Impacts of Forced Serious Game Play on Vulnerable Subgroups

*Carrie Heeter
Michigan State University
carrie.heeter@gmail.com

Yu-Hao Lee
Michigan State University
minke33@gmail.com

Brian Magerko
Georgia Institute of Technology
brian.magerko@lcc.gatech.edu

Ben Medler
Georgia Institute of Technology
benmedler@gatech.edu

International Journal of Gaming and Computer-Mediated Simulations (IJGCMS), 3(3), 34-53. doi:10.4018/jgcms.2011070103

ABSTRACT

Three vulnerable subgroups of players (non-gamers, resistant players, and females) were studied to understand how each approaches and plays serious games. We explore forced (required) play using four different online casual games. Our research strongly suggests that the most important threat to a serious game's impact is when players dislike the game. Serious games are likely to be least effective for players who dislike a game and most effective for those who like the game. Different people like and dislike different games, so simply advising designers to make a great game does not eradicate the problem. Additionally, non-gamers were at a distinct disadvantage as far as gameplay performance. They experienced more negative affect in two of the four games. Finally, males tended to seek more difficult challenges in games than females did. The optimal amount of challenge may be the most important gender difference to consider when designing serious games.

Keywords:

game analytics, gender , individual differences, non-gamers, serious games.

Acknowledgements:

Michigan State University graduate students Rosalie Blank and Younghwa Yun assisted with the literature review and online data collection. Jennifer Mangels, Associate Professor of Psychology at Baruch College & Graduate Center, City University of New York, served as a consultant. Hero Interactive LCC, FiLAMENT Games, iCivics, and the Michigan State University GEL Lab (games for entertainment and learning) provided the games in this study. This research is partially supported by grant #0943064 from the National Science Foundation. The opinions expressed are those of the authors and do not represent Michigan State University, Georgia Institute of Technology, or the National Science Foundation.

INTRODUCTION

Computer games for entertainment are downloaded or purchased from stores, played online, or borrowed from friends as part of a free market culture of choice. Game players decide what games they want to play, and when or where they will play them. In other words, playing games for fun is a highly voluntary and selective experience. Games for learning, on the other hand, can be quite the opposite. There are informal learning games mainly distributed on the internet, which players can voluntarily choose to play. However, many other serious games are required components of a school or training curriculum. These serious games are played within a military training context, as part of a high school curriculum, or used in corporate training. There are also games for health prescribed to maintain and improve cognitive health. Some of these games are mandatory learning experiences that are equivalent to assigned lab experiments, interactive training videos, simulation exercises, etc. These games are assigned in order to obtain desired serious outcomes other than just for fun.

When games are assigned rather than self-selected by the players, serious games face a much less enthusiastic, more diverse player audience than entertainment games due to the lack of the players' ability to choose which game to play. This often includes players who are disadvantaged because of the game format. A serious game's audience may include those who rarely play any kind of game (i.e. inexperienced "non-gamers") and those who dislike and normally avoid playing the genre used by that particular serious game (i.e. genre inexperienced and/or "resistant players"). The ramifications of this are obvious, though surprisingly overlooked in the digital game-based learning community at present: inexperienced players face an extra barrier of figuring out how to play, and resistant players miss out on the intrinsic motivation games afford to eager players. While certain games may be fun for many people (e.g., the bestselling *Civilization* series which is used for education; Squire, 2005), they are not "fun," "engaging," and "motivating" for everyone. Even the most wonderful serious games will undoubtedly fail to delight all members of a class or other target audience. This would not be an important concern if the only consequence was lack of fun or if this were in a culture where players could choose which educational game they will use in a formal learning environment.

However, being a non-gamer or resistant player may present more serious disadvantages in formal digital game-based learning situations.

We have previously written about hypothetical ways that vulnerable subgroups of players might face disadvantages compared to other players when assigned to play a serious game (Magerko, Heeter, & Medler, 2010). In this study we use large-scale survey and gameplay data to examine our hypotheses about how three vulnerable subgroups of players (non-gamers, resistant players, and females) approach and play serious games. The subsequent findings from this study are intended to help inform serious game designers about the population who may play their games and to potentially enable the intelligent adaptation of games to better fit individual learners.

THREE VULNERABLE SUBGROUPS

Non-gamers

Non-gamers are people who have little or no digital gaming experience. Unfamiliarity with gaming in general or with a particular game genre can present barriers to achieving learning goals. When a serious game is assigned for learning, a player must effectively master how to play the game in order to experience the desired learning content. From the perspective of cognitive load, more mental attention devoted to figuring out how to play means less cognitive attention available to devote to learning the intended material (Mayer, 2005a, 2005b; Sweller, in press). Non-gamers need to exert more effort figuring out how to play most games than experienced gamers, making it harder for non-gamers to benefit from serious games. Furthermore, feeling lost and incompetent while trying to play a game introduces negative thoughts that can create performance deficits by diverting cognitive load (Cardinu, Maas, Rosabianca, & Kiesner, 2005; Croizet et al., 2004), resulting in negative consequences for learning (Covington, Omelich, & Schwarzer, 1986; Thomas et al., 2006). For example, when a non-gamer participant in our study (explained in detail later) was asked why she quit early without playing for the assigned 10 minutes, the participant expressed both frustration over learning to play a game, and also negative emotions: "I did not know what to do. I hate playing video/computer games." This kind of response is obviously undesirable within a learning context.

Resistant Players

Even players with extensive gaming experience may dislike playing particular serious games because they find the games uninteresting or even unpleasant. In fact, people who are avid gamers may be more apt to dislike a serious game than are non-gamers if the game violates their expectations of entertaining game play and high end commercial quality production values. Regardless of their interest in the subject matter or the positive outcomes that the game is designed to evoke. In other words, using serious games as a one-size-fits-all players approach ignores the reality that some players will be unmotivated. One resistant player from our study explained why he quit playing: “It was boring and my time was up. I'm not really into games like that.”

Female Players

Gender intersects with gaming in complex ways. Although both females and males play games, males devote more time to gaming. Winn and Heeter (2009) looked at gender and gameplay behavior over time, confirming that males played significantly more than females in middle school, high school, and college. Males played an average of 266 more hours per year in middle school, 305 more hours per year in high school, and 225 more hours per year in college (p. 11). According to a recent Pew Foundation study of teens and gaming (Lenhart, Kahne, Middaugh, Evans, & Vitek. 2008), 97% of American teens play games. Boys play for more time and they play more different genres of games than girls do (an average of eight different genres compared to an average of six genres). Boys play more action, strategy, sports, adventure, first-person shooter, fighting, role play, survival-horror, and multiplayer games. Girls play more puzzle games. These gender differences in genre preferences (and by extension, genre experience) have implications for creating universally effective serious games. Some female participants in our study expressed genre preferences such as “I'm not into war-esque games”, “I prefer story aspects or online multiplayer with things like *Call of Duty*,” and also unfamiliarity with games, for example, “I did not know what was going on. I don't like playing video games at all.”

FOUR GAMES FOR STUDYING VULNERABLE SUB-GROUPS

With the three vulnerable subgroups in mind, our present study explores forced play situations using four different online casual games. These games were chosen because they represented a variety of genre and design intentions. They were also chosen for research purposes: they could be played for 10 minutes, were built in Flash, were from different designers, had different design intentions (i.e. informal education, cognitive training, pure entertainment, etc.) and the source code was accessible to us for implementing measurement protocols.

Game 1, *Photoflaw*, is a brain game built by the Michigan State University GEL Lab to exercise visual-spatial cognitive functions and attention. Game 2, *Do I Have a Right*, is a game for learning to teach about the Supreme Court and constitutional amendments, created by Filament Games. Game 3, *Starshine*, is a commercial puzzle game by Hero Interactive that could be used as a brain game to exercise visual-spatial reasoning and executive function, but the primary design goal was entertainment. Game 4, *Stormwinds*, also by Hero Interactive, is a commercial “tower defense” game that would not be used in a classroom or prescribed as a brain game. We included *Stormwinds* because games for learning often borrow from existing genres. The tower defense genre is highly gendered; males are more likely to choose this kind of game. Hence, observing what happens when everyone is forced to play has implications for serious games that might be based similar shooter/fighting entertainment genres.

***Figure 1*. Photoflaw: Michigan State University GEL Lab's hidden flaws game in which the player attempts to find subtle differences between two photos. (http://gel.msu.edu)**

***Figure 2. Do I Have a Right: Run a law firm of lawyers who specialize in constitutional law. Decide whether potential clients "have a right," and if so, match them with the right lawyer, created by iCivics and FiLAMENT Games.* (http://icivics.org)**

Figure 3 Starshine: Hero Interactive's trajectory shooting, puzzle game where players attempt to hit all of the stars located on the screen with a single shot that causes a chain reaction as stars explode from getting hit. (http://herointeractive.com.)

*Figure 4. **Stormwinds: Hero Interactive's blend of arcade shooter and tower defense game. Players purchase weapons they can use to shoot enemies that appear during each round of play.** (http://herointeractive.com.)*

TWO HYPOTHESES AND ONE RESEARCH QUESTION

The primary goal of this research is to examine forced play experiences and outcomes within and across our four games. The independent variables in this research are each player's gaming experience, gender, and whether each game was one the participant might choose to play even if not assigned to do so. The four dependent variables for this study were selected to have meaningful implications for serious game effectiveness and to be as directly comparable across the four games as possible, the four dependent variables in this study are attention and affect, commitment, performance, and challenge.

Dependent Variables

Attention and Affect How much of players' attention is focused on the game potentially mitigate whether or not a serious game has the desired impact (such as learning or cognitive exercise). Likewise, experiencing positive and negative affect during play may mitigate impact. Although our research did not directly measure whether the intended serious impact of gameplay was achieved, we can infer ways that differential affect and attention might impact desired outcomes of play. Players who pay strong attention to the game, those who feel more positive affect, and those who feel less negative affect are more likely to experience the outcomes serious game designers intended. Positive affect may enhance learning. Negative affect may interfere with learning. Paying close attention may enhance learning.

Commitment Players who are less committed and quit the game before completing the required amount of play are less likely to experience the intended impact of a serious game. For required serious games to have their intended impact, a player's commitment to play for the prescribed duration or longer certainly matters a great deal. When people assigned to play for 10 minutes stop playing long before the assigned duration, the likelihood they will receive the intended benefit from playing is low. In fact, based on data of how long players' browser windows were open to the game web page, compliance with the 10 minute "assignment" was far from perfect: 23% of *Starshine* players played for less than five minutes, as did 18% for *Do I Have a Right*, 12% for *Photoflaw*, and 12% for *Stormwinds*.

Performance Players who perform well in a game are more likely to acquire the intended impacts from serious game.

Challenge Tuning optimal levels of challenge is a big deal for game design in general, and is likely very important for serious games. Game designers often discuss Csikszentmihalyi's (1991) idea of *flow* (Salen & Zimmerman, 2004) – just the right amount of challenge relative to the player's ability. Too much challenge leads to frustration. Too little challenge leads to boredom. Flow presents a moving target. As skill increases, the optimal amount of challenge goes up.

Educators also seek optimal levels of challenge. Vygotsky's *zone of proximal development* is "the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with a more capable peer" (Vygotsky, 1978, p.86). This point to the area just at the edge of a student's current knowledge, new knowledge that is exactly challenging enough that he or she is able to learn with guidance from more experienced peers or teachers (Vygotsky, 1978). In theory, cognitive games to enhance brain function fit the same prescription (Breznitz, 2010). Just the right level of challenge is expected to optimize cognitive benefits (Heeter, Winn, Winn, & Bozoki, 2008).

Optimal levels of challenge (not too hard, not too easy) are likely to be associated with the experience with flow during gameplay and with learning. Players who feel the game is pleasantly challenging are more likely to acquire the intended impacts of a serious game.

Two of our games (*Stormwinds* and *Starshine*) appeared to participants to be difficult to figure out how to play; between 25% and 28% chose that response category. The other two games (*Photoflaw* and *Do I Have a Right*) did not appear to be as hard to figure out; only 6% to 9% of participants felt they were hard to figure out. Nearly half (47%) of people who played the two "hard to figure out" games also said that they were challenging to play. Even though *Photoflaw* was easy to figure out, 44% of *Photoflaw* players said the game was challenging to play. *Do I Have Right* was neither hard nor challenging (only 20% of players found it challenging). And yet, *Do I Have a Right* was the most popular of the four games. (See Figure 5)

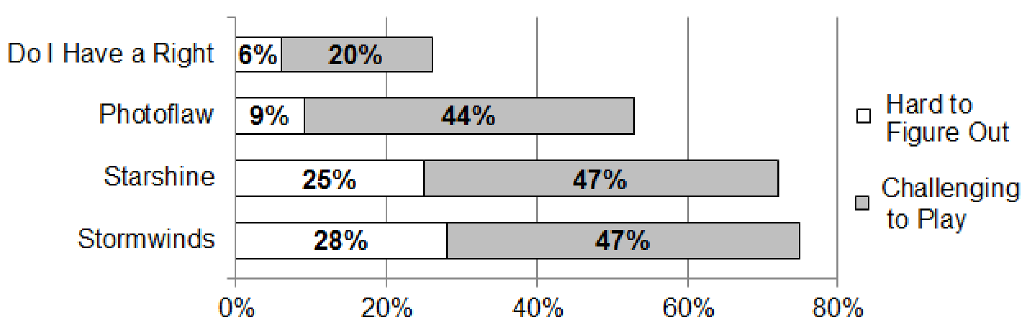

Figure 5. Difficulty of figuring out and playing game.

Hypotheses and Research Question

In order to examine how the three vulnerable subgroup of players may be disadvantaged in the four aspects (attention and affect, commitment, performance, challenge) that may affect intended outcomes of serious game, the hypotheses and research question for this study are:

H1: Non-gamers who rarely or never play games are less likely to experiences the benefits of playing a serious game compared with moderate and avid gamers. Specifically, non-gamers will experience less positive and more negative affect, pay less attention while playing, choose easier challenges and experience the game as more challenging, show lesser commitment to play, and perform worse than will avid and moderate gamers.

H2: Resistant players who do not like a serious game and would not play it if not assigned to do so are less likely to experiences the benefits of playing that game compared with players who like the game. Specifically, resistant players will experience less positive and more negative affect, pay less attention while playing, show lesser commitment to play, and perform worse than will eager players who like the game.

RQ1: Does a player's gender influence the likelihood of experiencing the benefits of playing a serious game? Specifically, how does gender intersect with affect, commitment, performance, and challenge?

METHODS

Research Design

We recruited 330 undergraduates interested in obtaining extra credit in four large telecommunication or history classes at a large Midwestern university for the research. Participants completed an initial survey, were randomly assigned to play two out of four possible games, we collected their gameplay behavior data directly from the game servers. Participants completed a post-play survey immediately after each of the two games. Figure 6 diagrams the research design. As a result, we collected actual play data from 152 *Stormwinds* players, 145 *Starshine* players, 146 *Photoflaw* players, and 112 *Do I Have a Right* players. The comparatively lower number of *Do I Have a Right* players was due to periodic failure to load that game.

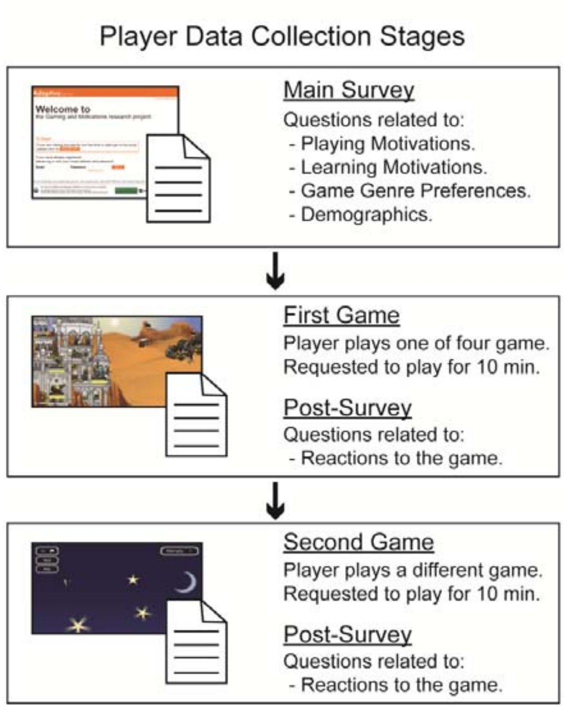

Figure 6. Research procedure.

Operationalizing the Vulnerable Subgroups (Independent Variables)

The independent variables in this study are (a) eagerness or resistance to play (player liking of the game), (b) non-gamers, moderate, or avid gamers (how much time per week participants normally spend gaming), and (c) gender (male or female). Each construct is described below, including descriptive sample statistics.

Eager and Resistant Players: A player's eagerness or resistance to play each particular game was measured using post-game survey questions that asked respondents whether they would play the game they had just played if they were not assigned to do so. Figure 7 below shows the proportion of participants who said they would play that game even if play was not assigned, the proportion that might play, and the proportion that would not play for each of the four study games. Those who would play anyway ("eager" players) ranged from a low of 29% for *Starshine* to a high of 49% for *Do I Have a Right*. Those who would not play ("resistant" players) were the plurality group for *Stormwinds* (51%) and *Starshine* (47%), while the proportion of players who would definitely not play was lower for *Photoflaw* (35%) and lowest for *Do I Have a Right* (27%). None of the games was universally disliked nor was any universally appealing; each had resistant and eager players.

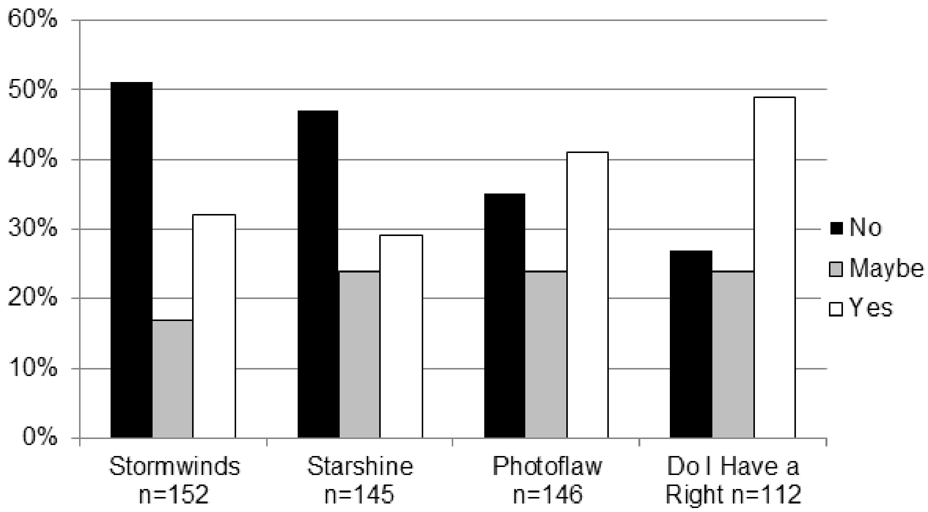

Figure 7. Whether participants would, might, or would not play game if not assigned to do so.

Non-gamer, Moderate and Avid Gamers: We categorized participants into three gamer categories based on our pre-survey question which asked respondents how many hours they played digital games per week. Twenty-five percent of participants were Non-Gamers who played games for one hour or less per week; 40% were Moderate Gamers who played between two to five hours per week; and 34% were Avid Gamers who played for six or more hours per week.

Gender: Thirty-two percent of study participants were female.

Four Dependent Variables

The dependent variables in this study were (a) Affect, (b) Commitment, (c) Performance, and (d) Challenge. Positive and negative affect and attention to the game as well as how challenging game play felt were measured using post-game survey questions immediately after participants finished playing each game. Commitment, Performance and Challenge were aggregated constructs derived from actual gameplay data.

Analyzing collected player data has been a common research area for many years (see Nacke, Stellmach, Sasse, & Lindley, 2009, for a detailed review). Some researchers used player data for analyzing player demographics and figure out the gameplay habits of certain population groups (Williams, Yee, & Caplan, 2008). Canossa and Drachen (2009) used gameplay data to determine how players can be matched to specific personas based on their gameplay. Other researchers collect physiological data to determine a player's emotion responses to gameplay (Mandryk, Atkins, & Inkpen, 2006) or track the eye gazes of a player to establish which game world features the player fixates upon inside a 3D level (Nacke et al., 2009). Our research follows similar lines of reasoning: collected game play behavior can be analyzed to determine certain characteristics about those players.

Games offer players choices and enable players to take actions over time. Collecting data on each player choice throughout a game is quite easy. Constructing conceptually meaningful game play behavior measures is more difficult. Constructing measures that can be compared to parallel measures across completely different games is harder still. For a detailed discussion of this topic and numerous examples, see the forthcoming chapter by Heeter, Medler, Magerko, and Lee (in press).

Each specific player choice is idiosyncratic to a particular moment of play in a particular game, and therefore specific choices rarely serve to characterize gameplay in a conceptually meaningful or even useful way. Player choices make game play interesting but they complicate play style measurement. Some player choices have little impact on subsequent game play (such as customizing the physical appearance of an avatar). Other player choices overtly alter the difficulty of challenges that will be presented in the game (such as choosing easy, medium, or hard gameplay). Still other player choices covertly impact the nature of subsequent play. (For example, selecting the first amendment lawyer or the fourth amendment lawyer could change which legal cases a player can bring to court; choosing a cannon instead of a machine gun could change which enemies can be destroyed; choosing to reveal a hint makes solving a puzzle easier but reduces points earned for solving that puzzle.)

The actions a player can take are also idiosyncratic to each game and they vary throughout the game experience, based on choices and actions that have already unfolded. Different player choices and actions combine to construct a unique play experience each time that game is played. Often, games are designed to become more challenging in later levels. Actions and their outcomes are altered by this progression. To complicate measurement even more, players play at different speeds and they play for different amounts of time.

We began by playing all our games many times and brainstorming every possibly important variable that could be measured during play. This approach yielded 81 variables for *Stormwinds*, 15 for *Starshine*, 27 for *PhotoFlaw*, and 68 for *Do I have a right*. We revisited the measurement approach and rethought the variables, organizing them around eight conceptual areas (approaching the game, challenge, response to failure, attention to feedback, exploration, strategy, performance, and commitment). This yielded 102 variables for *Stormwinds*, 28 for *Starshine*, 25 for *PhotoFlaw*, and 65 for *Do I have a right*. In this article, we will analyze three gameplay variables for each game (commitment, performance, and challenge), carefully constructed to be conceptually meaningful and comparable across players within a game and across games.

Detailed descriptions of how each survey and gameplay construct was measured appear below.

Affect and Attention: Participants completed a short post-survey immediately after playing each game which included 10 items used by Rodriguez, Guerra, Higgins, & Mangels (2011) to measure emotional reaction to an experience, such as calm, focus, wellbeing, involvement. Seven of the ten items loaded onto three factors. An Attention scale combined “I was completely focused on this game” and “my attention was monopolized by this game” (Cronbach’s alpha ranged from .86 to .88 across the four games). The Positive Affect scale was constructed combining “I felt happy while playing,” “I felt good while playing,” and “I enjoyed playing this game” (Cronbach’s alpha ranged from .82 to .86 across the four games). The Negative Affect scale combined “I felt bad while playing” and “I felt sad while playing” (Cronbach’s alpha ranged from.76 to .87 across the four games). These items used a 1 to 9 scale where 1 equaled disagree strongly, 5 was neutral, and 9 equaled agree strongly.

Commitment: The study interface instructed participants to play each game for 10 minutes, but the research system did not enforce this request. Some played for much less than 10 minutes and a few played for much longer. This complicated other measurements, but provides a dramatic indicator of how committed each player was to each game. It also reflects what happens when serious game play is assigned: many only partially comply with the assignment.

Conceptually the participants' game time seems like the most straightforward measurement of commitment. Our data collection system did track how long participants' internet browsers were open to the game web page. But we did not use that variable because some duration seemed much longer than the other recorded gameplay variables would suggest the person actually spent playing. We surmise that some participants were not playing the game that whole time. They could have left the room with the browser open or been doing other things on their computer, which would introduce potentially large inaccuracies in the data. Therefore, commitment was measured in an alternative way most fitting for each game. In the case of *Stormwinds*, commitment was operationalized as total levels attempted; this ranged from 0 to 36 (*m*=4.7, *s.d.*=5.2). Players had to have explored enough to select a campaign to be included in the analysis; three percent selected a campaign but did not go so far as to actually start playing and hence had zero levels attempted. For *Starshine*, commitment was also operationalized as total levels attempted. This ranged from 1 to 29 (*m*=11.9, *s.d.*=6.8). In *Photoflaw* commitment was measured as total photos attempted, which ranged from 1 to 69 (*m*=6.9, *s.d.*=6.5). *Do I Have a Right* participants' varying commitment was evident in the number of in-game "days" they played, which ranged from a low of 1to a high of 43 days (*m*=5.6, *s.d.*=6.0).

Performance: For the performance game play constructs, we needed to calculate variables that could be compared across players even though they had played for different amounts of time. One approach could be to consider performance only in the first round or level or day of play. A problem with that approach is that first levels are usually especially easy in order to help players understand game mechanics. An alternative approach is to calculate performance scores as a percent of attempts that succeeded or average number of successful attempts. The problem with percents and averages is that challenge in games is typically designed to be progressive, growing more difficulty over time. Therefore, players who played the longest might have lower average success, yet actually be playing better than those who quit the game after a single round after earning perfect scores. A research design that enforced rather than requested 10 minutes of play would have offered more flexibility in constructing comparable variables.

Even when 10 minutes of play is mandatory, player selection of difficulty level makes relative performance comparisons less than ideal. For example, *Photoflaw* scores (photos successfully solved, number of hints requested, and mistakes made) are confounded by the influence of player-selected difficulty. It is easier to earn a perfect score at the easy level. Similarly, comparisons of performance scores in *Stormwinds* and *Starshine* are confounded by difficulty and level.

Performance measures were constructed to be as independent of duration of play as possible. Conceptually, our performance variables represent how well the game is telling the player she is doing. They are absolute measures. They correspond exactly to the feedback the player can see about how successful he was at accomplishing the goals of the game. Our study addresses the conflagration of performance and challenge by examining them separately. Performance is conceptualized as how well the player is doing in the game (regardless of difficulty level). For *Stormwinds*, performance was computed as percent of levels attempted that that the player won. The range was from a low of 0% to a high of 100% (*m*=41%, *s.d.*=37%). For *Starshine*, where players are challenged to hit every star on the screen from a perfectly aimed single shot, performance was the percent of shots taken that successfully solved a puzzle. The range was from a low of 3% to a high of 100% (*m*=12%, *s.d.*=13%). For *Photoflaw*, we used the average number of mistakes as a reverse measurement of performance (i.e. fewer mistakes indicated better performance). The range was from a low of 0 to a high of 137 (*m*=12.9, *s.d.*=22.3). For *Do I Have a Right*, performance was the prestige point score at the end of the first day in the game. Day 1 prestige scores ranged from a low of 0 to a high of 425 points (*m*=294, *s.d.*=164). Unlike the other three games whose round 1 lasted less than one minute, one day in court for Do I Have a Right usually took two to three minutes to complete.

For three of the games, performing well in the game is directly tied to the intended impact of the game. *Do I Have a Right* is an educational game to teach about the U.S. Constitution Bill of Rights and the courts. Cases are won in the game by correctly judging whether a case does or does not fall under the auspices of particular constitutional amendments. Hence, high scores should parallel intended learning. Performance in *Photoflaw* reflects successful exercise of visual spatial attention. Although *Starshine* was not designed to be a cognitive game, it could conceivably be used to exercise visual-spatial cognition and executive function. Here too, performance in the game (in our study, percent of stars successfully targeted) parallels successful visual spatial reasoning and executive function. Since *Stormwinds* is purely an entertainment game, there are no intended serious impacts. Performance scores are suggestive of the potential influence of game genre if there were a *Stormwinds*-like learning game (for example, a game like *Math Blasters*).

Challenge: Two measures, perceived and actual, are used to analyze challenge for each game. Perceived challenge was assessed the same way for each game. On the post-game survey, respondents were asked “how challenging was the game for you?” using a five point scale from 1=Not Challenging at All to 5=Very Challenging.

Our four games incorporate actual choice of challenge in different ways. Two of them (*Photoflaw* and *Stormwinds*) overtly offer players the choice of easy, medium, or hard challenge before they begin to play. In *Photoflaw*, 23% of players consistently selected easy challenges, 14% chose a mix of easy and medium, 43% chose medium, 13% chose a mix of medium and hard, and 7% always chose hard challenges. *Photoflaw* challenges are different in the easy, medium, and hard difficulties, but photo difficulty does not become progressively more challenging as players continue to solve more photos. In *Stormwinds*, 69% of players only played easy campaigns; 16% only played hard campaigns, 5%

selected medium campaigns, and the remaining 10% tried more than one difficulty. Even though players choose a starting difficulty, each subsequent level of play is more challenging.

Starshine and *Do I Have a Right* did not offer player-selected difficulty. In *Starshine*, the beginning level presents two stars to shoot, and the number and type of stars and complexity of trajectory increase as play continues. Self challenge was operationalized as the highest level attempted. The range was from a low of 0 to a high of 29 (*m*=10.2, *s.d.*=6.9). This is unavoidably confounded with duration of play. (Players who play for longer are more able and more likely to attempt higher levels. Levels can be repeated as often as the player wishes.)

In *Do I Have a Right*, the player challenges herself by going after clients and bringing cases to trial. The amount of challenge in *Do I Have a Right* is covertly driven by player choices throughout the game. Players have control over how many cases they bring to trial. Each trial offers the chance to win or lose prestige points. Self challenge in Do I Have a Right was operationalized as average number of cases tried per day of play. The range was from a low of 0.5 to a high of 5.3 (*m*=1.9, *s.d.*=1.2).

FINDINGS

One-way ANOVA with post-hoc Tukey HSD test was used to compare Eager-Resistant Player and Gamer comparisons. T-tests were used to test the difference between gender and each dependent variable.

Hypothesis 1 Results (Non-Gamers)

Attention and Affect: Positive affect was not different by gamer status for any of the four games. In fact, three games showed no affect (*Starshine F*[2, 137]=.40, *n.s.*; *Photoflaw F*[2, 135]=1.55, *n.s.*; *Do I Have a Right F*[2, 112]=.70, *n.s.*) or attention (*Starshine F*[2,140]=.04, n.s.; *Photoflaw F*[2, 139]=.97, n.s.; *Do I Have a Right F*[2, 114]=1.92, n.s.) differences between gamers and non-gamers. For *Stormwinds*, non-gamers felt more negative affect while playing (*F*[2, 157]=3.97, *p*<.05). Thus, gaming frequency may sometimes be associated with negative affect in serious games, but this is highly dependent upon the nature of the serious game. Avid gamers who played *Stormwinds* also experienced more focused attention than did moderate gamers and non-gamers (*F*[2, 157]=4.62, *p*<.05).

Commitment: The frequency with which participants played games overall was not significantly related to gaming commitment for three of the four games (*Starshine F*[2, 141]=.81, *n.s.*; *Photoflaw F*[2, 143]=1.58, *n.s.*; *Do I Have a Right F*[2, 123]=2.13, *n.s.*). Post hoc Tukey HSD analysis showed that non-gamers played significantly fewer *Stormwinds* levels than did moderate or avid gamers (*F*[2, 160]=5.33, *p*<.01) and they played *Do I Have a Right* for significantly fewer days than did moderate gamers (p<.05). These results confirm that lower commitment to playing games is often found among non-gamers. Whether avid or moderate gamers played the most depended on the game. The two entertainment games showed avid games playing slightly more than moderate gamers. The brain game and learning game were both played slightly (though not significantly) more by moderate gamers. The relationship between gaming frequency and commitment to play assigned games is nonlinear and appears to be influenced by the nature of the game. (See Figure 8.)

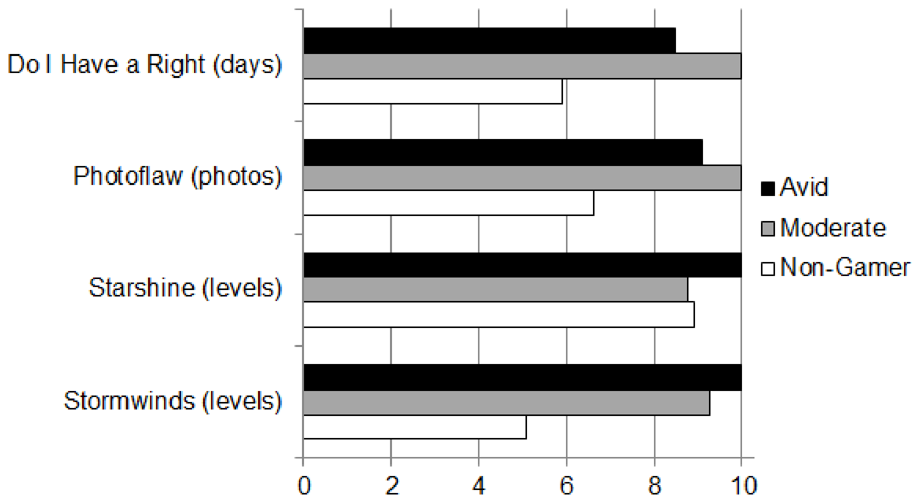

Figure 6. Average commitment to play by gaming frequency, normalized so that the highest average commitment in each game equals 10.

Performance: Being a non-gamer was related to significantly worse performance in two of four games (*Photoflaw* $F[2, 143]=2.58$, $p=.04$; Stormwinds $F[2, 160]=12.32$, $p<.001$). The exceptions were *Do I Have a Right* ($F[2, 123]=2.09$, *n.s.*) and *Starshine*, $F(2, 141)=2.07$, *n.s.*. Even there, the data conforms to the expected direction. (See Figure 9.)

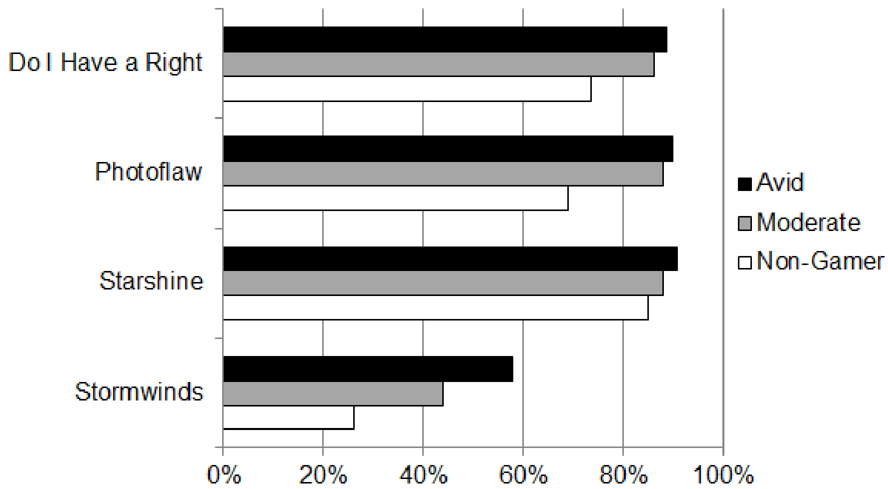

Figure 7. Average performance by gaming frequency, normalized to approximate how closely performance approached perfect play.

Challenge: No significant differences between gaming frequency were found regarding how challenging participants reported the four games. For three of the four games, self-selected difficulty was also not different by gamer status (*Starshine F*[2, 141]=1.63, *n.s.*; *Do I Have a Right F*[2, 123]=1.24, *n.s.*; *Stormwinds F*[2, 154]=1.53, *n.s.*). The one exception was *Photoflaw*, F(2, 143)=4.17, *p*<.05). Avid gamers chose significantly harder difficulty settings than did moderate gamers, who chose significantly harder difficulty settings than did non-gamers.

Summary: Hypothesis 1 is partially supported. Most importantly, non-gamers evidenced worse performance on three of four games. If performance is analogous to impact, then non-gamers are clearly a vulnerable subgroup. They also tended to have lower commitment to playing assigned games. There was some evidence of more negative affect for non-gamers on one of four games. But for the most part, positive affect, attention, perceived challenge, and self-selected challenge were not impacted by gaming frequency.

Hypothesis 2 Results (Resistant Players)

Attention and Affect: There were significant differences in attention level between resistant and eager players for three of the four games (*Photoflaw F*[2, 139]=.018, *p*<.05; *Do I Have a Right F*[2, 115]=20.59, *p*<.001; *Stormwinds F*[2, 148]=8.61, *p*<.001), the exception was *Starshine*, *F*(2, 141)=1.91, *n.s*. (See Figure 10.) Resistant players were least likely to say their attention was focused on the game. Eager players were most likely to say their attention was focused on the game. And those in between were in between. Clearly, attention suffered when players disliked a game.

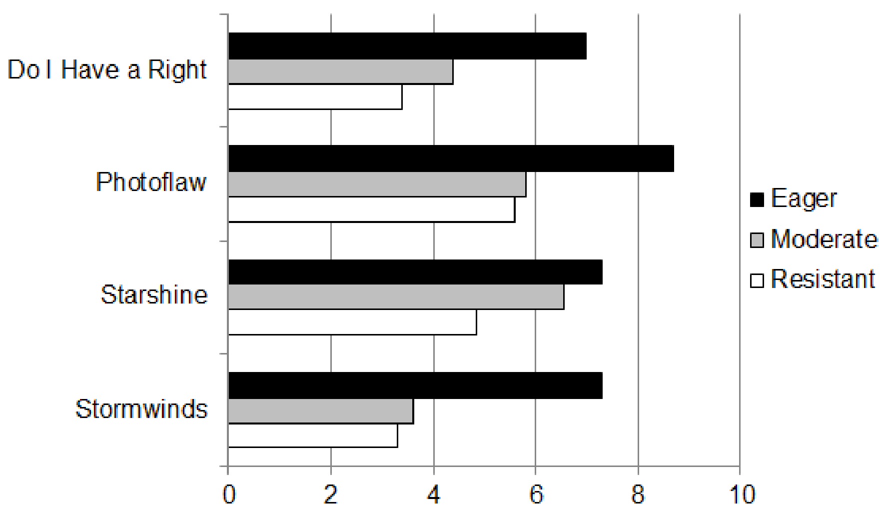

Figure 8. Average attention to each game by resistance to play (maximum = 9).

There were significant differences in positive affect between eager and resistant players (*Starshine F*[2, 138]=4.70, *p*<.05; P*hotoflaw F*[2, 135]=7.25, *p*<.001; *Do I Have a Right F*[2, 113]=17.52, *p*<.001; *Stormwinds F*[2, 149]=10.04, *p*<.001. Resistant players had the lowest levels of positive affect. Eager players had the highest level of positive affect, and those who might play anyway were in between. (See Figure 11.) There were no significant differences in negative affect with the sole exception of *Photoflaw* (*F*[2, 138]=7.94, *p*<.01). (See Figure 12.)

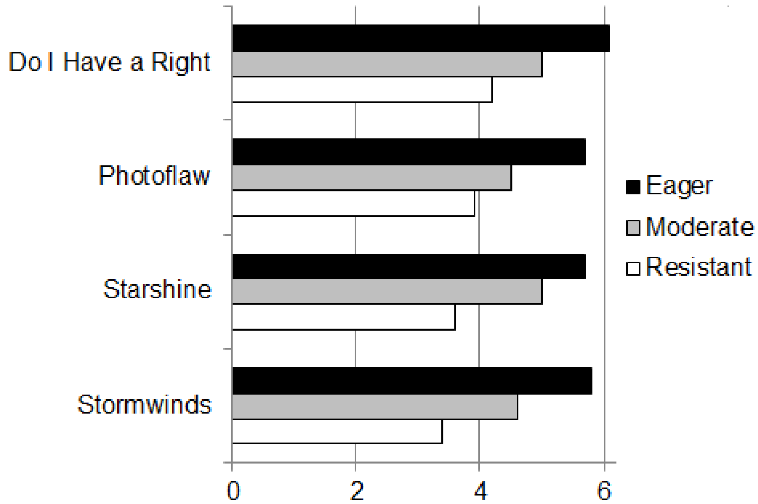

Figure 9. Average Positive Affect experienced in the game by resistance to play (maximum = 9).

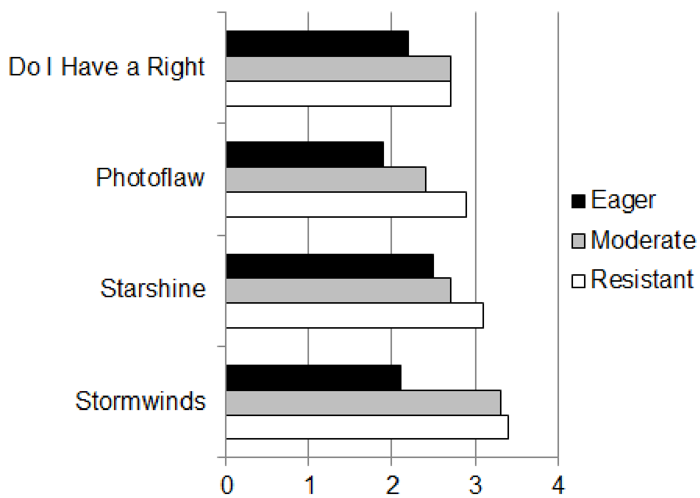

Figure 10. Average Negative Affect experienced in the game by resistance to play (maximum = 9).

In short, resistant players devoted significantly less attention, experienced significantly less positive affect and sometimes experienced more negative affect.

Commitment: Eagerness or resistance was significantly related to their commitment to the game across all four games, *Starshine* *F*(2, 142)=329.07, *p*<.01; *Photoflaw* *F*(2, 143)=31.37, *p*<.001; *Do I Have a Right* *F*(2, 123)=46.30, *p*<.001; *Stormwinds* *F*(2, 151)=79.92, *p*<.001. (See Figure 13.) Post-hoc Tukey HSD analysis showed that for three of the four games, the locus of the significant difference was that eager players played significantly more than those who might and those who would not play anyway. In the case of *Starshine*, resistant players played significantly less than moderate and eager players. In other words, resistant player were significantly less committed and are under the risk of not experiencing the intended impact.

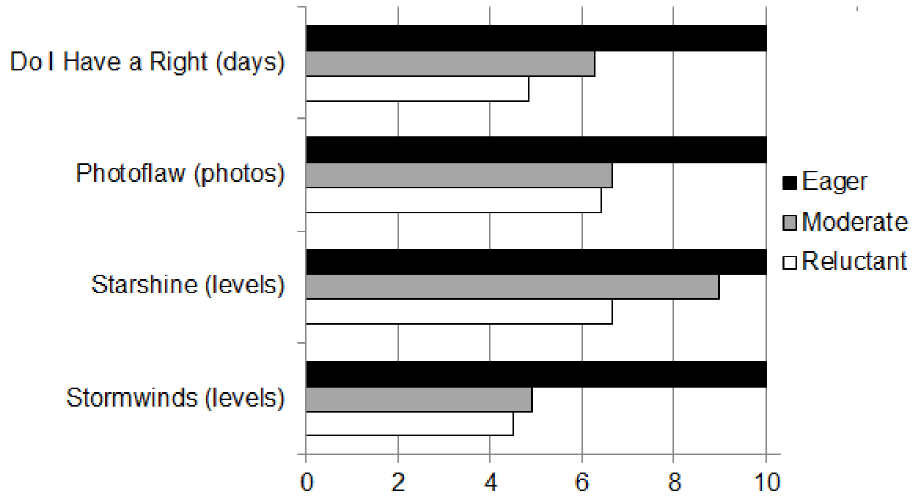

Figure 11. Average commitment by resistance to play, normalized so the highest average equals 10.

Performance: Performance was significantly lower among those who would not play the game anyway (resistant players) than for those who would definitely play anyway (eager players) across three of the four games. *Photoflaw* $F(2, 143)=5.42$, $p<.01$; *Do I Have a Right* $F(2, 123)=38.96$, $p<.001$; *Starshine* $F(2, 142)=3.40$, $p<.05$. The exception was *Stormwinds* ($F[2, 151]=.38$, *n.s.*). (See Figure 14.)

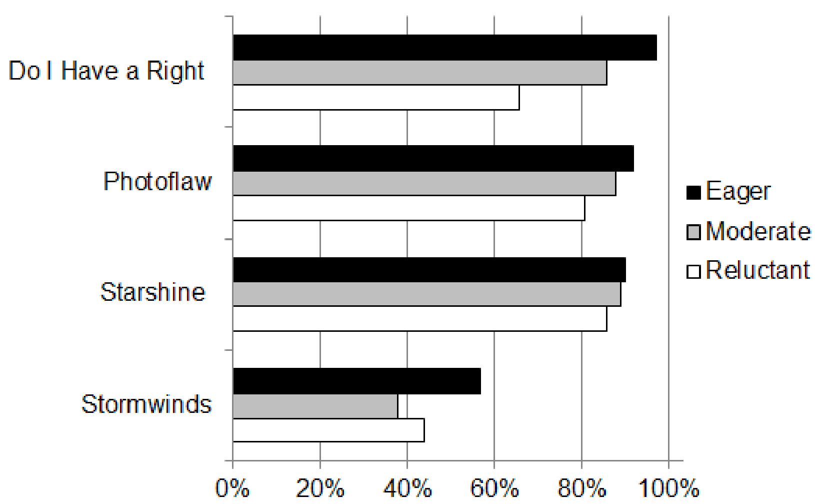

Figure 12. Average performance by resistance to play, normalized to approximate how closely performance approached perfect play.

Challenge: There was no significant difference between eager and resistant players in the level of self-reported difficulty. For *Starshine* and *Do I Have a Right* (the two games for which difficulty was not set at the beginning but instead was determined by player actions throughout the game), gameplay data shows that eager players pursued significantly more challenging play than did resistant players

(*Starshine F*[2, 142]=84.88, *p*<.001; *Do I Have a Right F*[2, 123]=35.28, *p*<.001). On the other hand, when presented with an overt choice of easy, medium, or hard play in *Stormwinds* and *Photoflaw*, there were no significant eager-resistant differences (*Stormwinds F*[2, 151]=.92, n.s.; *Photoflaw F*[2, 142]=.06, n.s).

Research Question 1 Results (Gender)

Attention and Affect: Of the four games, *Stormwinds* was the closest to a traditional, male-oriented genre. Females experienced less positive affect (*t*[159]=-3.07, *p*<.01)and less focused attention (*t*[159]=-2.91, *p*<.01) when assigned to play *Stormwinds*, most likely because of a lower enjoyment of violence in games. Surprisingly, females did not report more negative affect in *Stormwinds* (*t*[159]=.96, *n.s.*).

Commitment: Gender was significantly related to commitment for three of four games, but showed opposite results between games, suggesting that gender difference in commitment is due to preference rather than familiarity with gaming. *Stormwinds* and *Starshine* both showed stronger commitment among males (*Stormwinds t*[161]=2.38, *p*<.05; *Starshine t*[142]=2.31, *p*<.05). (See Figure 15.) These two games also involved shooting, in the case of *Stormwinds*, shooting to destroy enemies, and in the case of *Starshine*, shooting so that the trajectory intersected will all of the star "targets." It is possible (though not proven from our analysis) that shooting as a game mechanic tilted the games toward male players, even when the shooting was essentially non-violent. *Do I Have a Right*, the educational courtroom game evoked more female than male player commitment (*t*[124]=-2.07, p<.05). Entertainment game genre preferences have been shown to be highly gendered (Lenhart et al., 2008). Therefore, it is not surprising that individual games also appeal more to one or the other sex.

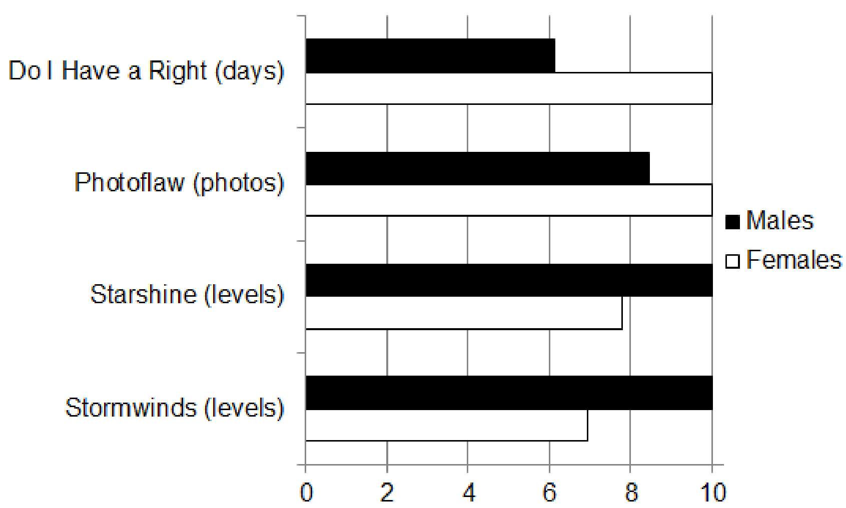

Figure 13. Average commitment by gender, normalized so the highest average equals 10.

Performance: Gender was not related to performance in any of the games.

Challenge: For both games in which players selected a difficulty level at the start of gameplay (*Stormwinds* and *Photoflaw)*, males chose significantly harder difficulty (*Stormwinds* $t[155]=1.97$, $p<.05$; *Photoflaw* $t[143]=3.35$, $p<.01$). Males also pursued more challenging puzzles than females did in *Starshine*, $t(142)=3.55$, $p<.01$. Gender was unrelated to seeking challenging play for *Do I Have a Right*, $t(124)=1.48$, *n.s*. There were no gender differences in how challenging female and male participants rated any of the four games. Males sought out play that was more difficult. Despite playing easier challenges, females reported that they experienced the same level of challenge as males reported while playing.

DISCUSSION

Hypothesis 1 states that non-gamers will have lower commitment, affect and performance. This hypothesis was partially confirmed. Non-gamers were at a serious disadvantage as far as performance. Because getting the intended impact from a serious game depends upon playing well, non-gamers were mostly to be left behind. Non-gamers experienced similar amounts of positive affect but more negative affect in one of the four games, which might be expected to interfere with learning or cognitive benefits.

Hypothesis 2 states that resistant players will have lower commitment, affect, performance. Our data was consistent with this hypothesis. Our research strongly suggests that the most important threat to a serious game having its intended impact is when players dislike the game and would not play it on their own if not assigned to do so. Resistant players were less committed (play for less time), experienced less positive affect, less focused attention, and more negative affect than eager players who would play even if they did not have to. Resistant players performed worse and played in such a way that they experienced less challenge during the game. Serious games are likely to be least effective for players who dislike a game and most effective for those who like the game.

Declaring that serious games are less effective for players who dislike the game on the surface seems to imply that serious games simply need to be great games, and they will have the desired impacts. The problem with that solution is that liking a game is highly idiosyncratic. Every one of our games in the study had eager players and resistant people who would never play the game intentionally. Even if a game were a great example of its genre, the appeal of such a game would certainly not be universal. Different people like different games. Even the most wonderfully designed serious game will fail to appeal to some players. This also points to the need for more future work on understanding the relevant individual differences between game players, particularly for formal learning environments.

Hypothesis 1 and 2 are expressed in terms of vulnerable subgroups being potential disadvantaged when assigned to play a serious game. Based on our findings, one could also appreciate the inverse interpretation: serious games are a good way to reach gamers, and serious games are a potentially great way to reach players who really enjoy the game.

Our results showed that males tended to seek more difficult challenges in games than females did. On two of four games (the two harder games, both of which involved shooting), males were more committed (played for longer). On the learning game (which was also the easiest game), females showed significantly stronger commitment. Performance was rarely different by gender. Affect was only different for one of the games (*Stormwinds*). Incorporating ways to permit different players to experience their own optimal amount of challenge may be the most important design consideration for mixed gender serious games audiences.

Eager and resistant players reacted very differently to all four games, these differences occurred consistently, regardless of the design of the game. But how well a game attracts and supports non-gamers and both genders depends upon the design of the game. *Do I Have a Right* and *Starshine* seem to be well designed to accommodate both non-gamers and avid gamers. Although those two games showed significant differences by gender as far as commitment, for *Starshine* males played more and for *Do I Have a Right*, females played more.

Photoflaw favored moderate gamers, with non-gamers faring least well. *Stormwinds* favored avid gamers, who experienced more commitment, better performance, stronger attention, and less negative affect than non-gamers. *Stormwinds* also yielded less play by females, more negative affect, and less attention. Serious game designers should consider how their game will be received by non-gamers as well as avid gamers and females as well as males.

By conducting parallel analyses of four different games, this research strongly documents that resistant players and non-gamers face disadvantages when playing serious games. Our analyses go beyond simply comparing performance among three potentially vulnerable groups. We included three more dimensions of serious game play (player commitment, player affect, and challenge seeking) that psychology and learning theories imply would have an impact on learning. Our study used only a single measure of performance to represent potential intended benefits of each game. These findings should be replicated with particular serious games, using more elaborate and complete measures of performance/intended impact. Future research should consider how serious games might be designed or introduced to players to mitigate these kinds of predictable resistant player and non-gamer issues. Deeper understanding of subtle gender differences related to challenge in games could lead to design ideas to optimize gameplay for diverse challenge-seekers.

REFERENCE

Breznitz, S. (2010). Cogfit training: The key to cognitive vitality. Accessed July 13, 2010 from http://www.cognifit.com/science/cogtraining

Canossa, A., & Drachen, A. (2009). Patterns of play: Play-personas in user-centred game development. In proceedings of *DiGRA 2009*. London, UK.

Cardinu, M., Maas, A., Rosabianca, A., & Kiesner, J. (2005). Why do women underperform under stereotype threat? Evidence for the role of negative thinking? *Psychological Science, 16(7)*, 572-578.

Covington, M., Omelich, C., & Schwarzer, R. (1986). Anxiety, aspirations, and self-concept in the achievement process: A longitudinal model with latent variables. *Motivation and Emotion, 10*, 71-88.

Croizet J. C., Després G., Gauzins, M. E., Huguet, P., Leyens, J. P., & Méot, A. (2004). Stereotype threat undermines intellectual performance by triggering a disruptive mental load. *Personality and Social Psychology Bulletin*, *30(6)*, 721–731.

Csikszentmihalyi, M. (1991). *Flow: The psychology of optimal experience.* New York: Harper Collins.

Heeter, C., Medler, B., Magerko, B., & Lee, Y.H. (in press). Six of one, half a dozen of the other: Meaningful metrics for comparing player processes and outcomes within and across games. In El-Nasr, M., Drachen, A., Canossa, A., and Isbister, K., (Eds)., *Game telemetry and metrics: Maximizing the value of your data*, Routledge Press.

Heeter, C., Winn, B., Winn, J., & Bozoki, A. (2008). The challenge of challenge: avoiding and embracing difficulty in a memory game. *Meaningful Play Conference*, East Lansing, MI.

Lenhart, A., Kahne, J., Middaugh, E. Evans, C., & Vitek, J. (2008). Teens' gaming experiences are diverse and include significant social interaction and civic engagement. Retrieved April 2, 2009, from Pew Internet and American Life Project: http://www.pewinternet.org/~/media//Files/Reports/2008/PIP_Teens_Games_and_Civics_Report_FINAL.pdf.pdf

Magerko, B., Heeter, C., & Medler, B. (2010). Tapping into the hidden potential of serious games: Accommodating individual differences. In R. Van Eck, (Ed.), *Interdisciplinary models and tools for serious games: Emerging concepts and future directions.* IGI Global.

Mandryk, R.L., Atkins, M.S., & Inkpen, K.M. (2006). A continuous and objective evaluation of emotional experience with interactive play environments. In Proceedings of *CHI 2006*. 1027-1036. Montréal, Québec, Canada.

Mayer, R. E. (2005a). Principles for managing essential processing in multimedia learning: Segmenting, pretraining, and modality principles. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning.* (pp. 169-182). New York: Cambridge University Press.

Mayer, R. E. (2005b). Principles for reducing extraneous processing in multimedia learning: Coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning*. (pp. 183-200). New York: Cambridge University Press.

Nacke, L., Drachen, A., Kuikkaniemi, K., Niesenhaus, J., Korhonen, H., Hoogen, W., Poels, K., Ijsselsteijn, W., & Kort, Y. (2009). Playability and player experience research. In Proceedings of *DiGRA 2009*. London, UK.

Nacke, L., Stellmach, S., Sasse, D., & Lindley, C.A. (2009). Gameplay experience in a gaze interaction game. In Proceedings of *COGAIN 2009: Gaze Interaction For Those Who Want It Most*. 49-54. Lyngby, Denmark

Rodriguez, S., Guerra, B. Higgins, T., & Mangels, J. (2011). Task goals and achievement mindset influence attention to feedback and learning success in a challenging memory task. *Cognitive Neuroscience Society 18th Annual Meeting, San Francisco.*

Salen, K., & Zimmerman, E. (2004). *The Rules of Play*. Cambridge: MIT Press.

Squire, K.D. (2005). Changing the game: What happens when videogames enter the classroom? *Innovate, 1(6).*

Sweller, J., Low, R., & Jin, P. (2010). Learner's cognitive load when using educational technology. In R. Van Eck (Ed.), *Gaming & cognition: Theories and perspectives from the learning sciences*. Hershey, PA: IGI Global.

Thomas, J., Bol, L., Warkentin, R., Wilson, M., Strage, A., & Rohwer, W. (2006). Interrelationships among students' study activities, self-concept of academic ability, and achievement as a function of characteristics of high-school biology courses. *Applied Cognitive Psychology, 7(6)*, 499-532.

Vygotsky, L. (1978). *Interaction between learning and development, in mind in society*. (Trans. M. Cole). Cambridge, MA: Harvard University Press, p. 79-91. Accessed on July13, 2010 from: http://www.comnet.ca/~pballan/Vygotsky%281978%29.htm

Williams, D., Yee, N., & Caplan, D. (2008). Who plays, how much, and why? A behavioral player census of virtual world. *Journal of Computer Mediated Communication, 13*, 993–1018.

Winn, J., & Heeter, C. (2009). Gaming, gender, and time: Who makes time to play? *Sex Roles*, *61*, 1-13.