The Effects of Age on Player Behavior in Educational Games

Eleanor O'Rourke, Eric Butler, Yun-En Liu, Christy Ballweber, and Zoran Popović
Center for Game Science
Department of Computer Science & Engineering, University of Washington
{eorourke,edbutler,yunliu,christy,zoran}@cs.washington.edu

ABSTRACT
Casual games attract a diverse group of players with varied needs and interests. In order to effectively tailor games to specific audiences, designers must consider the effects of demographics on player behavior. This is particularly important when developing educational games for children, since research shows that they have different design needs than adults. In this work, we develop in-depth metrics to capture demographic differences in player behavior in two educational games, Refraction and Treefrog Treasure. To learn about the effects of age on behavior, we use these metrics to analyze two player populations, children on the educational website BrainPOP and adults of the popular Flash website Kongregate. We show that BrainPOP players make more mathematical mistakes, display less strategic behavior, and are less likely to collect optional rewards than Kongregate players. Given these results, we present design suggestions for casual games that target children.

Categories and Subject Descriptors
K.8.0 [**Personal Computing**]: General – Games; H.5.0 [**Information interfaces and presentation**]: General

Keywords
games, player behavior, analytics, children.

1. INTRODUCTION
Video game players are a diverse group of people with varied interests, backgrounds, and motivations for playing. A recent study shows that women make up 47% of players, and that player age is evenly split between children under 18, young adults ages 18 to 35, and adults older than 35 [10]. Games are also rising to prominence as a way to motivate people to achieve serious goals, such as education [11], health [26], or scientific discovery [8]. Serious games must appeal to widely varied audiences to successfully achieve these goals. Despite this diversity of players, little is known about how demographics affect in-game behavior. This presents challenges for designers, since it is difficult to create games that appeal to diverse audiences without first understanding how players differ. It is also challenging for researchers to generalize experimental results without understanding the expected behaviors of the populations they study.

These challenges are particularly prevalent in the design of games for children. Young children make up a large portion of video game players, in part due to the growing interest in games as educational tools [11, 20]. Child-computer interaction researchers have shown that the design needs of children and adults differ, and that technology for children should consider these requirements [22, 9]. Game researchers have developed guidelines that suggest design considerations for children's games [19, 16, 3]. This work provides a valuable theoretical grounding, however very few of these guidelines have been verified through empirical studies comparing the behavior of adults and children in widely released games.

In this work, we study the effects of age on player behavior in natural settings with two casual online games developed by our research group, Refraction and Treefrog Treasure. Previous research has shown that it is challenging to collect demographic information from online players, and as a result most existing studies rely on optional surveys or bring players into unnatural settings such as labs [27, 21]. Instead of collecting demographic data explicitly, we note that many websites target a demographically distinct population of players. We released our games on two websites: Kongregate, which targets males ages 18 to 35, and BrainPOP, which targets elementary school children. We compare the behavior of these two player populations by analyzing detailed telemetric data. Though we do not know the demographics of any particular player, we can measure systematic differences in behavior and make inferences about the possible causes based on the target population of each website.

We describe the fine-grained metrics we developed for Refraction and Treefrog Treasure to measure differences in player behavior, and present results from an analysis of 8,000 players showing that the Kongregate and BrainPOP populations behave very differently. Although they show different levels of engagement in the two games, we find that BrainPOP players make more mathematical mistakes, display less strategic behavior, and are less likely to collect optional rewards than Kongregate players in both games. We provide explanations for our results and discuss possible implications for the development of games that target children.

2. BACKGROUND

Human computer interaction researchers have studied how children use technology for decades, showing that they have unique design needs [22, 9]. This research suggests that cognitive development affects children's interactions with technology. Gelderblom and Kotzé present design guidelines for children based on cognitive theories and describe concrete lessons from their research [12, 13]. They suggest that children's limited short-term memory affects their ability to remember instructions, and that their problem-solving strategies are influenced by brain maturation, conceptual understanding, and past experiences [13]. Other work has compared how children and adults search the web [14, 4], finding that children repeat the same search queries frequently, possibly due to their poor cognitive recall.

Researchers have applied cognitive theories to game design as well, developing guidelines for children's games. Moreno-Ger et al. explore methods of balancing fun and pedagogical goals [19], Linehan et al. look at ways to incorporate the Applied Behavioral Analysis teaching method into games [16], and Baauw et al. design a question-based evaluation methodology for assessing fun and usability in children's games [3]. This work provides a valuable foundation, however these high-level guidelines are not based on empirical evidence. Very few studies explore the effects of age on in-game behavior, even though Yee found that age is the most effective predictor of player's behavioral roles in a study of adults in online role-playing games [27]. Most closely related is a study by Pretorius et al. that compared how adults over 40 and children 9 to 12 learn an unfamiliar game by recording eye-tracking data and observing interactions with the tutorial. They found that adults approached the new game systematically and read instructions, while children ignored instructions in favor of a trial-and-error approach [21].

With this work, we build on existing research in child-computer interaction and cognitive development. We design in-depth behavioral metrics, perform an empirical analysis of how adults and children play games, and provide design considerations for games for children.

2.1 Game Website Audience

We released our games on two casual game websites that attract very different types of players. The first, Kongregate, is a popular portal for free Flash games. Developers can upload games to the site, which currently has over 50,000 titles available. Kongregate provides a variety of social features, and players can create optional accounts to become part of this community. Kongregate attracts 15 million unique players per month, which they report are 85% male with an average age of 21 [15]. The Alexa rankings for the website support this data, showing that visitors are predominantly males between ages 18 and 24 [2].

The second, BrainPOP, is a popular educational website [6]. BrainPOP is best known for its curriculum resources, including content for students and support materials for teachers. The BrainPOP Educators community has over 210,000 members [7], and the website is used as a resource in around 20% of elementary schools in the United States (Traci Kampel, personal communication). BrainPOP's educational game portal GameUp was designed for use in the classroom, and offers 54 games. While BrainPOP does not collect demographic information about players, the Alexa rankings show that the website is frequented by children and women ages 35 to 44, who are most likely teachers [1].

Given the distinct target audiences of these websites, we expected to observe differences in the behavior of the two populations. Many factors could produce these differences, such as variations in the website interfaces or differences in the social incentives and achievements provided by each site. Kongregate offers more games than BrainPOP, giving players many alternatives if they become bored with their current game. BrainPOP is primarily used in the classroom, which could influence player interactions. Despite these differences, we believe that player demographics will have the strongest effects on behavior. Kongregate attracts young male adults with prior gaming experience, while BrainPOP attracts children from diverse backgrounds. These populations will have different skill sets and developmental abilities, strongly affecting how they interact with games.

2.2 Expected Behavioral Differences

While there are many demographic differences between the populations we studied, we expected player age to affect behavior the most. Previous work has shown that age can influence play more strongly than gender [27], which will be most apparent when comparing very young players and adults. This intuition informed our hypotheses about the differences in behavior we expected to observe, which are based on in-person observations of children and adults and relevant theories in education and cognitive development.

Both Refraction and Treefrog Treasure are games about fractions. Fractions are a challenging concept for most elementary school children, and are considered one of the first serious roadblocks in math education [25]. As a result, we expected BrainPOP players to make more mistakes relating to fractions than Kongregate players. Although many adults find fractions difficult, we expected them to have stronger mathematical skills than children on average.

Hypothesis 1: *Kongregate players will make fewer mathematical mistakes than BrainPOP players.*

Children are also likely to display fewer strategic and problem-solving skills than adults. Extensive research in cognitive development has shown that problem-solving skills take years to develop [23, 24]. Children develop these skills slowly as they refine strategies and develop the metacognitive abilities needed to reflect on problem solutions [5]. As a result, we expected that adult players would display more strategic behavior than children.

Hypothesis 2: *Kongregate players will display more strategic behavior than BrainPOP players.*

Achievement systems are a huge part of the casual game industry for websites like Kongregate [18], and previous research has shown that Kongregate players will go out of their way to collect optional rewards [17]. However, during playtesting we have observed that children are less interested in rewards, often ignoring them entirely. We expected adult players to care about optional rewards more than children.

Hypothesis 3: *Kongregate players will collect more optional rewards than BrainPOP players.*

We explore these hypotheses by conducting an in-depth analysis of player behavior in two educational games.

3. METHOD

In this work, we study player behavior in two educational games developed by our research group: Refraction and Treefrog Treasure. Both games were designed to teach fraction concepts to elementary school children, however they have found success with older audiences as well. Both games are implemented in Flash and played in a web browser, and while both cover mathematical topics, they come from different genres and provide distinct gaming experiences.

3.1 Refraction

Refraction is a puzzle game that involves splitting lasers into fractional amounts. The player interacts with a grid that contains laser sources, target spaceships, and asteroids, as shown in Figure 1. The goal is to satisfy the target spaceships and avoid asteroids by placing pieces on the grid. Some pieces change the laser direction and others split the laser into two or three equal parts. To win, the player must correctly satisfy all targets at the same time, a task that requires both spatial and mathematical problem-solving skills. Refraction has 62 levels, some of which contain coins, optional rewards that can be collected by satisfying all target spaceships while a laser of the correct value passes through the coin. It has been played over 200,000 times on BrainPOP since its release in April 2012, and over 500,000 times on Kongregate since its release in September 2010.

3.2 Treefrog Treasure

Treefrog Treasure is a platformer that involves jumping through a jungle world and solving numberline problems to reach an end goal. The player navigates sticky, bouncy, and slippery surfaces and avoids hazardous lava to win. Numberline problems serve as barriers that the player solves by hitting the correct target location, as shown in Figure 2. The player can collect gems placed throughout the level, and gains additional gems by solving numberline problems. Gems are lost when the player dies, but they are scattered around the area of death and can be recollected. Treefrog Treasure has 36 levels, covering increasingly complex numberline concepts. The game is newer than Refraction; it has been played over 90,000 times on BrainPOP and over 17,000 times on Kongregate since its release in November 2012. On 4399.com, the most popular Flash game portal in China, it has been played over 5 million times since July 2012.

3.3 Data Collection

We collected two Refraction data sets, one from Kongregate and one from BrainPOP. The Kongregate data set contains 6,174 players and was collected from April 1, 2012 to December 2, 2012. We note that Kongregate awards players a badge if they collect all Refraction coins, which could influence player motivations. The BrainPOP data set contains 4,756 players and was collected from November 26, 2012 until December 3, 2012. Since these data sets had different numbers of players and were collected over different time

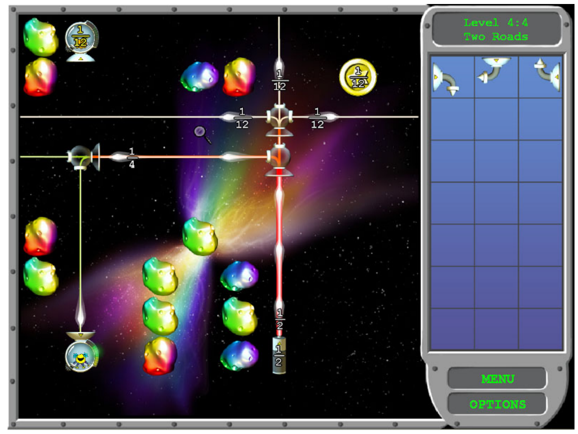


Figure 1: A level of Refraction. The pieces on the right are used to split lasers into fractional amounts and redirect them to satisfy the target spaceships. All ships must be satisfied at the same time to win.

Figure 2: A screenshot of Treefrog Treasure. The player navigates sticky, slippery, and bouncy surfaces and solves numberline problems to progress to the finish. Gems are collected along the way.

periods, we randomly sampled 3,000 players from each set to control for possible timing effects.

We also collected two Treefrog Treasure data sets. Due to time constraints, we collected less Kongregate data than we wanted. The Kongregate data set contains 1,412 players, collected from December 12, 2012 to December 13, 2012. Kongregate does not award any badges for this game. The BrainPOP data set contains 30,893 players, collected from December 7, 2012. to December 13, 2012. We randomly sampled 1,000 players from each data set for our analysis.

Both games recorded and logged every interaction players made with the game or its interface. We only included new players who were not familiar with the games in our analysis, and only used data from a player's first session to control for issues with shared computers in schools. In both games, we tracked players by storing their progress in the Flash cache, allowing us to selectively include new players and determine whether players returned to the game. One drawback of this method is that players who clear the cache or change computers will be treated as new players. However, since the Flash cache is inconvenient to clear and this action deletes all game progress, we considered this risk to be small.

4. DATA ANALYSIS AND RESULTS

We explore our hypotheses by performing a statistical analysis of game-specific behavioral metrics. The evaluation of the Kolmogorov-Smirnov statistic for each of our metrics was statistically significant, indicating that they violate the assumptions of normality. We therefore use non-parametric statistical methods: a Mann-Whitney U Statistic and a r measure of effect size for continuous variables, and a Chi-square statistic and a Cramer's V measure of effect size for categorical variables. We report effect sizes in addition to p-values to show the magnitude of the difference between our populations, since we are likely to find significant trivial differences due to our large sample sizes. For both tests, effect sizes with values less than 0.1 are considered trivial, 0.1 are small, 0.3 are moderate, and 0.5 or greater are large.

4.1 High-Level Behavior

First, we calculated descriptive statistics to gain an understanding of any high-level differences in player behavior. We looked at three metrics in both games: total time played, number of unique levels played, and return rate.

We calculated total time played by counting the number of active seconds of play, excluding menu navigation and idle periods with more than thirty seconds between actions. In Refraction, we found that Kongregate players play longer than BrainPOP players, with a median time of 437.00 seconds compared to 189.00 ($p<.001$, $r=0.36$). However, in Treefrog Treasure they play for less time, with a median time of 312.00 seconds compared to 476.00 ($p<.001$, $r=0.15$).

We calculated the number of unique levels completed for each player by counting levels with at least one game action. We exclude players who quit a new level before making any moves. In Refraction, Kongregate players play more unique levels, a median of 12 levels compared to 6 for BrainPOP ($p<.001$, $r=0.31$), but in Treefrog Treasure they play fewer unique levels (medians of 8 and 9, $p<.008$, $r=0.06$).

We calculated the return rate for each population by computing the percentage of players who come back to the game within three days. We found no significant difference in the return rate for Refraction, but found that 29.2% of BrainPOP players return to Treefrog Treasure, while only 9.2% percent of Kongregate players return ($p<.001$, $V=0.25$).

The Kongregate and BrainPOP populations reacted differently to the games. Kongregate players complete more levels and play longer than BrainPOP players in Refraction, but complete fewer levels and play for less time in Treefrog Treasure. The median values for these metrics indicate that Kongregate players are more engaged by Refraction, while BrainPOP players are more engaged by Treefrog Treasure.

4.2 Mathematical Understanding

We expected Kongregate players to make fewer mathematical mistakes than BrainPOP players. To explore this hypothesis, we designed game-specific metrics to capture mathematical mistakes in Refraction and Treefrog Treasure.

4.2.1 Refraction

Refraction players must understand fractions to split lasers appropriately and satisfy target ships. We measure two common mistakes to capture mathematical understanding. The first occurs when players try to satisfy a ship with an incorrect laser value, and the second occurs when an unsatisfied ship cannot be satisfied given the remaining lasers and pieces. Conducting a clean analysis of mathematical ability is challenging because it is often entangled with both spatial ability and the player's attempts to collect coins. To control for this, we analyzed two early levels that have no coins and require minimal spatial reasoning.

In this analysis, we wanted to capture the player's tolerance of mathematical mistakes rather than the raw number of mistakes made. Often, players place pieces on the grid to see what effect they produce, which could artificially inflate the total number of mistakes. We defined an "erroneous game state" as any state in which a mistake was present on the grid. If a player makes many moves without correcting a mathematical error, her trace will contain a large number of erroneous game states. Both of our mistake metrics count the total number of states in which the mistake is present.

The first mistake metric counts the number of game states in which a laser with the incorrect fractional value enters a target ship. A larger proportion of BrainPOP players' states contained these mistakes, with medians of 0.06 and 0.17 mistakes for the two levels compared to 0.00 and 0.00 for Kongregate players ($p<.001$, $r=0.31$ and $r=0.47$). We also calculated the proportion of active time spent in states with this mistake. Again, BrainPOP players perform worse, with medians of 0.09 and 0.20 compared to 0.00 and 0.00 for Kongregate players ($p<.001$, $r=0.36$ and $r=0.43$).

The second mistake metric counts the number of game states where the pieces are placed such that an unsatisfied ship cannot be satisfied with the remaining lasers and dividers. For example, if the player splits to create two 1/2 lasers, it becomes impossible to satisfy a ship that requires 1/3 power. We found that a larger proportion of BrainPOP players' states contained these mistakes, with medians of 0.17 and 0.12 compared to 0.00 and 0.00 for Kongregate players ($p<.001$, $r=0.38$ and $r=0.37$). BrainPOP players also spent a greater proportion of time in these states, with medians of 0.23 and 0.14 compared to and 0.00 and 0.00 for Kongregate players ($p<.001$, $r=0.38$ and $r=0.37$).

These findings support our first hypothesis. BrainPOP players are more tolerant of mathematical mistakes than Kongregate players, and spend more time in erroneous states on average. The effect sizes for all of these calculations were moderate to large, indicating that the two populations have very different levels of mathematical understanding.

4.2.2 Treefrog Treasure

Players must understand fractions to solve the numberline problems in Treefrog Treasure. To solve a problem, the player jumps into the numberline multiple times until hitting the correct location. Each failed attempt can be viewed as a mathematical mistake. The severity of the mistake can be captured by calculating the distance between the guess and the correct solution. We performed our analysis on three numberline problems from the first non-introductory level.

First, we calculated the proportion of correct attempts to total attempts across all three numberlines. Players who solved all three problems correctly on the first try have a proportion of 1. We found that Kongregate players have a greater proportion of correct attempts than BrainPOP players, but the medians for both are 1.0 ($p<.001$, $r=0.29$). Next, we calculated the average error, or distance from the correct solution, for failed attempts across all three numberlines. We found that BrainPOP players have a greater average error, with a median of 6% compared to 3% for Kongregate players ($p<.001$, $r=0.37$).

One explanation for these results is that Kongregate players are more careful and take the time to place their jumps correctly. To explore this possibility, we calculated the number of seconds spent between the state preceding the first jump into the numberline and the first jump. We found that BrainPOP players spend more time considering their first jump on average, with a median of 16.37 seconds compared to 12.21 seconds for Kongregate players ($p<.001$, $r=0.27$).

These results also support our first hypothesis. BrainPOP players spend more time planning their jumps than Kongregate players, but still make more mistakes with a greater margin of error. The effect sizes are moderate, indicating that the differences are considerable. It is possible that BrainPOP players make mistakes because they have have lower mouse proficiency than adults, however we believe that mathematical understanding plays a greater role because we see the same effect in Refraction.

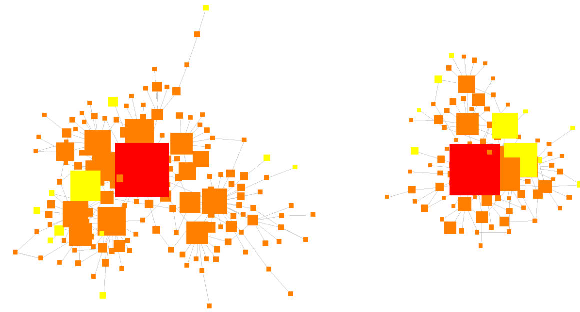
4.3 Strategy

We expected Kongregate players to act more strategically than BrainPOP players. To explore this hypothesis, we designed game-specific metrics to capture strategic behavior in Refraction and Treefrog Treasure. Refraction is a puzzle game that requires the use of problem-solving skills, so we were able to perform a deep analysis of strategic behavior in this game. Treefrog Treasure targets a younger audience, and was designed to require less strategy. However, we identified a few key tasks in this game that require planning, a central component of strategy, to study in our analysis.

4.3.1 Refraction

To complete a refraction level, a player must search the space of all possible moves to find a correct solution. One way to visualize her search through this space is as a graph, where the nodes represent unique board states and the edges represent moves. Given this representation, we can calculate metrics that capture characteristics of the player's search strategy. In this analysis, we use the same two levels without coins used to study mathematical mistakes.

First, we calculated the proportion of visited states that are unique. High values indicate that the player does not revisit states often, while low values show that the player returns to states over and over again. We would expect strategic players to have a high proportion of unique states, because they



(a) BrainPOP (b) Kongregate

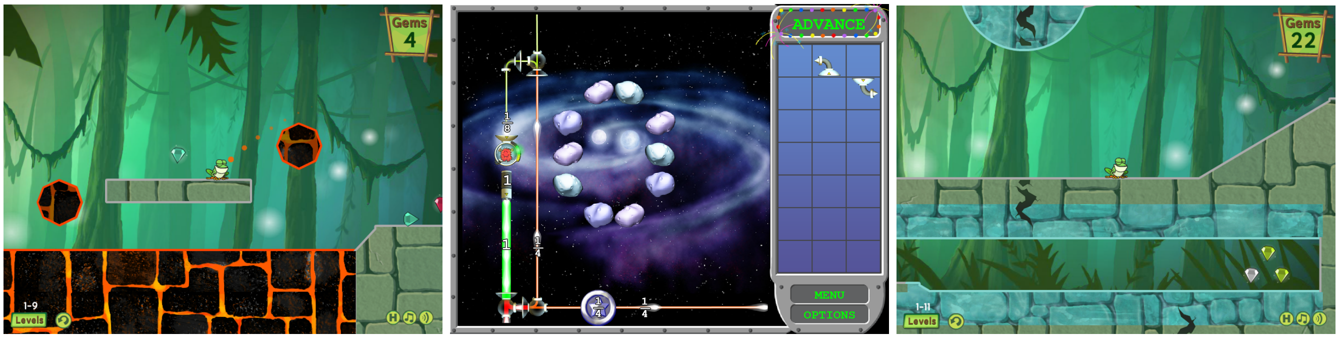
Figure 3: Refraction search graph representations, showing that BrainPOP players search less efficiently. Red is the start state, yellow boxes are win states, and orange boxes are intermediate states.

search efficiently and remember previously viewed states. We found that Kongregate players have a higher proportion of unique states than BrainPOP players in both of the analyzed levels, with medians of 0.5 and 0.5 compared to 0.36 and 0.37 ($p<.001$, $r=0.27$, and $r=0.26$).

Next, we calculated each player's search degree. Search degree as the average out degree of the player's graph, where out degree is the number of edges leaving a particular node. This metric captures the type of search the player is using. A high out degree indicates the use of a breadth-first-like strategy, and a low out degree indicates the use of a depth-first-like strategy. Lower search degree could also show that the player is thinking ahead and investigating multi-move hypotheses. We found that Kongregate players have a lower search degree, with medians of 1.5 for both levels, compared medians to 2.0 and 2.5 for BrainPOP players ($p<.001$, $r=0.35$ and $r=0.43$).

Finally, we calculated the number of "dead ends" a player reaches. A dead end is a node with only one neighbor, created when the player immediately takes back a move and returns to the previous state. A node is also a dead end if the player resets the board immediately after reaching that state. We would expect players who search inefficiently and have trouble predicting the effects of moves to have a large number of dead ends. BrainPOP players visited more dead end states, with medians of 1.0 and 3.0, compared to 0.0 and 0.0 for Kongregate players ($p<.001$, $r=0.38$ and $r=0.43$). During this analysis, we noticed that BrainPOP players return to the start state frequently. We counted the number of times players remove all pieces from the board, and found that BrainPOP players return to the start state more, with a median of 9 returns to Kongregate's 3 ($p< .001$, $r=0.43$).

These results support our second hypothesis. Kongregate players display more efficient and strategic search behavior, returning to states less often, searching more deeply, and reaching fewer dead ends than BrainPOP players. Kongregate players also reset the board less frequently, indicating that they are better able to reason about intermediate states. Figure 3 shows aggregate search graphs for a single level. These graphs were created by randomly selecting 240 players from each population and displaying states visited by



(a) (b) (c)

Figure 4: Screenshots of analyzed levels. Figure 4(a) shows moving hazards in Treefrog Treasure, Figure 4(b) shows a solution to a coin level in Refraction, and Figure 4(c) shows inconvenient gems in Treefrog Treasure.

at least two players. Size represents the volume of players and color indicates state type. BrainPOP players are less focused; they search more broadly, visit a wider variety of states, and reach more dead ends than Kongregate players.

4.3.2 Treefrog Treasure

Treefrog Treasure was explicitly designed to require minimal strategy. As a result, we explored our second hypotheses by identifying three level regions where players will perform better if they plan ahead and consider multiple moves in advance. The first region, shown in Figure 4(a), involves planning jumps to avoid moving hazardous objects. The other two regions involve jumping back and forth between walls to reach a better vantage point.

To measure performance in the hazardous region, we count the number of times players die by hitting the moving hazards. We found that BrainPOP players die more frequently than Kongregate players, although the median number of deaths for both was 1.0 ($p<.001$, $r=0.19$). For the other two regions, we count the number of jumps players use to reach the top of the wall, expecting players who plan ahead to require fewer jumps. BrainPOP players use more jumps in the first region, but the medians for both are 4 jumps ($p<.001$, $r=0.21$), and we found no statistically significant difference in the second region.

These results also support our second hypothesis. While the effect sizes are small and this analysis does not provide as deep of a picture of strategic thinking as our Refraction analysis, these results suggest that Kongregate players are better able to plan moves in advance, and therefore avoid obstacles and jump between objects more efficiently.

4.4 Optional Rewards

We expected Kongregate players to collect more optional rewards than BrainPOP players. To explore this hypothesis, we designed game-specific metrics to capture player interactions with coins and gems in Refraction and Treefrog Treasure. While both games include optional rewards, each requires different behavior to collect them. In Refraction, coins are collected by solving a more challenging problem with additional constraints, while in Treefrog Treasure, in-level gems are collected by taking the time to visit inconvenient locations.

4.4.1 Refraction

Refraction coins are static pieces with fractional values that appear on the game grid. Coins are collected by satisfying all the spaceships while a correctly valued laser passes through the coin, as shown in Figure 4(b). In this analysis, we calculated whether players successfully collect coins and the total time spent working to collect them. We analyze three early levels with coins, including the one shown in Figure 4(b).

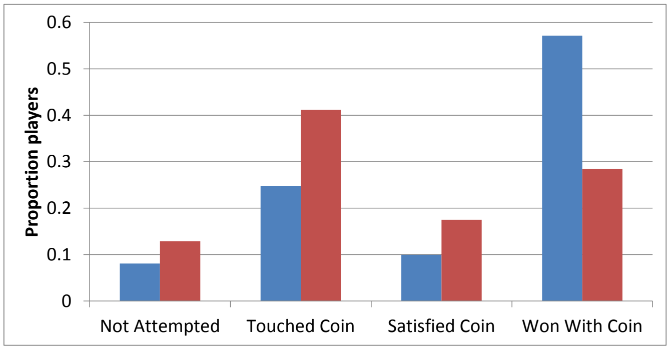
First, we looked at how players interact with coins in the three levels. We bucketed players into four interaction groups: players who never touch the coin with a laser, players who only touch the coin with an incorrectly valued laser, players who touch the coin with the correctly valued laser but do not win it, and players who win the coin. A graph showing this distribution for one of the levels is shown in Figure 5(a). Next, we calculated the raw number of coins that the two player groups collected on average, and found that Kongregate players collect more coins, with a median of 2.0 out of three possible coins collected, compared to 1.0 out of three for BrainPOP ($p<.001$, $r=0.42$).

We also calculated the amount of time players spent attempting to collect each coin on average. We looked at each play of the level during the player's first session, and only included plays in which the coin was touched with the laser at least once. We summed the amount of time spent across all level plays with these characteristics, and found that Kongregate players spent more time on average working to collect coins, with a median of 413 seconds compared to 268 for BrainPOP players ($p< .001$, $r=0.24$).

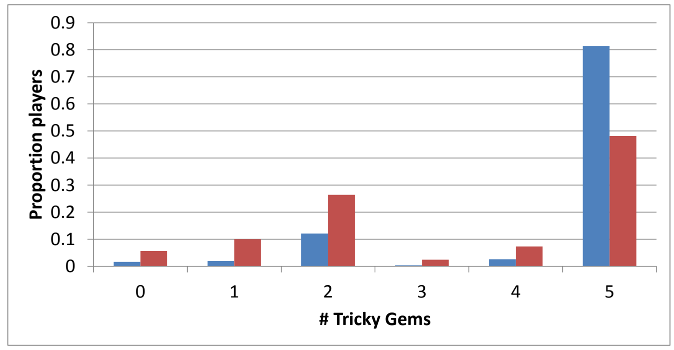
These results support our third hypothesis; Kongregate players collect more coins than BrainPOP players. Coin collection is conflated with mathematical ability, which could increase the size of the observed effect. However, BrainPOP players spend less time than Kongregate players working towards coins, indicating that they are less interested in collecting them. It is important to note that Kongregate players may be motivated to collect coins primarily due to the badge awarded to players who collect them all.

4.4.2 Treefrog Treasure

Gems in Treefrog Treasure can be collected in two ways. Players receive gems when they complete numberline prob-



(a)



(b)

Figure 5: Optional rewards graphs. Figure 5(a) shows four types of interaction with Refraction coins. Figure 5(b) shows the total number of tricky gems that Treefrog Treasure players collect, of five possible gems. In both graphs, Kongregate is blue and BrainPOP is red.

lems correctly, and they can also collect "fixed" gems that appear in the level. For this analysis, we look at the total number of gems collected in nine early levels and the number of gems players recollect after losing them by dying. We also identified a few fixed gems that are particularly inconvenient to collect, and examined these specifically in our analysis.

First, we calculated the percentage of the total possible gems that players collected across the nine levels, and found that BrainPOP players collect fewer gems on average, with a median percentage of 93% compared to 95% for Kongregate players ($p<.003$, $r=0.12$). Next, we looked at the percentage of lost gems that players regain. When players jump or fall into hazardous lava, they die and are re-spawned. Players lose a few gems every time they die, which are scattered around the area of death. We calculated the percentage of these gems that players recollected after dying on average, and found no significant difference.

We also analyzed two sets of gems that are inconvenient to collect. One set, shown in Figure 4(c), is only visible after the player has gone past the tunnel where the gems can be collected. We hypothesized that players would only go back to collect these tricky gems if they cared strongly about optional rewards. We found that BrainPOP players were less likely to collect these gems, with a median of 4.0 gems collected out of five compared to 5.0 out of five for Kongregate players ($p<.001$, $r=0.33$), shown in Figure 5(b).

These results support our third hypothesis. Kongregate players collect more gems than BrainPOP players, and while they do not recollect more lost gems, they are more likely to collect inconvenient gems. Kongregate players do not receive any achievements for collecting gems in Treefrog Treasure, indicating that they are genuinely more motivated or able to capture these optional rewards than BrainPOP players.

5. CONCLUSION

Our analysis of Kongregate and BrainPOP players in two casual games shows that the populations behave very differently. The games we studied, Refraction and Treefrog Treasure, come from different genres and provide distinct gaming experiences. High-level behavioral statistics indicate that the two player populations experience different levels of engagement with each game; Kongregate players finish more levels and play longer in Refraction than BrainPOP players, but have the opposite behavior in Treefrog Treasure. Despite these differences, all three of our experimental hypotheses were empirically supported in both games. BrainPOP players make more mathematical mistakes than Kongregate players, creating more incorrect lasers in Refraction and requiring more tries to complete numberline problems in Treefrog Treasure. They also display less strategic behavior. BrainPOP players use less efficient search strategies in Refraction, returning to states more frequently and reaching more dead ends than Kongregate players, and they are less efficient in the parts of Treefrog Treasure that require planning. Finally, BrainPOP players were less interested in optional rewards in both games, spending less time working towards Refraction coins and returning to collect tricky gems in Treefrog Treasure less frequently.

While the specific demographics of the Kongregate and BrainPOP player populations are only loosely confirmed, our results suggest that age had the strongest effect on player behavior. While our results should be repeated and confirmed by future studies with verified demographic data, the observed behavioral trends suggest design considerations for games that target children. Our data show that children have trouble searching large state spaces, suggesting that game designers should restrict the amount of searching and planning their games require. This could be achieved by limiting the size of the search space or by using tutorials and scaffolding to teach search strategies directly. Children also struggled with the mathematical concepts in our games. While this is not surprising, game designers should offer scaffolding or visual feedback when errors are made to ease the cognitive load for young players. Finally, children were less likely to collect optional rewards, indicating that they are not a strong motivator for this population. Designers may want to exclude optional rewards from games for young audiences.

With this work, we provide a model for studying the effects of demographics on player behavior that could be generalized to study other player populations. Our results show that each casual game website attracts players with distinct demographics, and that the effects of those demographics on player behavior should be considered when generalizing research results to different populations. It would be valuable to study the effects of demographics such as age, gender, and education on in-game behavior further, to allow designers and researchers to make principled predictions about how specific populations will react to a given game. Our research method could also be applied to work in adaptive games. In order to appropriately adapt a game to a particular player, the adaptive game must appropriately cluster players with similar needs and behaviors. A classifier trained on data sets with distinct demographics, such as the Kongregate and BrainPOP data sets, could produce successful clustering algorithms. We will explore this direction in future work.

This work highlights significant differences in the behavior of Kongregate and BrainPOP players, and represents the first in-depth comparative analysis of in-game behavior based on demographic characteristics. Our results support our prediction that any observed differences would be primarily caused by the age of the players, which suggests design guidelines for games that target children based on our results.

6. ACKNOWLEDGMENTS

We would like to thank the creators of Refraction and Treefrog Treasure for making this work possible. We also thank Anthony Pecorella of Kongregate and Allisyn Levy and Norman Basch of BrainPOP for promoting our games and helping us gather data. This work was supported by the NSF Graduate Research Fellowship grant DGE-0718124, the Bill and Melinda Gates Foundation grant OPP1031488, the Office of Naval Research grant N00014-12-C-0158, the Hewlett Foundation grant 2012-8161, Adobe, and Intel.

7. REFERENCES

[1] Alexa Ranking, BrainPOP.

[2] Alexa Ranking, Kongregate.

[3] E. Baauw, M. M. Bekker, and W. Barendregt. A structured expert evaluation method for the evaluation of children's computer games. In *Proceedings of the 2005 IFIP TC13 international conference on Human-Computer Interaction*, INTERACT'05, pages 457–469, Berlin, Heidelberg, 2005. Springer-Verlag.

[4] D. Bilal and J. Kirby. Differences and similarities in information seeking: children and adults as web users. *Inf. Process. Manage.*, 38(5):649–670, Sept. 2002.

[5] D. F. Bjorklund, editor. *Children's Strategies: Contemporary views of cognitive development.* Lawrence Erlbaum Associates, Inc., Hillsdale, New Jersey, USA, 1990.

[6] BrainPOP. http://www.brainpop.com/.

[7] BrainPOP Educators. http://www.brainpop.com/educators/home/.

[8] S. Cooper, F. Khatib, A. Treuille, J. Barbero, J. Lee, M. Beenen, A. Leaver-Fay, D. Baker, Z. Popović, and F. Players. Predicting protein structures with a multiplayer online game. *Nature*, 466(7307):756–760, August 2010.

[9] A. Druin. Cooperative inquiry: developing new technologies for children with children. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, CHI '99, pages 592–599, New York, NY, USA, 1999. ACM.

[10] Entertainment Software Association. Essential facts about the computer and video game industry, 2012.

[11] J. P. Gee. *What Video Games Have to Teach Us About Learning and Literacy.* St. Martin's Press, revised and updated edition. edition, March 2008.

[12] H. Gelderblom and P. Kotzé. Designing technology for young children: what we can learn from theories of cognitive development. In *Proceedings of the 2008 annual research conference of the South African Institute of Computer Scientists and Information Technologists on IT research in developing countries: riding the wave of technology*, SAICSIT '08, pages 66–75, New York, NY, USA, 2008. ACM.

[13] H. Gelderblom and P. Kotzé. Ten design lessons from the literature on child development and children's use of technology. In *Proceedings of the 8th International Conference on Interaction Design and Children*, IDC '09, pages 52–60, New York, NY, USA, 2009. ACM.

[14] T. Gossen, T. Low, and A. Nürnberger. What are the real differences of children's and adults' web search. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, SIGIR '11, pages 1115–1116, New York, NY, USA, 2011. ACM.

[15] E. Greer and A. Pecorella. Core games, real numbers: Comparative stats for mmos & social games. Presented at the Game Developers Conference, San Francisco, CA, 2012.

[16] C. Linehan, B. Kirman, S. Lawson, and G. Chan. Practical, appropriate, empirically-validated guidelines for designing educational games. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 1979–1988, New York, NY, USA, 2011. ACM.

[17] Y.-E. Liu, E. Andersen, R. Snider, S. Cooper, and Z. Popović. Feature-based projections for effective playtrace analysis. In *Proceedings of the 6th International Conference on Foundations of Digital Games*, FDG '11, pages 69–76, New York, NY, USA, 2011. ACM.

[18] C. Moore. Modern gaming trends: Achievements, 2012.

[19] P. Moreno-Ger, D. Burgos, I. Martínez-Ortiz, J. L. Sierra, and B. Fernández-Manjón. Educational game design for online education. *Comput. Hum. Behav.*, 24(6):2530–2540, Sept. 2008.

[20] H. F. O'Neil, R. Wainess, and E. L. Baker. Classification of learning outcomes: evidence from the computer games literature. *The Curriculum Journal*, 16(4):455–474, 2005.

[21] M. Pretorius, H. Gelderblom, and B. Chimbo. Using eye tracking to compare how adults and children learn to use an unfamiliar computer game. In *Proceedings of the 2010 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists*, SAICSIT '10, pages 275–283, New York, NY, USA, 2010. ACM.

[22] J. C. Read, P. Markopoulos, N. Parés, J. P. Hourcade, and A. N. Antle. Child computer interaction. In *CHI '08 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '08, pages 2419–2422, New York, NY, USA, 2008. ACM.

[23] R. F. Sternberg, editor. *Thinking and Problem Solving.* Academic Press, San Diego, California, USA, 1994.

[24] S. Thornton. *Children Solving Problems.* Harvard University Press, Cambridge, Massachusetts, USA, 1995.

[25] U.S. Department of Education. The final report of the national mathematics advisory panel, 2008.

[26] Y. Xu, E. S. Poole, A. D. Miller, E. Eiriksdottir, R. Catrambone, and E. D. Mynatt. Designing pervasive health games for sustainability, adaptability and sociability. In *Proceedings of the International Conference on the Foundations of Digital Games*, FDG '12, pages 49–56, New York, NY, USA, 2012. ACM.

[27] N. Yee. Motivations of play in online games. *Journal of CyberPsychology and Behavior*, 9:772–775, 2007.