

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/220200730>

A survey method for assessing perceptions of a game: The consumer playtest in game design

Article · October 2005

Source: DBLP

CITATIONS

69

READS

1,060

3 authors, including:



Keith Steury

12 PUBLICATIONS 456 CITATIONS

SEE PROFILE



Randy J. Pagulayan

Microsoft

17 PUBLICATIONS 1,115 CITATIONS

SEE PROFILE



John P. Davis, Ph.D.
User Research
Engineer
Microsoft Game
Studios

John Davis is a User Research Engineer for Microsoft Game Studios and has been a member of the group since 2002. He has worked on several titles, including *Jade Empire* and *Inside Drive*. John has co-instructed a successful tutorial on conducting usability research for games at the Game Developers Conference (GDC), and has been a regular contributor to the Conference on Computer-Human Interaction (CHI). Before coming to the games user research group, John was a Usability Engineer in the Social Computing Group (SCG) at Microsoft Research and an assistant professor of psychology at Seattle University. John earned his B.S. in psychology from Texas A & M University (yes, he is an Aggie) and his Ph.D. in Experimental Social and Personality Psychology from the University of Washington.

Keith Steury, M.S.

Keith Steury was formerly a User Research Lead for Microsoft Game Studios, where he worked on over a dozen Microsoft games including *Halo* and *NFL Fever*. Keith has coauthored several articles about the role of human-computer interaction in games and has spoken at the Game Developers Conference (GDC)

A survey method for assessing perceptions of a game: The consumer playtest in game design

by John P. Davis, Keith Steury, and Randy Pagulayan

INTRODUCTION

The computer and video game industry has only relatively recently burgeoned into one that rivals the film industry in terms of consumer spending. In the United States alone, the games industry reported about \$6.9 billion in sales in 2002, and sales increased to \$7 billion in 2003 and \$7.3 billion in 2004 [1]. Increased sales have also led to increased competition among games developers, as they vie for a share of the growing wealth. Because higher-quality games tend to sell better, game developers are increasingly looking for ways to improve their games.

Consumers can provide valuable information about various aspects of a game to help make it more fun to play. Methods for obtaining consumer feedback have been applied to games with varying success. Some methods, such as traditional usability testing, can be an excellent source of behavioral information about consumers. Usability research can help identify problems or issues that block users from experiencing the fun of a game. Surveys and focus groups can provide limited information about how consumers perceive games or their attitudes toward them in some contexts.

Each of these methods, however, falls short in providing adequate information from consumers about how they perceive a particular game, specific, actionable feedback that game designers can use to make their games better. Usability tests, for example, are best at detecting problems. While we can glean some perceptual information from consumers in usability tests, the typically small sample sizes prohibit generalization to a wider population. Surveys and focus groups can collect consumers' evaluations of games, but they are often far removed from the immediate experience of a game.

In this paper we describe the *playtest method* we have developed, a method that combines traditional, scientific survey methods with a controlled laboratory environment to collect systematic, quantitative information about consumers' perceptions of games. The playtest method for gathering consumer perceptions of a game can be used to aid in the iterative development of games that has been described by Salen and Zimmerman [15] and other game designers. We describe some of the ways we have used playtests to provide feedback to game designers and also present a case study of how the method has proven effective at making a game better—that is, more fun.

Games vs. Productivity Software

and the Conference on Computer-Human Interaction (CHI). Before coming to the Games User Research Group, Keith contracted as a usability engineer at Microsoft Research. Keith holds a M.S. in cognitive psychology from the University of Washington where he studied language processing through functional magnetic resonance imaging. Keith currently resides in Sydney, Australia.

Randy Pagulayan, Ph.D.

User Research Lead
Microsoft Game Studios

Randy Pagulayan is a User Research Lead for Microsoft Game Studios and has been a member of the User Research group since 2000. He has previously worked on over a dozen Microsoft games, including *Halo 2*, *Brute Force*, *RalliSport Challenge* (1 & 2), and *Top Spin*. His most recent publications include several co-authored book chapters: *User-Centered Design in Games* (2003), *Designing for Fun: User-testing Case Studies* (2003), and most recently an invited chapter for the *International Encyclopedia of Ergonomics & Human Factors* (2nd ed). Randy has also been an invited speaker for the Nielsen Norman Group User Experience Event (2003). Randy has a B.A. in Psychology from the University of Maryland, and a Ph.D. in Experimental Psychology from the University of Cincinnati.

Games vs. Productivity Software

Games share some important similarities with productivity applications, similarities that make the process of improving the user experience with games similar in some ways to the one used with applications [4]. Like productivity software, games contain menu systems for changing options, interfaces to communicate status information, and input and control devices. Furthermore, games also require that users develop a conceptual understanding of the rules of use, and often offer tutorials and help systems to aid players in learning basic skills. Because of these similarities, methods for obtaining consumer feedback about productivity applications can also be used with games.

But games also differ in several fundamental ways from productivity software, and the differences make the process of acquiring and using information to improve the user experience in games more challenging than productivity applications [4]. On the most basic level, the primary goal in a game is to be enjoyed, something that is rarely the case in a productivity application. Users of productivity applications must be satisfied with the experience, but the primary concern is that they are able to do what they need to do (accomplish tasks) easily, quickly and effectively. In contrast, games must be “fun.” Games must also be challenging, but challenge is something that applications are typically designed to minimize. Lastly, games constantly try to innovate by trying novel things, including interfaces, controls and cutting edge technology; applications, on the other hand, attempt to be consistent and typically change only incrementally from version to version. The last point increases the challenge that accompanies user research in games: user researchers must work harder to identify the issues associated with the innovations in games and be equally innovative in adapting their methods to gather information from consumers to address the issues.

Games user researchers are naturally concerned that players have an enjoyable experience, but what makes games “fun”? The nature of “fun” experiences in games has been the topic of both speculation and more systematic inquiry for some time. Specific psychological principles, such as schedules of reinforcement, play a key role in keeping players engaged in a game [8]. There has also been some qualitative research on the characteristics that make games fun, such as their challenge, and the fantasies and sense of curiosity they evoke in the player [10]. Some previous research has sought to explain qualitatively the characteristics that make games engaging and their “playability” [e.g., 10; 13; 12]. Lazzaro [12] has also explored some of the underlying motivations people have for playing games using qualitative methods. In addition, practitioners have elaborated additional principles that should be used when designing games, and key pitfalls to avoid, have also been described [e.g., 9, 11].

Consumer Feedback and Games

While some of the factors that make games fun, and player motivations for playing games, are relatively well understood, systematic quantitative methods for measuring and assessing the fun of a game are rarely employed in the games development process. We earlier suggested, however, that some of the challenges found in games development are similar to those faced in the development of productivity applications. As a consequence, some familiar user research methods that have traditionally been

used to make applications better are also used to make games better [2]. We first briefly discuss such qualitative and quantitative methods, and then describe the combination of methods we have developed to collect evaluative information from consumers about their gameplay experiences.

Focus Groups

Focus groups have been used in the game development process for various purposes, including helping game designers get a better sense of how consumers feel about various aspects of a game or, more likely, a game concept. Focus groups typically comprise 6-12 consumers from a particular group (e.g., people who play racing games). Several such consumers meet together to discuss topics that the game design team is interested in, such as what features in the game are important to the consumers and what they think of various design concepts. Story boards that describe the game's premise or the game's design mechanics sometimes accompany the discussion.

Focus groups can be useful for concept generation in the initial stages of a project or for obtaining a better general understanding of a problem space in some circumstances. However, they are poor at providing specific, actionable data that help game designers make their games better for several reasons. In focus groups, consumers are often asked to give their reactions to abstract gaming concepts or ideas rather than the implementation of the concepts (though this may sometimes be the case). Judgments about the value of a concept can be dramatically different from judgments of the concept's practical implementation in a game. Focus groups are also susceptible to a host of group pressures that impact the quality of the information they yield. One or two group members may dominate the discussion, while contributions from less vocal members are lost.

Retrospective Surveys

Surveys have also been used frequently to gather information from consumers. For example, a game developer might conduct a survey of players to find out what features they like and dislike in a certain type of game, such as a driving game. Surveys, therefore, tap players' perceptions of games. Conducted correctly, they yield important data that designers can use to aid them in designing features and more fully understanding how people who play games think about them—what they like, dislike or are indifferent to.

Surveys are difficult to conduct well, however. Survey questions are tricky to write in a way that yields high quality, actionable data.

Importantly, the sample for the survey must be large enough to enable meaningful statistical analyses and comparisons, and it must be representative of the population of interest. If I want to know what "racing gamers"—the people who will potentially buy my game—think about my game, I must ensure that my sample truly represents the population of racing gamers. Surveys are also retrospective in nature. When players are asked in a survey to give their views about a game, they are asked to think about their past experiences with a game, experiences that may have varied widely. Survey respondents may have finished the game, quit because it was not fun or played very little of the game. It is often not clear, therefore, that survey respondents are reporting their

perceptions of the same gaming experience, which makes it difficult to draw generalizable conclusions from data collected in this manner. Furthermore, self-reported judgments about experiences that took place in the past are subject to biases and other difficulties that are well documented [6].

Beta Testing

Another popular source of information about gameplay comes from *beta testing* an early build of the game. Beta tests are used in all types of software development, and are often an indispensable source of data for developers. Beta testers are typically volunteers who are recruited through various sources to play an early version of a game. They are generally asked to provide information about technical issues or bugs in the software, but they can also provide feedback about issues relating to gameplay.

While the results from beta tests are crucial for identifying important bugs in games, they are less satisfactory in helping to identify and fix gameplay issues or issues relating to fun. Beta testers are typically very advanced players—the types who are eager to play a new game and volunteer to do so for hours at a time with minimal tangible reward. As a consequence, they are not members of the “typical” gamer population—they are experts. Further, gameplay feedback from beta tests is rarely (in our experience) systematically collected in a rigorous manner. Beta testers are asked to play the game as much as they can and to give feedback about any issue they desire, generally in any form they can. Lastly, developers have little control over the gameplay experience the beta testers have. While the developer can ask the testers to focus on particular issues or areas of the game to provide feedback, because beta testers play at home (or work), the developer has little or no control over the environment. Such variability hinders the value of beta tester feedback to determine whether an “issue” that is identified is general or is particular to the individual tester or his or her environment and experience.

Usability Testing

More traditional HCI methods have proven to be very useful in the design of games. For example, usability tests are used to identify problems and issues in games, much as they are used in productivity applications. We mentioned earlier that games, like productivity software, must generally be easy to learn and easy to use. For example, it is important that a new user of a console game (e.g., the Xbox or Playstation 2) be able to pick up the controller and immediately play the game and have fun. The player’s initial experience in the game is critical to ensure that he or she continues to play, usability tests can be used to ensure that new users (and more advanced users) can quickly and effectively start a game from the game’s menu system (the “gameshell”). In the game, the player’s controls must be intuitive and easy to learn, and usability methods can shed additional light on how to ensure that this is the case.

Usability methods can also be used in novel ways to identify problems and their causes during actual gameplay, after the player has successfully started a game. Many challenges or puzzles within a particular game require players to figure out the correct sequence of actions to take or the correct combination of items to use to solve the challenge. Confusion about such tasks can often

be uncovered and corrected with data gathered in usability sessions.

Another novel use of a usability test in games is to discover whether the experience the player has in the game matches the experience the designer *intended* the player to have. A game designer, for example, may craft a particular scenario in a game with the intent that the user will play the game a particular way to have the most fun. Usability tests can assess whether users really do play through the scenario as the designer intended them to, and if not, why not.

An example of utilizing usability testing methods to ensure that the designer's intent has been met in a game comes from the development of the Xbox game *Halo: Combat Evolved* [5]. Halo is a game where players play a soldier character, and they must engage alien enemies in combat using weapons of various varieties, including guns. During one "mission" in the game, the game's designers intended for players to engage the enemy from short range. In usability testing sessions, however, novice players engaged the enemies from much longer range than the designer intended; they would see the enemy in the distance and shoot them with their assault rifles. Note that this strategy was effective, in that players could successfully complete the mission in this manner. The designers, however, had created the missions so that they would be most fun if the player engaged the enemy at the desired shorter range.

Given the usability data that showed players engaging in combat from greater distances than they intended, the game's designers came up with some novel solutions to ensure that players moved closer to enemies before engaging them. For example, they changed the behavior of the aiming reticle, the graphical indicator on the screen that indicates where the user is aiming a weapon. In *Halo*, the reticle is normally light blue, but it turns red when it is positioned by the player over an enemy, indicating that the player's shot will hit the enemy. The designers adjusted the behavior of the reticle so that it turned red only when the player was closer to the enemy. The designers also adjusted the behavior of the enemies, so that when they were fired upon at long range, they would run for cover. Consequently, players had to get closer to enemies to engage them in combat. A few other adjustments to the game were also made to ensure that players engaged enemies from a reduced distance.

Follow up tests were performed to assess the impact of the changes the design team had made. As hoped, the changes encouraged players to engage enemies from closer range, as the designer team intended. Further, users reported (anecdotally) more satisfaction with the performance of their weapons, and they seemed to enjoy the game more.

As we have described, well-designed usability tests usually produce detailed, actionable feedback about some of the important issues or problems in a game, and are therefore valuable tools to make games better. They can also be used in novel ways. In the case just described, usability feedback was used to ensure that players experienced the game in the way the designers intended, and to gather information about why this was or was not the case. However, usability methods also have shortcomings that are apparent when they are used to provide consumer feedback about

games. They are not good at providing generalizable data about how consumers overall will respond to important dimensions of the game. For example, if one or two usability participants (out of 8) have difficulty completing a mission during a usability session, can we conclude that the mission is too challenging and recommend to the design team that they make the mission easier? Subjective reports from usability participants about their experiences can be useful in pointing out potential problems (e.g., repeated cursing on the part of a participant probably indicates he is frustrated). It can be risky, however, to generalize anecdotal information from a small number of participants to a population of gamers. The value of a usability test comes from its focus on a user's behavior, not from users' subjective evaluations. While usability tests can include enough participants to allow generalizations, running enough participants is extremely costly in terms of both time and money. In our experience, larger scale tests are seldom run in practice.

Surveys and Play: the Playtest method

We have described some of the methods that are typically used to collect information from consumers to make games more fun. Focus groups can generate ideas, but the information obtained is usually general and subject to the effects of the group environment. Surveys are useful in obtaining generalizable information from consumers about how they perceive games, but the information is often not specific enough to be useful for designers. Usability tests are extremely useful in gathering in-depth and actionable feedback from consumers about a game's design, but the information is typically about a user's behavior—what they do in the game—rather than about a user's perceptions of the game. Furthermore, the feedback usually comes from a small number of users, so the findings may not be applicable to all players in the population of interest.

To obtain specific information about how consumers perceive different aspects of a game, we have combined surveys with hands-on gameplay into a method we call the *playtest method*. The goal of the playtest is to obtain feedback from consumers about their experiences with a specific game in a systematic manner using scientific methods. The ultimate goal is to provide game designers with actionable feedback from consumers about how they perceive critical aspects of their games. While the methods that comprise the playtest are not themselves unique, we feel that their use in combination to test the "fun" of games is.

Characteristics of the Playtest

In a playtest, a sample of consumers is selected from our database of game players and asked to come to our lab facility to participate. When they arrive, they are met by trained playtest moderators who guide them through the procedure. They first complete a screening questionnaire to ensure that they meet the requirements to participate. They are then escorted to the playtest lab, where they are seated at a standard station with standard equipment. Though they are seated in the same lab, players cannot see what the others are doing, and all players wear headphones so distractions are minimized.

Players are then given a copy of the game and instructed to play it as they normally would at home. If the game is one that has already been released, players are given all materials that came

with the game, including the manual and other ancillary materials. After they have played, participants answer specific questions about various aspects of the experience, including overall fun, perceptions of the graphics, controls, story (if applicable), sound, and other play elements that are part of the game (e.g., ratings of the elements of "combat" if it is a fighting game or "driving" if it is a racing game, etc.).

The playtest has several characteristics that are essential for obtaining quality, actionable feedback from consumers about their perceptions of a game's key dimensions.

A Focus on the Initial Experience

The "typical" playtest focuses on the players' initial experience of a game, which we operationally define as the first hour of play. Why do we focus on the initial experience? When someone plays a game for the first time, the first hour or so of play is critical. If the player finds the first hour of play compelling and interesting, he or she is more likely to continue playing. With so many activities competing for consumers' attention and time, including television and other forms of entertainment, ensuring that the first hour is fun is essential for success.

Larger Samples

One of the weaknesses of usability tests as they are applied to games is their reliance on data collected about the behavior of only a few participants. Small samples work well in a usability setting, where detailed behavioral data often provide insight into usability issues. However, usability tests are not intended to provide generalizable data about how a population of gamers perceives a game. Instead, usability tests are designed to gather behavioral data about the problems and issues experienced by "these 8 gamers, playing this portion of this game."

In contrast to small-sample usability methods, the playtest method relies on data collected from larger samples. We generally rely on a sample size of 25-35 participants, which is based on power analyses of the statistical tests we employ. Meaningful comparisons can therefore be made between consumers' perceptions of different games of the same type (e.g., two different basketball games) or between versions of the same game (e.g., this year's version of the game compared to last year's version).

Comparisons are essential for gaining an understanding of which issues are truly problems to be addressed by the design team and which issues are not.

For example, if the results of a playtest show that 70% of the participants "liked" the character animations in a basketball game, is this good or bad? Should the game's designers spend valuable resources to improve the character animations, or is the 70% finding typical or even above average for games of this type? Only by comparing the current data to data collected about consumers' experiences with other games are we able to make informed decisions about what the "real" issues are.

Standardized Questions

In typical playtests, participants answer questions about their experiences immediately after they have played. Many of these questions are standardized based on previous tests. For example, players may be asked to give an overall rating of the "fun" they

experienced in the first hour of play, ratings of the game's pace, quality of the graphics, sound, controls, etc. In addition to the questions that are asked in most playtests, players are asked additional questions about dimensions that are unique to the type of game being tested. The exact questions that are asked depend on the type of game being tested and the focus of the playtest. In a test of a racing game, for instance, players may be asked to rate their perceptions of the different types of vehicles they operated, the sensitivity of the controls, and the tracks on which they raced.

Lastly, participants may be asked questions about their perceptions of elements of the game that are unique to the game being tested, questions that are usually developed by the user researcher responsible for the test.

The use of standardized questions in playtests makes several important things possible. First, standardized questions facilitate comparisons between consumers' ratings of the dimensions of different games. New versions of a successful game are often released that contain novel gameplay features, so standardized questions facilitate comparisons between ratings of the different versions to track the impact of changes to the game's design (e.g., *Halo* can be compared to *Halo 2*). Importantly, we are also able to compare ratings for successive versions of an in-development game to track the effects of changes to the game's design. For example, problem areas in a game can be identified through a playtest, and design changes are made as a result. Standard questions allow tracking of the effect of the changes; in other words, we are able to determine whether the changes improved the game as intended.

In addition to the standardized questions, the design of the playtest also gives the user researcher flexibility because he or she can add questions to assess dimensions of the game that are unique to the game being tested. Standard rating-type questions are limited, however, in that they typically yield only a number, which may not be helpful in understanding the reasons participants experienced a game in a particular way. Another important component of the playtest questioning process is the inclusion of open-ended questions in which participants can give additional feedback about their experiences. For example, participants may be asked to describe what they liked about the game, what they disliked and other factors. By examining responses to open-ended questions, the user researcher can develop a more comprehensive understanding of the reasons behind the ratings, and identify areas that may need to be further explored in additional tests.

Targeted Samples

We mentioned earlier that sampling from the population of interest is essential to conducting high-quality surveys that yield generalizable data. The same is true with the playtest method.

When conducting a playtest on a specific game (e.g., a basketball game), we first construct a "profile" of the participants that represents the target population of consumers. For example, if we are conducting a playtest of a basketball game, the profile may ask for participants who have played sports games (and specifically basketball games—but not the one they will be playing in the lab). To gauge players' experience playing particular types of games, we may ask them to tell us which games they have played and estimate how much experience they have with specific

games.

By collecting information from gamers who fit into the target consumer group, we can be more certain that the data we gather truly represents the thoughts and feelings of the player population. Remember, the goal of the playtest is to gather information about games to help the design team make their game better.

A Focus on Identifying Issues

The overarching goal of the playtest is to identify problems or issues in games by collecting data from consumers. Comparisons of ratings of the current game to ratings of other games are essential for identifying such issues. Playtests are not generally intended to serve as a test of the game's "quality." The tests focus on the first hour of gameplay, which we feel is critical. However, a great deal of play takes place after the first hour, play that is not generally assessed in playtests (an issue we address later).

A Controlled Environment

One of the hallmarks of the playtest is experimental control. All participants are treated in the same way from the time they are contacted about participating to the time they leave the lab. They are given the same instructions, play the same game, play individually (without interacting with others) and answer the same question about the game as they play. All participants in a playtest play the game using identical equipment, and personnel conducting playtests are thoroughly trained to ensure that they treat participants consistently. To the extent possible, the methods employed in the playtest limit the potential effects of environmental variables in the test that do not pertain to the game.

Low Cost and Rapid Turnaround

A usability engineer might wonder what real advantages playtests offer over usability tests with comparable sample sizes. Indeed, usability tests often provide richer and deeper qualitative information than playtests. However, one of the key characteristics of the playtest is its comparatively low cost. A playtest with a sample large enough to allow for statistical analyses of the results can be completed in a few hours and with minimal supervision. A comparable usability test would take far longer and be much more expensive. For instance, usability test with 30 participant sessions would take about 60 hours to complete (30 participants x 2 hours each), an estimate that does not include the time to prepare the test and analyze the results! (Note that the playtest's reduced time and resources is achieved in part because we have access to relatively large playtest labs where we can run multiple participants in the test at the same time; however, even if we were to playtest a few participants at a time, the cost would be greatly reduced compared to usability tests or individualized interviews because oversight is minimal). The relatively straightforward nature of the playtest (computer-aided, automatic data collection through surveys) also allows data to be analyzed and disseminated to the design team more rapidly. Usability tests and in-depth interviews often require substantial *post hoc* data organization in addition to interpretation, even when the *a priori* research questions are clear and well-conceived. In an industry where a hectic pace and tight development schedules often necessitate rapid decisions, speed and economy are particularly important.

Identifying Issues and Making Recommendations

As we said earlier, the broad goal of playtest data is to help identify issues with gameplay based on consumer feedback. For example, we ask specific questions about the “fun” of particular gameplay elements or mechanics (e.g., “How fun was combat?” or “How fun was Quest 1?”). In addition, we generally ask questions about elements that are less related to gameplay, like the music, graphics and sound effects. We can compare the results from the playtest of the current game to other games we have tested. However, it is not necessary to have a storehouse of data from other games, because the researcher can compare results from the current test to previous tests of the same game (iteration is key).

We sometimes offer recommendations for fixes to the issues that are identified in a playtest. Fully understanding the issue and why it occurred is essential to making good recommendations. The ratings or numbers from the playtest help illuminate potential issues, and the open-ended responses can help provide additional context. The recommended fixes are often based on the manner in which other games have successfully implemented a feature or based on our observations of watching players struggle with similar issues in other games.

The key value of the playtest, however, lies in identifying problem areas. We think of playtests as a tool to help designers understand an issue, so they can think up a workable solution. Importantly, if a designer makes substantial changes to the game based on playtest results, we can iterate to ensure the changes were effective in improving the game. Iterative testing is important to show that changes actually lead to improvement!

User Research in Games: A Case Study of Brute Force

We have described some of the methods that have been used to collect information from consumers about their experiences playing games, and have focused primarily on the playtest method. In the following paragraphs, we describe how usability and playtest methods have been used to identify potential design problems and ways to ameliorate them. In the Xbox game *Brute Force*, a squad combat based game, the design team made extensive use of feedback from users at several points in the design process.

Usability Testing and Brute Force

Being able to aim a weapon effectively is critical to succeed and have fun playing a combat game like *Brute Force*. Early in the development process for this game, the aiming/targeting mechanism consisted of two parts. First, the crosshairs in the aiming reticle would expand or contract to give the player feedback about how accurate their shots were likely to be. Wider crosshairs indicated less accuracy; narrower crosshairs indicated increased accuracy. The crosshair’s size increased as the player moved, feedback that was intended to tell the player that their shots would be less accurate. Second, the aiming mechanism allowed the player to “lock on” to a target when the aiming reticle was centered on the target for a certain period of time, allowing a more accurate shot. If the player kept the aiming reticle centered

on the target for the predetermined amount of time, a circle closed around the reticle. When the circle was complete, the player could move around but remained “locked on” to the target for more accurate shots.

To gather early usability information about the targeting features in *Brute Force*, the design team created a prototype of an area of the game that allowed users to move around and target enemies in a controlled environment. The results showed that users were confused about why they were unable to hit the enemies while they were running around, confusion that resulted in player deaths and frustration. Participants did not notice how the size of the reticle changed as they moved to reflect the effects of their movements on the accuracy of their shots. The circular lock mechanism was also confusing for players, who thought they were locked on to a target as soon as the circle closed (which was not the case). No participants were able to use the target lock feature, and player deaths were frequent.

The usability results suggested that people did not understand how the aiming reticle worked and how their behavior affected the reticle and their accuracy. However, they spent a lot of time and effort trying to figure it out (at the expense of other interesting gameplay mechanics). The movement of the reticle—both the crosshairs and the circle lock mechanism—were particularly problematic. As a result of this usability feedback, the design team implemented changes to the targeting features. In particular, they removed much of the movement feedback that had previously accompanied aiming, so the reticle became largely static. The lock on mechanism was removed completely, which allowed greater control of the character by the player, at the expense of ease of targeting. Subsequent usability tests found that these changes largely solved the aiming problems (participants were able to aim and shoot effectively) and consumers in later playtests reported that aiming, and combat in general, was fun.

Playtesting and Brute Force

The playtest method was also instrumental in identifying gameplay issues in *Brute Force*, and in gauging the effectiveness of design solutions that were implemented as remedies. For example, toward the end of the game’s production, it was hypothesized that some of the more interesting aspects of *Brute Force* were not being discovered quickly enough by players. The design team wanted players to learn about certain features more quickly, and they needed data to know if mechanisms they had designed in the game to teach players were effective. Furthermore, they needed data from consumers from the game’s target audience, a group with a widely varying skill-set and experience. A series of playtests was designed to help address this potential problem.

The first playtest was designed to assess the initial, out of the box experience of *Brute Force* by consumers. The game’s “missions”, which are scenarios in the game that players must complete to move on in the game (e.g., rescuing a prisoner), were presented in an order representative of the final version of the game, and participants played the game from the beginning. Consumers’ assessments of the game showed that the second mission was taking participants much more time to finish than the designers intended. In addition, participants felt that the mission was too long, too linear, and too restricting.

Based on the data from this playtest, the design team decided to shorten the second mission, to bring later missions (which were hypothesized to be more fun) closer to the beginning of the game. A second playtest was then conducted to assess whether the new mission ordering was indeed more fun than the previous one. By comparing consumer assessments from the second playtest to those obtained in the first playtest, the team confirmed that the new mission ordering was more fun. Players experienced more of the “fun stuff” earlier in play, and they learned more of the important skills that were necessary to successfully complete more difficult missions. The team was therefore able to ensure that players’ initial experiences in the game were fun and engaging.

Conclusion

We have described the playtest method, which is a way to collect information from consumers about how they perceive a game’s elements. The main goal in using this method is to give game designers quantitative evaluations from players about how they experience the game. The method combines surveys with gameplay in a controlled environment, and we have found it to be a valuable way to identify gameplay issues and gain insight into players’ perceptions.

One of the ways to use the playtest method described in this paper is to focus on the first hour of play or the initial experience, an interval we feel is critical to “get right” in the design of a game. Many players who do not like a game almost immediately will quit playing. But variations of the method can also elucidate specific issues in games, many of which take place after the first hour of play. For instance, players can be given a short tutorial, asked to play missions that occur later in the game, and then asked to rate their perceptions of those missions. Another example of a more interesting use of a playtest may be to compare different player control schemes on players’ perceptions of the game. Controls have a large influence on whether or not players enjoy a game, so playtest data can be invaluable to help designers choose the optimal scheme. Portions of games that are in development can be tested over time to evaluate whether changes to various aspects of the game have had a positive impact on consumers’ perceptions.

We also use playtest data to help with game tuning. For example, consumer data can help when designers tweak the handling of a car in a racing game or the difficulty of AI enemies in a shooter.

A limitation to the playtest method is that it loses some of the richness of feedback that is captured in well-designed usability studies. In a playtest, participants do not interact with an experimenter and do not provide insight into their experiences as they play as they do in a usability test. Unlike usability sessions, playtests sessions are not recorded and playtest moderators generally do not interact with participants except to give them standard instructions. Consequently, there are few opportunities to “drill down” on important experiences to figure out why players rated the game as they did. We are currently working on ways to address these concerns and gather richer information within the playtest framework. In addition, we are also experimenting with several additional methods, such as play diaries, that move beyond the laboratory to capture perceptions of the elements of deeper gameplay as players advance in a game.

Game designers can reap the benefits of better feedback from consumers about their games, but gathering such information in this domain is challenging. We have combined several methods into a tool that goes some distance toward achieving the goal of obtaining better information about consumers' perceptions of games. We hope that other researchers expand on the methods described here and apply them to additional challenges in game design.

ACKNOWLEDGMENTS

We would like to thank the Games User Research Group at Microsoft Game Studios (MGS) for their support and feedback, and particularly Bill Fulton and Ramon Romero for their pioneering work on the playtest method. We also thank Tim Fields for his feedback on the interaction between user research feedback and the design of *Brute Force*.

REFERENCES

1. Essential facts about the computer and video game industry. From the *Annual Report of the Interactive Digital Software Association* (2005).
2. Fulton, B. & Medlock, M. (2003). Beyond focus groups: Getting more useful feedback from consumers. Game Developer's Conference 2003 Proceedings, San Jose CA, March 2003.
3. Nielsen, J. (1993). *Usability Engineering*. New York: Morgan Kaufmann.
4. Pagulayan, R. J., Keeker, K., Wixon, D., Romero, R. L., & Fuller, T. (2003). User-centered design in games. In J. A. Jacko & A. Sears (Eds.), *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications* (pp. 883-906). Mahwah, NJ: Lawrence Erlbaum Associates.
5. Pagulayan, R. J., Steury, K. R. Fulton, B. & Romero, R. L. (2003). Designing for fun: User-testing case studies. In M. A. Blythe, K. Overbeeke, A. F. Monk & P. C. Wright (Eds.), *Funology: From Usability to Enjoyment* (pp. 137-150). Boston, MA: Kluwer Academic Publishers.
6. Pagulayan, R. J. & Fields, T. (2003). Design and user research: Dangerous alone, deadly together. Working Paper, Microsoft Game Studios.
7. Fowler, F. J. (1995). *Improving Survey Questions: Design and Evaluation*. Thousand Oaks, CA: Sage.
8. Loftus, G. R. & Loftus, E. F. (1980). *Mind at Play: The Psychology of Video Games*. New York, N.Y.: Basic Books.
9. Clanton, C. (1998). An interpreted demonstration of computer game design. In *Proceedings of CHI '98* (Los Angeles, CA, April 1998), ACM Press.
10. Malone, T. W. (1982). Heuristics for designing enjoyable user interfaces: Lessons from computer games. In *Proceedings of the 1982 Conference on Human Factors in Computing Systems* (Gathersburg, MD), ACM Press.
11. Crawford, C. (1982). *The Art of Computer Game Design*. Berkeley, CA: Osborne/McGraw-Hill.
12. Fabricatore, C., Nussbaum, M., and Rosas, R. (2002). Playability in action videogames: A qualitative design model. *Human-Computer Interaction*, 17, 311-368.
13. Kim, J., Choi, D. and Kim, H. (1999). Toward the construction of fun computer games: Differences in the views of developers and players. *Personal Technologies*, 3, 1-13.
14. Lazzaro, N. (2004). Why we play games: Four keys to more emotion without story. XEODesign, Inc. (http://xeodesign.com/xeodesign_whyweplaygames.pdf)
15. Salen, K. and Zimmerman, E. (2004). *Rules of Play: Game Design Fundamentals*. Cambridge, MA: The MIT Press.