connection strength which shows the progress towards the goal of connecting a certain number of wires, and Label E is the error log which displays any incorrect moves the player has made.



(McRoberts et al., 2016). The model of an implied game commentary audience has prompted Aidan into constantly thinking about the game and providing verbal comments.

Separating Facilitator From Feedback

Although most players stated that they enjoyed the game, the Play Aloud methodology allowed us to collect evidence showing small, crucial moments of elements that players did not enjoy, or found to be lacking. An example comes from Bethany, who although positive towards the game (indicating that she enjoyed it in her interview), still had several critical comments as she played. As Bethany jumps off the crashed boat, she sees the robotic dog for the first time, commenting, "*That dog looks stupid. I don't like it*." Similarly, she makes a sarcastic comment about a narrative implausibility (the player must get boat passes in the town on the island, and conveniently the shopkeeper there is offering two boat passes in exchange for fixing broken machinery) rolling her eyes, and saying "*The fact that he already knew that is* ***actually*** *crazy*," and after fixing the broken machinery jokes, *"Ok, I fixed your stuff now, you old,* ***old*** *man.*" The sort of feedback that Bethany provides is both candid, in the sense that she employs sarcasm and humor in describing her reaction to this character and goes so far as to directly critique the aesthetics of the robot dog. We argue that this sort of feedback would be far less likely with a facilitator present, especially when the facilitator is known to be the designer of the game and is afforded by the assumed audience of the Play Aloud methodology.

Allowing for Immersed Responses to Gameplay

*HEX of the Turtle Island*s seeks to immerse players in both gameplay and narrative of the game. In our analysis, we operationalized immersivity as players speaking and behaving as if they were in the games themselves, using a first-person language when reacting to in-game events. We use Eliza's playtesting data below as an example of that.

Eliza both narrated her thought process and responded in-character to NPCs frequently during her play session. One of the first goals in the game is to find research notes that have been locked in the ship's hold with a passcode, which necessitates talking to an NPC named Baz (Figure 1, label A) to retrieve the lock's code. Eliza explores this first section of the game, making comments to herself as she does. For example, there is an interactive element that shows the crew members of the expedition. Eliza notices it immediately, saying, "*Oh! What's this…*" and interacts with the bulletin board, studies it for a few seconds, repeats the characters' names and says "*Ok!"*. She moves forward, triggers the dialogue with the second lead scientist in this area, which gives several hints: confirming that the player should be looking for the research notes, that the technician (named Baz) has the passcode, and that the waters the ship is currently traveling in are plagued by pirates - hence the security measures to get into the hold. Eliza again expresses momentary confusion, "*Hmmm, where would I find him… what was his name again?*" She goes back to the bulletin board with the staff, interacts with it, sees the crew list again, and confidently says, "*Oh - OK. Baz!*" she pauses, smiles, and says, "*My goodness, this is the best game ever!*" She finds the exit to the lower deck where Baz is located, sees Baz's sprite (which she recognizes from the interactive element), and says, "*Oh! There he is! Yessss!*" Baz opens his introduction in the middle of the action, with the assumption that the player has been on the ship for several days as part of their work study program with dialogue that reads, "*Hi there! I'm Baz. I hope your first week on the ship has gone well.*" Eliza responds to this prompt by saying, "*It has!*" out loud. She then sees the keypad next to the locked door to the hold, and excitedly says, "*There it is!*" repeating the access code to herself again. She opens the door successfully, and vocally celebrates, saying "*Yay!*" as she enters the hold and sees the research data on the table. The action of picking up the notes triggers a cut-scene in the game, where the crew members are captured, and the player can overhear the pirates and crew fighting. This is prefaced by a shaking screen and flashing lights. Eliza sees this, raises her hands from the keyboard, and gasps audibly saying, "*Oh my God! Noooo, what's happening!?*" As she advances through this scene, she holds her hand to her mouth and says, "*Oh my god, this is so scary!*" laughing a bit as well. The cutscene ends with the boat running aground (Figure 1, label B), and an exit opening up through a broken vent. Eliza sees the exit, and says to the NPC friend characters, "*C'mon peeps, let's go!*"

In this example, we can see Eliza making use of the interactive elements in the game (e.g. the bulletin board), responding in character to Baz and the friend characters as well as to the drama of the events on the boat (e.g., laughing at characters, urging them to action, acting surprised at the kidnapping scene), and narrating her thought processes in the first puzzle sequence (e.g. thinking aloud about her goals as she progresses). This allows us to understand key moments that work in terms of triggering presence, and strive to recreate those experiences as we continue the game design and development. Eliza's reactions come unprompted from a facilitator and reflects her moment by moment responses to the game. In terms of immersion, we can see Eliza focusing on different elements of the game, commenting on them, with these comments often merging into one another, which is useful for examining the way that Eliza responded to story elements as they were situated within gameplay.

Rich Multi-Modal Data

Game commentary is a multimodal genre of media by design, incorporating elements of the game, the streamer, and the technology of the game and platform (Hamilton et al., 2014; Pellicone & Ahn, 2016;). This element of Play Aloud method is appealing for assessing the necessarily multi-modal data associated with player experience in a game. In our method, we can see player facial expressions as they play (e.g. smiling or frowning), their body language (e.g. leaning towards or looking away and being distracted), their inputs during play (e.g. mouse clicks and button presses), and all gameplay events (e.g. strategies to approach puzzles). This affordance was particularly useful for the participants who were not consistent in their verbal commentary. For example, one player, named Nicolas, who was not consistently verbal, but also experienced a great deal of frustration with the game, responded

frequently with non-verbal cues demonstrating his frustration. An example comes from when Nicolas approaches a pigpen cipher puzzle in the game. After finding the device that acts as a flashlight, Nicolas enters a cave, and sees a Q prompt, suggesting that he use the flashlight. He tries to interact with it like an object in the gameworld instead of connecting it with the 'Q' key on the keyboard, which was visible through our input module in the OBS scene. He struggles with this a bit and says, *"...What!?*", trying several combinations, leaning intently forward as he attempts to figure out the controls. Finally, he interacts with the friend NPC positioned in the cave, who provides a hint about the code being nearby, and he says, "*Bruh… you could've said that such a long time ago…*" He looks around for help at this point, but both facilitators are interviewing other players. He then walks around fruitlessly on the lower part of the beach, saying "*There's no code!*" in a frustrated voice, he goes back to the box puzzle area and says, "*Where's the code!?*".

This piece of data points towards the utility of having multi-modal data in a single unit of analysis. Throughout this series of interactions, Nicolas is vocalizing the source of his confusion (e.g. wondering where the code is), encountering difficulty in trying to interact with game elements (e.g. pressing various buttons to find the flashlight), and demonstrating concentration and frustration through non-verbal cues (e.g. leaning towards the computer and sighing heavily at times). Interactions such as these helped us identify several frustrating aspects of the game (e.g. needing clearer prompts, and better explicating puzzle instructions), even though Nicolas' responses were not as consistent as Aidan's. This indicates that the Play Aloud method, even for participants who are not engaging as fully with the game commentary model, still provide the sort of data that is useful in playtesting. Furthermore, we can connect Nicolas' physical actions, verbalizations, and activities in game easily, allowing us to understand how he is reacting to the designed gameplay experience by using the combined data sources of the OBS instrumentation.

Providing Rich Data at Scale

For this implementation, we were able to collect data from 11 participants with only two facilitators over the course of 2 90-minute sessions. In comparison to a traditional one-on-one implementation, which would take approximately 60 minutes per player, by employing the Play Aloud methodology, we were able to run multiple playtesting sessions in parallel. This allowed us to collect 11 playtesting sessions in a total of 3 hours, which is significantly less than the 11 hours it would have take for a one-on-one synchronized playtesting think-aloud approach.

DISCUSSION

The affordances of the Play Aloud methodology described above present several benefits for playtesting:

- **Reducing Facilitator Bias:** The Play Aloud methodology creates conditions where players can give candid feedback without introducing potential bias from a facilitator. We saw this both in the candid ways the players such as Bethany responded to elements of the game, as well as in Aidan's framing of his commentary specifically within the context of speaking towards an assumed audience. The method was successful at provoking constant ongoing player gameplay data, which blended fluidly between reacting to narrative, gameplay, and cognitive elements of the game as shown in our players' responses to the conceptual elements of puzzles. For players who did not respond as readily to the verbal aspects of the method, their data still provided information about their physical response to the game, gameplay inputs, and gameplay capture. This is demonstrated in Nicolas's data – although he did not give consistent verbal feedback, our method still provides useful data for analyzing playability and cognitive processes in puzzles.

- **Better Capturing Presence:** A main conflict of traditional one-on-one pairings for facilitators and players is the players' immersed experiences in games versus having to provide consistent feedback as one plays (Nielsen, 2002). Through Eliza, we can see the value of consistent, running feedback in evaluating the immersive properties of gameplay and that our narrative is effective at getting her to respond while playing the game.
- **Scaling Facilitators to Participants:** The traditional format for a think-aloud interview is to pair one facilitator to one participant (Nielsen, 2002). The Play Aloud methodology, by transferring the prompting to the cultural model of game commentary, allows for limited numbers of facilitators to collect detailed data from a large number of participants. We had one experienced facilitator who had done think-aloud style interviewing and one novice facilitator that were able to collect rich gameplay data from 11 participants in about 3 hours over the course of two days. For academic and independent teams to conduct rigorous one-on-one interviews, who often do not have access to either the staff or the expertise (Mizra-Babei et al., 2016), our method presents an effective alternative.

Limitations and Next Steps

While this method worked well for our team's present context, which is the alpha phase of a game in development with a small academic team, we recognize that the Play Aloud methodology is only one step in the larger playtesting process. Ideally playtesting is an iterative process (Fullerton, 2014), comprised of individual playtests tailored to the questions that a team is hoping to answer for that phase of development. Our questions for this round of testing were open, and related largely to initial player responses to mechanics and narrative - therefore the Play Aloud method worked for our goals. Therefore, the Play Aloud method would need to be tailored to the specific goals a design team has for their project.

For *HEX of the Turtle Islands*, we have employed this method in subsequent rounds of playtesting. This method was developed prior to the global COVID19 pandemic, but one benefit that we have found in testing is that this methodology moves to digital data collection very effectively. In future work, we plan to analyze how the method changes in online testing. However, our preliminary findings using this method are promising. We recognize that a full assessment of the method would come from a comparative analysis with the traditional one-on-one think-aloud method. Furthermore, we will conduct research and collect data to gain insight into how this method will interact with existing participant perceptions of game culture – e.g. participants who are unfamiliar with or indifferent towards game commentary practices.

Another limitation for this methodology is its dependance on participants familiarity with contemporary gaming culture and live-streaming and video blogging practices. While these are increasingly well-known among youth, not all potential study participants may be familiar with it. This is potentially an issue if part of the goal of a game-based learning environment is to recruit youth who are not interested in games or participate less often in game culture. Likewise, trying to bring this methodology into contexts at odds with gaming culture, such as a classroom, may would also introduce a tensions between gaming norms and classroom culture. A similar concern is that game commentary tends to encourage the performer to have exaggerated reactions to fit with their perceived audience (Pellicone & Ahn, 2017; Postigo, 2014; Taylor, 2018).

The concept of cognitive load is commonly recognized as a drawback to traditional think-aloud methodologies (Knoll, 2018). A valid critique of the Play Aloud methodology is that streaming has its own associated cognitive load (Pellicone & Ahn, 2017; Taylor, 2018). We recognize this, but point towards our data, which shows that for some players this mode of commentary came easily, thus indicating that (as with any method) a designer or researcher should keep in mind their target audience. Using the Play Aloud methodology will assist some participants in giving richer data, but may hinder others.

Finally, Play Aloud methodology has great utility for the game development process, but that playtesting is meant to be iterative, and adaptive (Fullerton, 2014). The Play Aloud methodology is therefore seen as a complement to more traditional one-on-one think-aloud protocols. We can see Play Aloud methodology as more effective in early stages of game development to get rich, multimodal data, which is then iterated upon and further investigated through targeted one-on-one playtesting.

CONCLUSION

Games simultaneously act as cultural objects, and as the locus for new media cultures that form around play and socialization (Steinkuehler, 2006). Therefore, we present our method as a way to mesh both the game as a designed object, and the existence of games and gaming communities as distinct cultures unto themselves. Game culture is rich and multi-faceted, and the way an individual plays and enjoys games is deeply tied to the process of game development (Dovey and Kennedy, 2006). The cultural forces of gameplay exist within a feedback loop that determines the types of games that are designed, and the way that those games are perceived by their players (Kirkpatrick, 2014). Game commentary has grown in popularity in the last decade and significantly changed what it looks like to participate in game culture, reshaping the way that both designers and players conceptualize gameplay (Taylor, 2018). In this paper, we presented one implication of this rapid change – the utility of game commentary in understanding player experience. Player experience is central to gameplay goals in developing skills (Ketelhut, 2006), inculcating learning (Gee, 2008), fostering mastery (Ketelhut, 2006), and building identities (Squire, 2006). The Play Aloud methodology gives designers the ability to capture this experience directly, using a form that is natural to many young players, in a way that allows for unfiltered reactions to gameplay, and produces rich quality game test data with limited outlays of resources and time. With this work, we introduced the Play Aloud methodology as a tool to be added to the methodological toolbelt of game designers. In doing so, we provide a way for game designers to draw from emerging trends in gaming culture as a means to generate new insights into how players experience their designed worlds.

FUNDING AND ACKNOWLEDGMENT

We acknowledge the funding support of the Department of Defense. The views and conclusions expressed in this paper are those of the authors and do not necessarily represent those of the Department of Defense. Most of all, we like to thank students who participated in this study for sharing their experiences with us.

REFERENCES

Anderson, M., & Jiang, J. (2018). *Teens, social media & technology 2018*. Pew Research Center. https://www.pewresearch.org/internet/2018/05/31/teens-social-media-technology-2018/

Aslam, H., & Brown, J. (2020). *Affordance theory in game design: A Guide toward understanding players*. Morgan & Claypool Publish.

Back, J., Segura, E. M., & Waern, A. (2017). Designing for Transformative Play. *ACM Transactions on Computer-Human Interaction*, *24*(3), 1–28. doi:10.1145/3057921

Berg, M. (2019). The highest-paid YouTube stars of 2019: The kids are killing it. *Forbes*. https://www.forbes.com/sites/maddieberg/2019/12/18/the-highest-paid-youtube-stars-of-2019-the-kids-are-killing-it/

Blumberg, F. C., & Randall, J. D. (2013). What do children and adolescents say they do during video game play? *Journal of Applied Developmental Psychology*, *34*(2), 82–88. doi:10.1016/j.appdev.2012.11.004

Boren, T., & Ramey, J. (2000). Thinking aloud: Reconciling theory and practice. *IEEE Transactions on Professional Communication*, *43*(3), 261–278. doi:10.1109/47.867942

Charmaz, K. (2014). *Constructing grounded theory* (2nd ed.). Sage.

Coenraad, M. Pellicone, A. Ketelhut, DJ. Cukier, M. Plane, J. Weintrop, D. (2020). Experiencing cybersecurity one game at a time: A systematic review of cybersecurity digital games. *Simulation and Gaming, 51*(5), 586 – 611)

Denning, S., Hoiem, D., Simpson, M., & Sullivan, K. (1990). The Value of Thinking-Aloud Protocols in Industry: A Case Study at Microsoft Corporation. *Proceedings of the Human Factors Society Annual Meeting*, *34*(17), 1285–1289. doi:10.1177/154193129003401723

Dovey, J., & Kennedy, H. W. (2006). *Game cultures: Computer games as new media*. Open University Press.

Eckardt, L., Grogorick, S., & Robra-Bizzantz, S. (2018). Play to learn: Conducting a playtest session for improving an educational game. *Proceedings of Twenty-Fourth Americas Conference on Information Systems*, 1–10.

Fierley, R., & Engl, S. (2010, September 1). User Experience Methods and Games: Lessons Learned. *Proceedings of HCI 2010*. doi:10.14236/ewic/HCI2010.26

Fullerton, T. (2014). *Game design workshop: A playcentric approach to creating innovative games*. Academic Press.

Game Freak. (1998). *Pokemon Red and Blue*. Nintendo of America.

Gee, J. (2008). Deep learning properties of good digital games: How far can they go? In U. Ritterfeld, M. J. Cody, & P. Vorderer (Eds.), *Serious games: Mechanisms and effects* (pp. 67–82). Routledge.

Hamilton, W. A., Garretson, O., & Kerne, A. (2014). Streaming on twitch: Fostering participatory communities of play within live mixed media. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1315–1324. doi:10.1145/2556288.2557048

Juul, J. (2013). *The art of failure: An essay on the pain of playing video games*. MIT Press.

Knoll, T. (2018). The think-aloud protocol. In A. Drachen, P. Mirza-Babaei, & L. Nacke (Eds.), *Games User Research* (Vol. 1, pp. 189–202). Oxford University Press. doi:10.1093/oso/9780198794844.001.0001

Louvel, G. (2018). "Play as if you were at home": Dealing with biases and test validity. In A. Drachen, P. Mirza-Babaei, & L. Nacke (Eds.), *Games User Research* (Vol. 1, pp. 189–202). Oxford University Press. doi:10.1093/oso/9780198794844.001.0001

McRoberts, S., Bonsignore, E., Peyton, T., & Yarosh, S. (2016). Do It for the Viewers!: Audience Engagement Behaviors of Young YouTubers. *Proceedings of the 15th International Conference on Interaction Design and Children - IDC '16*, 334–343. doi:10.1145/2930674.2930676

Mirza-Babaei, P., Moosajee, N., & Drenikow, B. (2016). Playtesting for indie studios. *Proceedings of the 20th International Academic Mindtrek Conference on - AcademicMindtrek '16*, 366–374. doi:10.1145/2994310.2994364

Nielsen, J. (1992). The usability engineering life cycle. *Computer*, *25*(3), 12–22. doi:10.1109/2.121503

Nintendo. (1986). *The Legend of Zelda*. Nintendo of America.

Pagulayan, R., Keeker, K., Wixon, D., Romero, R., & Fuller, T. (2008). User-centered design in games. In A. Sears & J. A. Jacko (Eds.), *The human-computer interaction handbook: Fundamentals, evolving technologies, and emerging applications* (2nd ed.). Lawrence Erlbaum Assoc.

Pellicone, A. J., & Ahn, J. (2017). The game of performing play: Understanding streaming as cultural production. *CHI '17 Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems,* 4863–4874. doi:10.1145/3025453.3025854

Postigo, H. (2014). The socio-technical architecture of digital labor: Converting play into YouTube money. *New Media & Society*. Advance online publication. doi:10.1177/1461444814541527

Quantcast. (2020). *Twitch.tv traffic and demographics*. https://www.quantcast.com/twitch.tv#demographicsCard

Reed, J., & Acosta-Rubio, J. (2017). *Innovation through inclusion: The multicultural cybersecurity workforce*. https://www.isc2.org/-/media/Files/Research/Innovation-Through-Inclusion-Report.ashx

Salen, K., & Zimmerman, E. (2003). *Rules of play: Game design fundamentals*. The MIT Press.

Schell, J. (2015). *The art of game design: A Book of lenses* (2nd ed.). CRC Press.

Squire, K. (2006). From Content to Context: Videogames as Designed Experience. *Educational Researcher*, *35*(8), 19–29. doi:10.3102/0013189X035008019

Steinkuehler, C. A. (2006). Why Game (Culture) Studies Now? *Games and Culture*, *1*(1), 97–102. doi:10.1177/1555412005281911

Taylor, T. L. (2018). *Watch me play: Twitch and the rise of game live streaming*. Princeton University Press. doi:10.2307/j.ctvc77jqw

Walker, A. (2014). Watching us play: Postures and platforms of live streaming. *Surveillance & Society*, *12*(3), 437–442. doi:10.24908/ss.v12i3.5303

Yarosh, S., Bonsignore, E., McRoberts, S., & Peyton, T. (2016). YouthTube: Youth Video Authorship on YouTube and Vine. *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing - CSCW '16*, 1421–1435. doi:10.1145/2818048.2819961

APPENDIX: DATA COLLECTION INSTRUMENTS

Parent Survey

Child's Name:
What is your child's age?
What is your child's grade?
What school does your child currently attend?
What is your child's gender?
Is your child of Hispanic, Latino, or Spanish origin?
How would you describe your child (choose all that apply)?
- [] American Indian or Alaska Native
- [] Asian or Pacific Islander
- [] Black or African American
- [] Native Hawaiian or Other Pacific Islander
- [] White

What languages does your child speak?
What languages do you speak at home with your child? (Please list based on how often each language is spoken)
What is your child currently doing in math?
Does your child usually struggle to finish their math homework?
Has your child done any computer programming classes in school? If so, what has he / she done?
Has your child done any computer programming during camps or after-school activities? If so, what has he / she done?
What is the highest level of school that your child's mother has completed?
- [] Some high school
- [] High school graduate
- [] Some college
- [] Associate degree in college (2-year)
- [] Bachelor's degree in college (4-year)
- [] Master's degree
- [] Doctoral degree
- [] Professional degree (JD, MD)

What is the highest level of school that your child's father has completed?
- [] Some high school
- [] High school graduate
- [] Some college
- [] Associate degree in college (2-year)
- [] Bachelor's degree in college (4-year)
- [] Master's degree
- [] Doctoral degree
- [] Professional degree (JD, MD)

Interview Protocol

1. Do you like to play games? If so, then what sort of games do you typically play in a week?
2. How would describe this game to a friend?
3. This game was designed for you to learn something. What do you think the game was designed for you to learn?
4. Have you ever learned computer science before or done programs like Code.org or Scratch?

5. What is your favorite subject?
6. Do you like playing video games?
7. Do you like doing puzzles?

Closing Question

Would you be interested in coming back to play more of the game in the future, or to give us feedback on designing the game?

Anthony Pellicone is a post-doctoral researcher at the University of Maryland. His work considers the ways that people learn, socialize, and connect through games and game-related spaces, and how to best design those experiences.

David Weintrop is an Assistant Professor in the Department of Teaching & Learning, Policy & Leadership in the College of Education with a joint appointment in the College of Information Studies at the University of Maryland. His research focuses on the design, implementation, and evaluation of effective, engaging, and equitable computational learning experiences.

Diane Jass Ketelhut is Professor of Science Education at the College of Education, and Director of the Center for Science and Technology in Education. Her research interests center on improving student learning and engagement with science through increasing access, particularly in urban contexts, to scientific inquiry and computational thinking (CT) experiences and through raising self-efficacy.

Ekta Shokeen is a PhD Student in the College of Information Studies at the University of Maryland. Her research focuses on the intersections of learning, technology, and embodied cognition. Her current research focuses on the design of educational videogames from the perspective of embodied cognition.

Michel Cukier is a professor of reliability engineering at the University of Maryland, College Park. He is also the director for the Advanced Cybersecurity Experience for Students (ACES). His research covers dependability and security issues. He has published more than 90 papers in journals and refereed conference proceedings in those areas.

Jan Plane is a principal lecturer in the computer science department at the University of Maryland, College Park. She is the director of the Iribe Initiative for Inclusion (I4C) and Diversity in Computing and the Maryland Center for Women in Computing (MCWIC). She is also the associate director for the Advanced Cybersecurity Experience for Students (ACES). Her research covers computer science education and inclusivity in computing fields.

Firoozeh Rahimian is a researcher in computer science with emphasis on cybersecurity. She has worked as an IT professional for 30+ years as a project manager and software developer for marketing and financial systems in several fields.