Playtesting with a Purpose

Judeth Oden Choi[1], Jodi Forlizzi[1], Michael Christel[2], Rachel Moeller[1], MacKenzie Bates[3], Jessica Hammer[1 2]

[1]HCI Institute, Carnegie Mellon University, Pittsburgh, USA
[2]ETC, Carnegie Mellon University, Pittsburgh, USA
[3]Sledgehammer Games, Foster City, USA
{jochoi, forlizzi, christel, hammerj}@cs.cmu.edu, rmoelle@andrew.cmu.edu, itbmac@yahoo.com

ABSTRACT

Playtesting, or using play to guide game design, gives designers feedback about whether their game is meeting their goals and the player's expectations. We report a case study of designing, deploying, and iterating on a series of playtesting workshops for novice game designers. We identify common missteps made by novice designers and address these missteps through the concept of purposefulness, understanding why you are playtesting as well as how to playtest. We ground our workshops in the development of rich player experience goals, which inform playtest design, data collection and iteration. We show that by applying methods taught in our workshops, novice game designers leveraged playtest methods and tools, employed playtesting and data collection methods appropriate for their goals, and effectively applied playtest data in iterative design. We conclude with lessons learned and next steps in our research on playtesting.

Author Keywords

Game design; playtesting; game design education; game user research.

ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous; I.2.1 Applications and Expert Systems (H.4, J); K.3.2 Computer and Information Science Education.

INTRODUCTION

Involving users in the process of game design and development is important, especially with the increasing complexity of digital games. Often this is done in the form of usability testing, which ensures that different disciplines develop a shared vision of the game, and that the features of the game are easily understood by users.

Game User Research, a growing community of user experience researchers, game developers and academic researchers, have developed a number of tools and methods to address the unique challenge of testing games. These methods help game developers better understand the player experience.

However, game user testing methods are not as widely applied or understood in industry as one would hope. A recent paper by Washburn et al. out of Microsoft Research, reviewed 155 postmortems shared on gamasutra.com [17] and discovered that most often when testing was listed under "what went wrong" developers cited a lack of testing. Our own review of the 72 postmortems shared on gamasutra.com since 2010 revealed that 28% of developers expressed some desire to have conducted more testing. Perhaps more telling is one of Washburn et al.'s key takeaways: "For a better development process, game developers should invest time in the beginning of the project planning and designing. Game developers should also build prototypes during development, and if possible continue building off of these prototypes using an iterative process" [17].

We observed first-hand in our work that student teams struggled with integrating playtesting into the iterative design process. Therefore, we set out to identify the biggest playtesting challenges teams faced, to develop a curriculum that helped teams meet these challenges, and to understand the curriculum's impact.

To do so, we assembled an interdisciplinary team of game designers, computer scientists, designers, and educators. We first conducted primary research to discover and assess what existing game design and development curricula and resources exist for game design teams who are not affiliated with a particular company. Informed by these resources, we conducted a literature review and interviews with game designers. We then used our findings to develop a series of playtesting workshops: Explore, Refine, and Prove. We ran these workshops in the context of a graduate program focused on game design and development twice during the 2014-2015 academic year, evaluating the playtesting outcomes, interviewing faculty and student stakeholders, and debriefing within our group. In observing students' progress with playtesting, we identified common missteps, such as trouble with applying data collected from a playtest to advance game design. The common theme of these missteps was a lack of *purposefulness* in playtesting.

CHI PLAY '16, October 16-19, 2016, Austin, TX, USA
 ISBN 978-1-4503-4456-2/16/10…$15.00
DOI: http://dx.doi.org/10.1145/2967934.2968103

Therefore, we created a second iteration of the workshops, focusing on "playtesting with a purpose." We emphasized 1) setting player experience goals, 2) making and testing hypotheses about how design decisions support those goals, and 3) using playtesting data in a persuasive way.

In this paper, we report a case study of designing, deploying, and iterating on these playtesting workshops. We show how purposeful playtesting, designing playtests to address a designer's specific set of questions, positively impacts an iterative game design process, and that novice game designers struggle in this area. We provide evidence that with support for purposeful playtesting, novices can leverage playtest methods and tools, selecting and executing testing and data collection methods appropriate for their goals, and effectively applying playtest data in iterative design. We conclude with lessons learned and next steps in our research on playtesting.

RELATED WORK

Games are experiences created by rules [3]. These rules set the parameters within which players make choices and perform actions, referred to as game mechanics [3]. Art, narrative, and system design create the context for the rules and mechanics, allowing the player to make meaning from the gameplay. Taken together, all the elements of a game holistically form a system that shapes the player's experience [14].

Player engagement with the rule system cannot always be predicted from the set of rules itself; there must be freedom for players to make choices, mess around, and, unsurprisingly, play [27]. Even in the most cohesive system, the player sometimes behaves in unexpected ways and finds unintended meaning in the game. These unexpected interactions between the player and the game system are a form of emergent play [4].

Game design employs an ecology of approaches, such as design, storytelling, performance, psychology, behavioral economics, and computer science to create this holistic experience [14, 28, 29]. It is inherently multidisciplinary and often requires the coordination of diverse practices and processes. Balancing the interdependencies within the game system and across the design process is challenging, and an iterative design process is commonly used [14]. In an iterative design process, the prototype is tested, analyzed, refined and then the cycle repeats [31]. Effective user testing is employed to engage both with elements that games share with other types of interactive software systems, such as art or usability, and also with game rules and mechanics [15].

Game User Research

Game User Research (GUR) is a growing community bringing together user experience researchers, game developers, and academic researchers to help improve the player experience [9]. GUR attempts to understand and measure how the play experience meets the designer's goals and the player's expectations [9].

Researchers have developed a number of heuristics to understand the player experience, such as flow [8], presence [22], immersion [16], engagement [5], enjoyment [30] and challenge and enjoyment [1]. Through the study of the player experience, GUR has developed and applied a number of quantitative and qualitative methods, instruments and tools to the study of games. For example, think-aloud protocols applied from cognitive science are used to understand the mental process of players [20]. The RITE (Rapid Iterative Testing and Evaluation) method was developed by Microsoft to fold usability testing into their iterative development process [23]. PLAY is a set of heuristics adapted from usability testing for the play experience [10]. Playtesting, the study of hands-on play [24], has been evaluated through physiological measures [2], and with a number of validated questionnaires based off of sets of heuristics, such as immersion, engagement and challenge [18].

GUR methods have been widely adopted by the gaming industry. Microsoft developed the RITE method during the development of *Halo I*. In RITE when a player identifies a problem, the developers address the problem immediately before testing the game further. Large game companies such as Valve incorporate observation, biometric data, and game metrics to draw conclusions about players [2]. Riot Games is famous for conducting A/B testing in *League of Legends* to understand what factors reduce negative player behavior [6]. Heat maps are used to visualize where and how often player deaths occur in first-person shooters [11] in order to identify excessively or unintentionally challenging areas of the game. These examples demonstrate not only the utility of GUR to the game industry, but also the variety of ways it can be applied.

Playtesting as a Method

Game development includes the process from a designer's first sketch through the marketing and release of a digital game. The game development process can be divided into four phases: concept, pre-production, production, and quality assurance [13].

Playtesting evolved from product research methods at firms such as Disney and LucasArts and, because of this, is a method often used late in the development process, as the game nears release [9]. However, game educators and game textbooks suggest that playtesting should be part of an iterative design process starting from the beginning of the design process. In her game design text, Fullerton says that, "Playtesting is something that the designer performs throughout the entire design process to gain an insight into whether or not the game is achieving your player experience goals" [12]. Schell also advocates for early playtesting, "The whole point of playtesting is to make clear to you that some of the decisions you were completely comfortable with are completely wrong. You need to find these things out as soon as possible, while there is still time to do something about them" [29].

Schell differentiates playtesting from other kinds of testing, such as usability testing or quality assurance testing, as "the kind of testing designers care about most" [29]. Playtesting guides the design of a playful experience by generating detailed feedback to the development team about if and how the game fulfills the player experience goal. In game design textbooks, "playtesting" seems to be the term of art. Although it was developed from the product research context, game design educators understand playtesting more broadly as a tool to gain insight on the player experience.

Teaching Playtesting

In teaching playtesting, we incorporated GUR methods and the insights of game educators about playtesting as an end-to-end process. We therefore researched both GUR curricula and playtesting curricula. We discovered that game design textbooks provide high-level descriptions of playtesting, but they do not detail the process. At best, the entire process is summarized in a chapter [12]. Courses are offered at the college and graduate level on game research methods, but those materials are not publicly available. Individual lectures that are public, such as those by Lewis-Evans or Ambinder [19, 2], provide little pedagogical support. Recently Pulsipher has published a lecture-based course on playtesting at udemy.com [26], but it provides no hands-on activities or feedback.

In response to these findings, we set out to develop a playtesting curriculum for game designers that was detailed, pedagogically appropriate for designers, hands-on, and publicly available. We provided a detailed focus on methods, from designing and running a playtest, to collecting data and applying data to design. We drew materials from game design texts and GUR methods. Within our workshops we provided hands-on opportunities to practice skills and apply methods taught to student game designers.

To develop the workshops, we engaged in an iterative design process of our own. Spanning three iterations over three years, we have developed insights into effective playtesting teaching and the common missteps of novice game designers.

CONTEXT OF THE WORKSHOPS

The Entertainment Technology Center (ETC) is a two-year professional graduate program that focuses on game development, interactive entertainment research, design, and production. In the spring of year one and in each semester of year two, students complete a semester-long interdisciplinary design project focused on designing and developing a digital game or interactive experience. Teams typically are comprised of 5-8 members with diverse backgrounds in art, programming, game design, writing, and producing.

Each project team partners with a real-world client to produce a working digital prototype of an original game or experience. Clients range from local schools and hospitals, to energy companies and AAA video game design studios. Similarly, the types of experiences teams produce vary from games for entertainment, to interactive museum installations, to serious and educational games.

At the beginning of each semester, teams receive a project brief from the client. Project briefs vary in specificity. It is typical for teams to meet with their clients weekly or every two weeks throughout the semester to ensure that they are meeting the client's expectations. In this context playtesting data is an important tool for communicating with clients, as project teams act much like small design studios.

Two faculty advisors are assigned to meet regularly with project team. Additionally, four checkpoints during the course of the semester involve the entire ETC faculty in evaluating and providing feedback on games in progress.

Curriculum design

To address the lack of publicly available playtesting curricula and to complement game design perspectives on playtesting with GUR methods, we developed a detailed, process-driven and hands-on playtesting curriculum. To determine the specific issues that our curriculum should address, we conducted interviews with game design students and faculty.

We heard from the faculty that students did not seem invested in playtesting. They seemed to be "checking a box," and had difficulty allocating time and resources to playtesting. Students expressed their anxieties about playtesting: they did not know where to start, had not received hands-on practice with playtesting and were afraid of making mistakes.

Faculty agreed that student playtesting needed to start earlier in the design process. However, both students and faculty revealed a tension in the perception of playtesting: some viewed it as intuitive observation of play, while others viewed it as a rigorous research method. There was also a deep concern that student teams did not know how to apply data collected in playtesting to the next iteration of their prototypes.

Informed by the literature, interviews with game designers, and our collective expertise, it became apparent that playtesting did not require a single method, but a set of methods that could be applied differently at different stages in the game design process.

Workshop design process

The workshop series was developed in collaboration with experts from a variety of disciplines, including game designers, HCI researchers, and educators. The workshop was piloted in Fall 2014 with students in a game design studio course. The workshop series was significantly redesigned and ran during the Spring 2015 semester with students completing the ETC's semester-long game design projects. The three playtesting workshops were scheduled to align with project timelines. They were redesigned again

and run in the Spring 2016 semester with ETC students completing semester-long projects.

Participants

At the beginning of the Spring 2015 and 2016 semesters, we were provided with a list of 18 ETC student teams and projects. All student teams were invited to attend the voluntary workshop series. After the workshop series, we observed a subset of teams conducting public playtests and followed up with interviews when possible.

	Attended at least 1 workshop	Attended all workshops	Playtest observations	Interviews
2015	16 teams	7 teams	8 teams	6 teams
2016	18 teams	8 teams	8 teams	2 teams

Table 1. Overview of teams who participated in the Spring 2015 and Spring 2016 workshops, including how many teams were observed playtesting and were interviewed.

2015 WORKSHOPS

Based on our research, we developed a series of playtesting workshops that focused on playtesting methods and addressed the faculty's key concerns: 1) playtesting should start earlier in the design process; 2) students needed practice and feedback on playtesting skills; 3) students needed to apply more rigor to playtesting; and 4) students did not always seem invested in playtesting. To address these concerns, we: developed one workshop dedicated to exploratory playtesting, built hands-on activities and opportunities for feedback into every workshop, emphasized rigorous experiment and data collection methods, and demonstrated how playtesting was relevant to each member of a multidisciplinary team.

Skills across three workshops

The game development process can be divided into four phases: concept, pre-production, production, and quality assurance [13]. We identified questions that designers are commonly faced with in each stage and explored how they use play to investigate those questions. For example, we asked: how do game designers use play to conceptualize a game? To iterate on a game? And to evaluate a game?

These questions became the basis for three workshops: Explore, Refine and Prove. The workshops continued to develop around the theme of "The Right Playtest at the Right Time." In each workshop, we practiced three core skills: asking good questions, choosing appropriate methods, and applying data to design. We used these skills to show similarities between the Explore, Refine, and Prove playtest approaches.

Physical Set-up

Tables and chairs were arranged around the room. On each table were a variety of post-it notes, paper and markers. Project teams sat together at a table. At the front of the room we played our slide presentation. Teams displayed their activity posters (e.g. composition box) on the wall near their tables. During the workshop, we alternated between lectures, team discussions and exercises, and adding to or notating the posters on the wall.

Explore

In the Explore workshop, we introduced exploratory playtesting techniques. Exploratory playtesting incorporates a broad set of practices used by game designers to better understand the player population and their motivations, the environment or context where the game will be played, and/or the design space within which a game is situated. Exploratory playtest methods include "playstorming," a process where game rules are generated on the fly during play; creating and testing low fidelity prototypes of a game or of a subset of game features; modifying existing game systems; or creating a scaled down digital prototype of a game. By studying players playing the game prototypes, designers discover and clarify their player experience goals and the game design features they believe will achieve those goals.

We developed a visual organization tool called the "composition box" (based loosely on a method of devising works for the theatre [7]) to help designers visualize their design space and goals. Student designers framed their composition box poster with a description of the game. Within the box the designers list the "ingredients" that might go into designing the game. These ingredients can be divided into three basic categories: 1) needs, 2) inspiration, and 3) experience. After listing the ingredients, designers outlined their player experience goals in terms of type of play (e.g. cooperative or competitive), tone, or emotions. This encouraged collaboration between the designers by sharing their expectations, creating common references, and identifying holes in their collective knowledge for further exploration.

Refine

The Refine workshop introduced playtesting methods to help teams iterate on an existing game during development. At this stage in the design process, teams had homed in on a set of game design and player experience goals and were iterating on ways to achieve those goals through their design choices. In this workshop we covered 1) posing important, accurate and answerable questions, 2) designing playtests to investigate those questions, and 3) practicing data collection techniques.

The Refine workshop began with a continuation of the composition box exercise. This time teams were asked to bring their composition box back to the workshop setting and to identify one player experience goal to frame a new composition box poster. By narrowing down the list of possible ingredients they generated in "Explore," they filled the new box with a recipe of features that they hypothesized would achieve the player experience goal, including game mechanics, narrative elements, art, and important context about the stakeholders and the play environment. Teams then brainstormed questions about the possible interactions and effectiveness of the features in their recipe. We used

these questions throughout the workshop to shape each team's research questions and playtest design.

In the second half of the workshop, we introduced data collection methods and skills. We practiced these skills by running playtests on three modified Uno decks. Participants practiced observing behavior and asking follow-up interview questions. We drew connections between observed behavior and the different mechanics of each modified Uno deck. Participants then crafted survey questions which we workshopped to reduce bias. For example, one participant asked, "How fair do you think the game is?" which after some discussion we revised to, "How much do you feel chance plays a role in the game?"

Prove

The Prove workshop focused on conducting playtests with an eye toward communicating with stakeholders. When designers have refined their games to the point where they can make claims about the player experience, it is possible to run playtests to evaluate those claims. Playtest results might be informative to players and clients, but may also serve to resolve conflicts about design within the team. We discussed experiment design methods including A/B testing, pre- and post-tests, and using in-game metrics to reveal patterns of play. We practiced communicating the strengths of your game by situating it within a body of literature and explicating your design process.

Each team began the workshop by listing the claims they could make about their game, then the evidence that it would take to support that claim, and finally the stakeholder to whom they needed to communicate the claim. Next, teams were asked to consider what that stakeholder would find persuasive. After introducing different models of experiment design, the majority of the workshop was spent in consultation with the teams as they designed a playtest that could best test their claims.

Data Collection

Data on teams' development process was collected during the Spring 2015 semester. We investigated uptake of the workshops using a variety of methods: observations of playtesting processes, interviews with students and faculty advisors, and collection and analysis of a variety of artifacts, including materials created during the playtesting workshops, weekly activity logs, and weekly reports generated for faculty.

Team name	Game topic
Team S	Game on socially sensitive topic for undergraduates
Team F	Cooperative game for families
Team J	Science game for elementary school classrooms
Team K	Enhanced storybook for middle school readers
Team B	Game to empower children with a health condition
Team D	A live game on global issue

Table 2. Overview of teams and game topics from 2015.

Midway through the semester, we approached a cross-section of teams who attended at least one workshop for permission to study their game design process. Eight teams agreed to be observed while conducting playtests, and six of these teams agreed to interviews and to having their workspace documented. Faculty advisors for the six teams interviewed were approached for additional interviews; four teams' advisors agreed. For the purpose of this study, team names are anonymized, but the topic of their game is not.

Observation

We observed eight teams playtesting for two hours at a public playtesting event, which took place mid-semester. During this event, each team conducted 2-6 playtest sessions with unique sets of playtesters. Playtesters were members of the public who were age-appropriate for each game. We took structured notes and photographs of playtesting set ups, procedures and data collection methods. We collected teams' playtesting documents, including scripts, gameplay instructions, surveys and interview protocols.

Late in the semester, design teams invited faculty into their studios to play and critique games. We shadowed two faculty members as they critiqued the teams' games and experiences included in our study. We documented the state of the game, the faculty's critique, and the design teams' discussion of their process.

Interviews

In the final two weeks of the design process, we conducted closing interviews with six teams (Table 1) and the advisors of four of those teams. Topics included the role of playtesting in their design process and how playtesting data was used to iterate on teams' games. Interviews with advisors were used to provide an expert view on the use of playtesting in the game design and development process.

Data analysis

We triangulated our findings from these data sources. Using Atlas TI software, we quantitatively analyzed interview transcripts, design artifacts and observations of playtesting sessions, using a grounded theory approach [25]. We analyzed the data to identify important concepts including how playtesting was applied and how the resulting data was used to drive the design. This resulted in 86 concepts, which were organized into three final themes.

PITFALLS TO PLAYTESTING

Our workshops were designed to teach game designers a suite of playtesting methods to be applied at different points in the design process. In each of the three stages, we guided students through the process of 1) asking good questions; 2)

choosing appropriate playtesting methods; and 3) applying playtesting data to their designs. However, through our post-workshop observations of each team's playtests, evaluation of playtesting and game design materials, and interviews with 6 teams and the advisors to 4 of the teams, we discovered that novice design teams faced stumbling blocks during each of these steps.

Setting goals and asking questions

Asking good playtesting questions was difficult for some teams because they did not have a clear concept of what their games should achieve and had not set specific player experience goals. Player experience goals help designers conceptualize "the type of experience" they want the player to have [12]. Player experience goals provide a lens through which to make design decisions and to evaluate a game. Pagulayan discusses the role of the designer in interpreting user data, stating that "the designer holds the vision for their creation as well as the vision for what makes their game fun. […] it is only the designer who can recognize when the player experience is not being experienced as intended" [23].

We noticed that setting specific player experience goals provided a frame for designing good playtests and interpreting playtest data.

Team B was charged by their client to make a game to empower children with a health condition. The team struggled with the idea of empowerment, which could have been a rich area for inspiration. Instead of asking targeted research questions around feelings of empowerment, their playtests hinged on survey questions drawn from an established engagement questionnaire. Engagement questions such as these are intended to give the designers an overall feel for how well their game is received; however, research has shown that users commonly over-report satisfaction, and this effect is likely exacerbated in children [13]. In the absence of other data, Team B used high engagement scores as a defense of their game design. When asked what the design team learned from the survey, the producer replied, "We mostly got that the game was engaging, for the most part." When pushed to identify data from their playtesting that informed a design decision the producer said, "We based a lot of parts of our game on stuff that we know that work for children." In the absence of usable data and with the reassurance of high engagement scores, the team had no reason to iterate further on their design based on playtesting. Instead they relied on internalized assumptions about what children like to guide them. In the end, they squandered the opportunity to learn about their player population and iterate on their gameplay.

Choosing appropriate methods

Some teams set player experience goals, but did not connect those goals to the design of their playtests; they therefore struggled to choose appropriate playtesting methods. The data that they collected was divorced from their core gameplay experience, so they struggled to use the data to inform the next iteration of their games.

Designing playtests that tested gameplay proved difficult for several novice teams. For example, Team K created a playful storybook experience, in which they wanted the player to feel as though "the page is magic." Despite designing a number of mechanics that activated the page, the development team had difficulty identifying and testing the features of the game that contributed to the "magical" feel. Instead they designed tests that were divorced from the core mechanic. For example, they designed an A/B test to measure reading comprehension of the game versus reading comprehension of a digital text. However, they lacked the resources to create a controlled experiment with a large enough sample size to claim any difference in learning. Not only were their results inconclusive, but they walked away from the test without generating any data that helped them iterate on their game design.

Next they designed a usability test, where they asked children to go on a "scavenger hunt" to find and play with all of the responsive features in the app. The scavenger hunt created an environment one designer described as "competitive reading." By introducing this element of competition and reducing the emphasis on the reading experience, the usability test divorced the game mechanics from the player experience they were designing for. In the usability test, hypertext features were rated as highly usable, and the team used this data to defend their design decision. Their faculty advisors, however, felt strongly that hypertext, by directing the player to a new page, went against their design aesthetic that the page itself is magic and alive: "Your design goal is not usability. Your design goal is a magical experience. And — 'but our demographic testing showed that nobody had trouble with it!' — that's not the question." In the end, the team followed the advice of their advisors, which focused on player experience goals, over their playtesting data, which measured only usability, and removed the hypertext features.

In fact, we found that development teams across the board struggled to playtest their game's core mechanic. It is easier to test art, usability, narrative, character design, or other features of a game. When teams were able to playtest game mechanics, they did so by connecting specific player experience goals to game mechanics they believed would support that experience. In other words, player experience goals seemed to be an important mediating factor in creating successful playtests.

Applying data to an iterative design process

Other teams, especially those with low-fidelity prototypes or attempting to measure the intended impact of their game, also struggled to apply the data they collected to the design of their games.

Playtesting with low fidelity prototypes

Teams embraced the idea of running early playtests — in theory. For example, Team S commented on how running early playtests was the most valuable thing they learned across the playtesting workshops: "I understand the importance of playtesting. And I understand what it can

teach you and show you about your product. But it hadn't occurred to me to playtest the idea first. So for me personally, once you talked about it, it was, like, 'duh!'" More broadly, eleven of the fifteen teams expressed intentions of conducting exploratory playtests in their weekly newsletters. However, only four teams actually followed through.

One reason for this discrepancy was that teams did not know what they were trying to learn from their prototypes. Schell asserts that "Every prototype should be designed to answer a question and sometimes more than one" [29]. Because teams did not define the problems the prototype could help them solve, they either decided against designing low-fidelity prototypes, unable to justify investing the time and resources, or they built prototypes but did not know how to interpret data generated from playtesting with the prototype.

Team D created a paper prototype for a location-based mobile game. After running the playtest, the designers were frustrated with the results, reporting that "the paper prototyping just couldn't stand up to the processing that needed to happen in a mobile experience." However, once they built the digital prototype the problems revealed in paper prototyping persisted. After some reflection, they realized that "if we had dug into the reasons why that (the failure of the paper prototype) was happening we probably would have identified a lot earlier that that part was just too complicated in the first place."

Measuring for impact and gameplay

Some games in our sample were transformational games, intended to change the player in some way. While measuring the impact of transformational games is important, focusing too tightly on impact during the semester-long design process often distracted teams from iterating on game mechanics and core interactions. This tension was demonstrated above in case of Team K. However, for some teams, impact and gameplay were closely intertwined, and for these teams interpreting data was particularly complicated. Team S designed a game on a socially sensitive topic for undergraduates. Several questions in their survey and interview protocol attempted to measure learning or a shift in attitude. Often players reported little learning: "some people are very confident… and they know what they would do, so when we asked, 'Did this teach you anything?' They said 'not really.'" These results were disappointing for the team and caused conflict about how to move forward. In order to iterate on their design, they had to ask themselves hard questions about what the mechanics were designed *to do* and what kind of experience they wanted players to have. They came to understand that: "I think being able to gauge the effectiveness is going to be difficult, because it's going to be long term. We're trying to effect long-term change." They realized that instead of trying to cause attitude change, the real goal of the game was to spark conversation. With this knowledge they shifted their playtests to measure the willingness of the player to discuss the sensitive subject matter of the game and was able to use this data to iterate on important features of the characters and dialogue.

PLAYTESTING WITH A PURPOSE

We designed the first round of playtesting workshops to encourage game design students to playtest early and often as part of an iterative design process. Based on the literature and our problem-finding interviews with faculty and students, we focused on introducing and practicing a range of playtesting methods, selecting the appropriate method to fit one's playtesting question and stage of development and then applying the data to the next iteration of a game prototype.

We initially believed that students needed more information and practice with playtesting tools and methods; however, we found that what students were actually lacking was *purposefulness*, a deep understanding of the purpose of their games, their design choices and how to test them. Game designers, whether designing for entertainment or for transformation, make design choices based on their goals; their playtest design should be just as intentional.

The pitfalls we identified across the game design teams suggested that designers did not necessarily struggle with *how to playtest*, but rather with understanding *why they were playtesting*. Novice game designers demonstrated an understanding, for example, of data collection methods and reducing experimenter bias, but struggled at a more fundamental level with problems of purpose. Without good goals, teams could not articulate the purpose of their games, much less test them. Without purposeful questions, teams playtested using methods they were comfortable with, instead of choosing the best method to fit their purpose. When teams collected data without having a clear plan, they struggled to make sense of the data and apply it to their design.

In response to our findings, we re-designed the workshops for the Spring 2016 semester. We address the major changes below.

Workshop redesign

We redesigned the workshops to fit the theme "playtesting with a purpose". Table 3 summarizes the main design iteration made in the workshops.

These changes were implemented both as uniform changes across all three workshops, and as specific changes to the content of each workshop to address the three pitfalls we observed. For example, to address choosing appropriate methods across all three workshops, we added pro and con lists to each method we presented, and an in depth example of how professional or novice designers have applied it. We also addressed "choosing appropriate methods" head-on in the Refine workshop with a hands-on activity to help students connect goals to specific questions and methods for answering them. Each of the three pitfalls informed the redesign of the workshops' content in the following way:

	Fall 2014 workshop	Revisions in 2014-2015	Revisions in 2015-2016
Overall		• Employ equal case studies from industry and academic projects • Create an overarching structure: 1) introduction, 2) asking good questions, 3) choosing appropriate methods, 4) applying data to design, 5) drafting an action plan	• Emphasize player experience goals • Develop pre-workhsop activities; leveraged to reveal playtesting goals • Develop a rubric to guide in method selection
Explore	• Introduce playtesting as part of an iterative design process • Explore a design space through prompts, bodystorming, & improv • Perform exercises in playful bodystorming and improvisation techniques.	• Focus on problem selection and asking open-ended questions • Remove bodystorming exercises • Add pre-workshop composition box activity • Add Plex Cards ideation exercise around player experience	• Provide more structure in the composition box activity. • Link exploratory methods to "what you know" about your design space • Streamline Plex Card player experience exercise
Refine	• Focus on asking an important, answerable and accurate questions • Introduce designing with stakeholder needs in mind • Exercise on persona development • Introduce affinity diagram as method to interpret playtest data	• Focus on developing player experience goals • Uno game mod used as a playtest observation exercise • Practice crafting and asking interview questions	• Bridge activity to craft a "recipe" of ingredients that serve player experience goals • Integrate methods more fully • Allow time to discuss action plans in the workshop.
Prove/ Persuade	• Focus claims/evidence structure • Introduce designer's judgment as a way to advance game design	• Add lecture on reliability and validity • Add examples of experimenter bias • Add pros and cons for each testing method	• Change workshop name from "Prove" to "Persuade" • Divide methods into two groups: experimental (A/B, pre/post) and expertise (expert panels, designer's judgment).

Table 3. Overview of revisions across three iterations of workshops.

Explore

Realizing both how difficult and how important it is for designers to set meaningful and specific goals for their games, we redesigned the explore workshop to better support setting player experience goals. While keeping the composition box exercise as a pre-workshop activity, we restructured it with a focus on exploring player experience goals, what you want the player to feel, think or do.

In the workshop, we suggested four starting points for the game design process and linked them to pre-production research methods: 1) Observe: if you know your target population, conduct field observations or design a playtest of a typical game for your population, 2) Probe: if you are starting with an idea for art or other assets, begin testing it to learn how potential players respond to it, 3) Test: if you know your core mechanic, create a low fidelity prototype or mod of an existing game and begin playtesting, 4) Co-design: if you are working with stakeholders to create a specific experience or impact, introduce play and the language of games into your relationship early.

Next, we used Plex cards to introduce different types of play experiences, such as competition, control or fellowship [21]. Students selected a card and created a mind map with this player experience in the center. They distilled their insights and added these notes to the output of their composition box, adding to their understanding of their player experience goals.

We repeated this simple framework of starting with what you know, identifying the gaps in your knowledge through research, and then adding to your understanding of the desired player experience in the pre-workshop and Plex card activities. We also demonstrated how to apply exploratory playtesting methods to this framework and led students in designing an exploratory playtest.

Refine

To help students select appropriate playtesting methods, we first revisited their player experience goals, hypothesized how specific game features affect the player experience and then determine which playtesting method to use.

In the pre-workshop activity, we provided more structure on how to create a composition box recipe for a prototype: 1) Frame the composition box with a player experience goal; 2) Fill the box with the set of features, or "ingredients," you hypothesize will support the player

experience goal (consider affect, themes, environment, objects, mechanics, relationships, and events); and 3) Pull from the composition box sketches for a prototype incorporating these features.

Throughout the workshop, students practiced asking questions about how their set of ingredients supported their player experience goals. These questions were developed first through observation of play (e.g. What evidence are you looking for?), then through asking interview questions of players, and finally by writing survey questions.

Prove to Persuade

We reframed the Prove workshop as Persuade because we recognized that novice designers, especially those designing transformational games, felt intense pressure to prove that their games were successful and impactful. This need to validate one's designs can stymie the iteration process and make it difficult to pivot. In Persuade we focused on making a persuasive claim about a game based on evidence in order to elicit the support of stakeholders.

Of the four methods we taught, two were experimental methods, A/B testing and pre/post testing, and two were open-ended methods based on expertise, expert panels and designer's judgment. In practice, even with extensive playtesting, design decisions come down to judgment calls. We recognized the role of expertise in the design process to dissuade novice designers from running unproductive playtests, or simply "checking the box" without purpose. We provided examples of making a persuasive case for your game both based on data and through a clearly developed design rationale.

Data Collection

We investigated uptake of the workshops using a variety of methods: observations of playtesting processes, interviews with students, and collection and analysis of a variety of artifacts, including materials created during the workshops, materials displayed in student's workspace, and materials created in support of playtests after the workshops.

Data collection for this semester is ongoing. To date we have observed 8 teams playtesting and have conducted semi-structured interviews with 2 teams.

Team name	Game topic
Team L	VR public speaking simulation
Team H	Problem-solving game for boys

Table 4. Overview of teams and game topics from 2016.

EVIDENCE OF IMPACT

By observing the activities of design teams (Table 4), we can see that intervening at the level of *purpose* rather than *method* leads to the design of playtests that are more relevant to the designers and inform further iteration on their game prototypes. We have looked at design teams' activities rather than outcomes because we aimed to improve the process by which playtests are designed.

Specificity

We have observed that teams can identify more specific design elements within their prototype. Specificity demonstrates an awareness of their design choices and reflects the purposefulness of those choices. We see evidence of specificity in the materials generated across the workshops.

Providing more structure to the composition box exercise allowed teams to think about their game concept, their player experience goals and how their design choices support those goals more analytically. We observed students express more unique and purposeful player experience goals in their workshop materials. When asked to develop player experience goals the 2015 teams tended toward blanket statements, such as "fun and engaging," "to empower children," or "enhance a storybook." 2016 teams, on the other hand, demonstrated more nuance in their goals, showed greater consideration of the player's emotional experience, and pointed toward actionable design choices. For example, a team developing a VR experience to practice public speaking sought to create a game that dealt with feelings of "tension and anxiety" and gave the player "self confidence for presentations." A game geared toward elementary school boys wanted players to feel "powerful, accomplished and smart."

We observed an increase in analytical thinking, manifested through the specificity of students' playtest design. For example, when asked what evidence they needed to collect to test their claims, a 2015 team designing a storybook experience answered "observation," "surveys," and "test reading comprehension," whereas a 2016 team designing a VR training simulation, said they would measure "volume, pauses, eye gaze tracking, and time management."

Ability to pivot

Pivoting in the design process helps designers get unstuck and innovate. In order to pivot, a designer needs a deep understanding of both their goals and the design space. Pivoting is evidence that a development team is making a purposeful turn based on what they have learned through the iterative process of prototyping and playtesting.

From the teams who attended the workshops, we identified teams who attempted to pivot. We selected teams for case studies on pivoting based on availability for interviews.

Pivoting goals

Team L (2016) demonstrated an ability to pivot, shifting player experience goals and design aesthetic based on player feedback. As with many VR experiences, Team L believed that immersion, the feeling that the player was present in a real world, was the most salient feature of their VR training simulation. In their first set of playtests, they focused on how natural, realistic and believable the interaction was. However, they realized that to operate as a training simulation, the readability, credibility and applicability of the feedback provided to the player was of paramount importance. They then ran playtests that focused

on how and when feedback was displayed to the player and noticed that because head-tracking stood in for eye-tracking (a constraint of the mobile VR technology), players had difficulty trusting feedback on their gaze. To alleviate this problem the team added a crosshair to the screen, signaling to the player where they are looking at all times. This reduced the realism of the experience, and arguably detracted from a sense of immersion in a virtual world. The claim/evidence/stakeholder framework introduced in the workshop assisted the reevaluation of their goals. They then dedicated their second round of playtesting to evaluating feedback to the player, and turned their focus to the quality of the experience as a training simulation.

In contrast, because Team K (2015) tested for factors such as reading comprehension and usability divorced from the player experience, they were unable to prioritize the player experience in their design choices. Instead it was their advisor who intervened to preserve the player experience of interacting with a "magic" page. Team K struggled to design playtests that served their purpose, while Team L designed purposeful playtests.

Pivoting gameplay

Team H (2016) designed a problem-solving game for elementary school boys. Their goal was for players to feel empowered, to demonstrate self-determination and feelings of competency, but also to practice computational thinking. They began iterating on a paper prototype based on a Pac-man-like chase mechanic. Team H playtested their prototype once a week at an elementary school. Playtesting with paper prototypes went well, but when they transferred to a digital prototype, players overwhelmingly wanted to attack the enemy and lost sight of the problem-solving objectives. To address this issue they returned to the paper prototype and went through seven more iterations of mechanic and level design, playtesting each one, before finding one that struck the right balance between a fighting game, thus fulfilling the goals of feeling empowered, and a strategy game, fulfilling the goal of practicing computational thinking. Team H went into each of their weekly playtests with a prototype of their "best guess" at this balance, and every week they collected actionable data from their playtests that informed the next iteration.

Compare Team H's methodical and analytical process of finding the right balance between empowerment and challenge in their game to Team B's (2015) reliance on user engagement surveys and their own heuristics. Team H approached each playtest as an opportunity to iterate meaningfully on their design, while Team B, despite their efforts to apply validated methods, was unable to generate data that truly informed their design.

LIMITATIONS

We believe that the workshops represent an approach to playtesting that contribute to our understanding of game development and to game design education. However, our research has some limitations: workshops were conducted at one institution with a relatively small sample of student game designers; and, although faculty did not express any discernable differences in ability or attitude between the class of 2015 and 2016, we cannot rule out the possibility that students in the 2016 workshops started the semester with greater expertise.

CONCLUSION

We conducted research to improve the process of game design and development. We first focused on training students in testing methods, but found that the real need was to understand the overarching player experience, then to create a research plan that supported delivering that experience. We demonstrated how purposeful playtesting, designing playtests to address a designer's specific set of questions, positively impacts an iterative game design process. We provide evidence that by applying methods taught in our workshops, novices leveraged playtest methods and tools, selecting and executing playtesting and data collection methods appropriate for their goals, and effectively applied playtest data in iterative design.

Purposefulness is the hard part of playtesting. Being purposeful requires more than sound experimental methods, survey writing skills or good heuristics to incorporate feedback from the player in an iterative game design process. It requires that the designers have a vision for their game, can translate that vision into design choices, and can design purposeful playtests to test their goals and the player's expectations. Our workshops provided evidence that structuring playtesting methods with the purpose of understanding the overarching game experience is useful, and may even result in more innovative games. Designing these workshops in the context of a graduate program, provided us insight into how much there is still to understand about when, how and why game developers apply testing methods.

Future work in this area can address several research trajectories. First, we can conduct similar work with other populations of game designers who face resource challenges (e.g. indie and transformational game designers) to further validate our findings across populations. Second, we can develop tools and methods to support game designers, such as we have begun to do with the composition box, and recommend best practices for playtest processes. Third, we can develop testable models of the playtest process that will help us generalize our insights. We seek to understand not just *how* to support game designers in playtesting in ways that produce better games, but in understanding *why* certain types of playtesting processes work. Finally, we will continue to iterate and evaluate our playtesting workshops. We offer this work to the community, with the hopes that we can collectively reach the long-term goal of improving games and game mechanics across many domains.

ACKNOWLEDGEMENTS

This research was funded by the Simon Initiative. We thank the students and faculty of the ETC.

REFERENCES

1. Abuhamdeh, S. and Csikszentmihalyi, M. 2012. The Importance of Challenge for the Enjoyment of Intrinsically Motivated, Goal-Directed Activities. *Personality and Social Psychology Bulletin 38*, 317–330.
2. Ambinder, M. 2011. Biofeedback in Gameplay: How Valve Measures Physiology to Enhance Gaming Experience. http://www.valvesoftware.com/publications/2011/ValveBiofeedback-Ambinder.pdf.
3. Anthropy, A. and Clark, N. 2014. *A Game Design Vocabulary: Exploring the Foundational Principles Behind Good Game Design*. Pearson Education.
4. Barr, P., Noble, J., and Biddle, R. 2007. Video game values: Human–computer interaction and games. *Interacting with Computers 19*, 2, 180–195.
5. Brockmyer, J.H., Fox, C.M., Curtiss, K.A., McBroom, E., Burkhart, K.M., and Pidruzny, J.N. 2009. The development of the Game Engagement Questionnaire: A measure of engagement in video game-playing. *Journal of Experimental Social Psychology 45*, 4, 624–634.
6. Cabrera, E. 2014. League of Legends - The Potential of A/B Testing. *gamasutra.com*. http://www.gamasutra.com/blogs/EduardoCabrera/20141020/227964/League_of_Legends__The_Potential_of_AB_Testing.php.
7. Choi, J., Forlizzi, J., Christel, M., Bates, M., and Hammer, J. 2015. Compositions: Moving from Ideation to Prototyping. Games + Learning + Society 11.
8. Csikszentmihalyi, M. 1990. *Flow: The psychology of optimal performance*. Cambridge University Press.
9. Desurvire, H. and El-Nasr, M.S. 2013. Methods for game user research: Studying player behavior to enhance game design. *IEEE Computer Graphics and Applications 33*, 4, 82–87.
10. Desurvire, H. and Wiberg, C. 2009. Game usability heuristics (PLAY) for evaluating and designing better games: the next iteration. *Online communities and social computing*. Springer Berlin Heidelberg, 557-566..
11. El-Nasr, M.S., Drachen, A., and Canossa, A. 2013. *Game Analytics: Maximizing the Value of Player Data*. Springer Science & Business Media.
12. Fullerton, T. 2014. *Game Design Workshop: A Playcentric Approach to Creating Innovative Games, Third Edition*. CRC Press.
13. Hull, R. and Reid, J. 2003. Designing engaging experiences with children and artists. *Funology: From usability to enjoyment*, 179–188.
14. Hunicke, R., LeBlanc, M., and Zubek, R. 2004. MDA: A Formal Approach to Game Design and Game Research. *Workshop on Challenges in Game AI,* 1-4.
15. Isbister, K. and Schaffer, N. 2008. *Game Usability: Advancing the Player Experience*. CRC Press.
16. Jennett, C., Cox, A.L., Cairns, P., et al. 2008. Measuring and defining the experience of immersion in games. *International Journal of Human Computer Studies 66*, 9, 641–661.
17. Jr, M.W., Sathiyanarayanan, P., Nagappan, M., Zimmermann, T., and Bird, C. 2016. "What Went Right and What Went Wrong": An Analysis of 155 Postmortems from Game Development. *Proceedings of the 38th International Conference on Software Engineering Companion*, ACM Press. 280-289.
18. Lankoski, P. and Björk, S. 2015. Game Research Methods: An Overview. Lulu.com. http://press.etc.cmu.edu/files/Game-Research-Methods_Lankoski-Bjork-etal-web.pdf.
19. Lewis-Evans, B. 2013. Intro to Games User Research Methods. http://www.slideshare.net/Gortag/intro-to-games-user-research-methods-march-2013.
20. Lewis, C. and Mack, R. 1982. Learning to use a text processing system. *Proceedings of the 1982 conference on Human factors in computing systems - CHI '82*, ACM Press. 387–392.
21. Lucero, A. and Arrasvuori, J. 2010. PLEX Cards : A Source of Inspiration When Designing for Playfulness. *Fun and Games 15*, 17, 28–37.
22. McMahan, A. 2003. Immersion, Engagement, and Presence. *The video game theory reader Immersion,* 67–86.
23. Medlock, M.C., Lead, U.T., Studios, M.G., et al. 2002. Using the RITE method to improve products ; a definition and a case study. *Usability Professionals Association*.
24. Pagulayan, R.J., Keeker, K., Fuller, T., Wixon, D., Romero, R.L., and Gunn, D. V. 2002. User centered design in games, 1–46.
25. Patton, M.Q. 2005. Qualitative Research. In *Encyclopedia of Statistics in Behavioral Science*, 1633–1636.
26. Pulsipher, L. Game Playtesting: the Heart of Game Design - Udemy. *udemy.com*. https://www.udemy.com/game-playtesting/.

27. Salen, K. and Zimmerman, E. 2004. *Rules of Play: Game Design Fundamentals*.MIT Press.

28. Salen, K. and Zimmerman, E. 2005. Game Design and Meaningful Play. In *Handbook of computer game studies* The MIT Press. 59–79.

29. Schell, J. 2015. *The Art of Game Design: A Book of Lenses*. CRC Press.

30. Vorderer, P., Klimmt, C., and Ritterfeld, U. 2004. Enjoyment: At the Heart of Media Entertainment. *Communication Theory 14*, 4, 388–408.

31. Zimmerman, E. 2003. Play as Research: the iterative design process. In *Design Research: Methods and Perspectives*. 176–184.