# Exploring the Influence of Demographic Factors on Progression and Playtime in Educational Games

Amogh Joshi
joshi134@purdue.edu
Purdue University
West Lafayette, Indiana, USA

Christos Mousas
cmousas@purdue.edu
Purdue University
West Lafayette, Indiana, USA

D. Fox Harrell
fox.harrell@mit.edu
Massachusetts Institute of Technology
Cambridge, Massachusetts, USA

Dominic Kao
kaod@purdue.edu
Purdue University
West Lafayette, Indiana, USA

## ABSTRACT

Games are now ubiquitous, and educational games are becoming increasingly prevalent. Like other games, educational video games attract participants from different ethnicities and with different gender expressions. As such, educational game designers face a necessity to develop inclusive games. In this paper, we focus on inclusivity, diversity, and equity (DEI) issues by investigating if the computer programming game *Mazzy* benefited participants from broad demographic backgrounds. We highlight inclusive features present in *Mazzy*, and, focusing on the participants' self-reported gender and race/ethnicity, reflect on their play experience and learning outcomes. We found evidence that the game supported learning outcomes and facilitated an engaging play experience for participants from diverse demographic backgrounds. We discuss challenges and implications for the broader literature.

## CCS CONCEPTS

• **Human-centered computing → HCI design and evaluation methods**.

## KEYWORDS

inclusivity, educational game, demographics, playtime, game progression

**ACM Reference Format:**
Amogh Joshi, Christos Mousas, D. Fox Harrell, and Dominic Kao. 2022. Exploring the Influence of Demographic Factors on Progression and Playtime in Educational Games. In *FDG '22: Proceedings of the 17th International Conference on the Foundations of Digital Games (FDG '22), September 5–8, 2022, Athens, Greece.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3555858.3555873

## 1 INTRODUCTION

Educational games are widely present in today's school curricula [10, 111, 116]. They can be powerful tools to learn a wide range of topics such as computer programming [71, 84], math [145], physics [6], public health literacy [82], etc. Researchers recommend taking promising aspects of popular video games to design educational games, such as immersive and immediate feedback [45, 46, 135]. However, developing games for learning can be a more challenging task than developing games for fun. Educational games need to support inclusivity [57]. This is because educational video games can improve learning outcomes [130], which ultimately affects educational achievement, occupational success, and well-being [126].

Educational video games need to account for real and virtual experiences of the students. Female and ethnic minority groups often report being linguistically profiled when playing online multiplayer games and hence, face racist and sexist comments [49, 57, 123]. These experiences can negatively affect their sense of belonging in games and sour future gameplay experiences in the classroom and online settings [123]. Students may differently receive the learning material inside educational video games because of prior negative experiences while learning in the classroom. Both gaming and Science, Technology, Engineering and Math (STEM) fields, such as Computer Science, are male-dominated and have a lower representation of Black, Latinx, and female students [4]. The lower rates of enrollments in STEM fields mean that under-represented groups may find it challenging to cultivate peer relationships as effectively as other groups, affecting their sense of belonging [22, 100]. Additionally, the under-represented groups face a cold reception in the classroom. They receive negative stereotypical judgments about their abilities that contribute to low self-efficacy beliefs [90, 125] which are difficult to overcome and affect how they interact with computer programming games.

In this paper, we explore approaches taken by researchers to design educational games that are inclusive (see Section 2). We situate our game, *Mazzy*, which was previously used to facilitate computational learning among learners from currently underrepresented groups (see Section 3). We then aggregate data from various experiments conducted using *Mazzy* as a testbed to reflect on instances where *Mazzy* supported learning of computational concepts for participants (n=3574) and places where learning experience can be made more inclusive. We identify implications for design and evaluation of educational games (see Section 6).



*FDG '22, September 5–8, 2022, Athens, Greece*

ACM ISBN 978-1-4503-9795-7/22/09.
https://doi.org/10.1145/3555858.3555873

## 2 RELATED WORK

We turn to the literature that has investigated game preferences, game design frameworks, and the use of games user research (GUR) to make inclusive games.

### 2.1 Understanding Game Preferences for Inclusivity

Game companies research their gamer demographic and tailor games and experiences according to their players' preferences. For children, some games are designed to promote active play [92] (such as Pokemon Go [107]) and foster curiosity about the world. Games for the elderly, on the other hand, sometimes have been developed to stimulate brain activity and prevent the onset of various illnesses, such as Alzheimer's disease [23]. In academic literature, researchers have further investigated how certain games (and their genres) appeal to a specific demographic [58, 62]. Game players from many different demographics appreciate many types of games. That said, there are trends in the games currently taken up by player in different demographic groups. One study has indicated that a majority of female players show more appreciation for casual and social aspects of games [62]—e.g., often found in massively multiplayer online role-playing games (MMORPGs), puzzle, and party games—compared to the first person shooter, fighting, and sports genres that players identifying as male seem to prefer [23, 58]. Players over the age of 55 prefer playing intellectually stimulating games, such as strategy, puzzle, and educational games [16]. On the other hand, younger players prefer games with fast reflexes such as action, racing, and shooter [50, 58, 114]. Demographic preferences for *educational video* games follow a similar pattern. Studies have shown that boys preferred to play educational games that have an active style of play (e.g., quick actions [25, 88] and a rapid sequence of events) or strategic play (e.g., manipulating resources over time). Girls preferred educational games with creative play (e.g., storylines, customizing avatars, building or modifying environmental elements) and interaction (e.g., with non-player characters (NPCs), pets, or friends) [25, 88]. A study found that the player's age played a significant role in determining the strategies in an educational math game [110]. School children adopted a trial-and-error strategy, made more mistakes, and collected fewer optional rewards than young adults.

By tailoring games according to a specific demographic group, some educational game designers have hoped to create a more engaging learning experience for their intended audience. However, many research scholars have noted the lack of literature surrounding race/ethnicity in relation to playing games [35, 57, 68]. Analysis in existing research papers is often limited to descriptive statistics broken down by nationality, income, and education [51, 148]. With a limited sample size for populations other than self-identifying White participants, it is difficult to ascertain gameplay preferences for participants belonging to diverse ethnic backgrounds. An absence of research in understanding play motivations and game preferences of under-represented groups could mean that game designers may not design inclusive games. Conversely (and perhaps perversely), educational game designers may only know how to design games that appeal predominantly to White-male players, thereby reinforcing harmful stereotypes regarding play preferences of under-represented groups [113]. We should be careful not to only suggest making games in genres tailored to the trends above because there may be value in people playing other types of games and, also, we want to support players who do not fit the trends. Passmore highlights how generalizations can be created for under-represented groups because only certain game genres—such as sports and military-themed games—have adequate representation for Black or African American gamers [113]. Even if game genre preferences are known, and the game players adhere to genre preference trends (keeping in mind that there might be a diverse number of reasons for such trends, e.g., the communities of other players who play the game), it is immediately clear how game designers can design inclusive games. This is because a game genre only provides a top-level perspective that does not explain how smaller building blocks of the game should be designed [141]. For instance, an educational math game can be designed in a first-person perspective or a third-person perspective. In both these game genres, there are additional design considerations—such as design of learning activity, feedback to the learner—that cannot be readily ascertained by knowing the genre of the game. Overall, this suggests that while understanding demographic preferences can be a useful starting point to develop educational video games, designing inclusive games requires a much more nuanced perspective.

### 2.2 Game Design Frameworks and Inclusivity

Over the years, scholars have proposed a multitude of game design frameworks that help game designers develop specific aspect of games, such as achievements [53], narrative [15], game feel [117], mechanics [60], etc. Game design frameworks have also been proposed for educational games [7, 28, 66, 147]. A common theme for many game design frameworks is that they help elicit a design solution by structuring specific concepts relating to game design. However, only a handful of frameworks have focused on how to develop inclusive games [42, 61, 121].

Ibrahim developed a Gender Inclusive Framework (GIF) that combined elements from the game genre (e.g., action, strategy, racing), content (e.g., storyline, graphics), and gameplay (e.g., feedback system, personalization) [61]. The framework highlights how various game elements can be combined so that the educational games can be more engaging for *male and female* learners. Other frameworks focus on the relationship the game designer shares with the process of game development [42] and with the intended audience [42, 121]. Flanagan and Nissenbaum presented a "Values at Play" methodology wherein designers follow a three-step process when they develop and design games: discovery, translation, and evaluation [42]. The authors highlight how they designed an inclusive computer programming game (Rapunzel) for female learners by first discovering the values that the game should contain (e.g., inclusivity, diversity). In the translation process, the authors show how the values discovered during the discovery phase were turned into features in the game (e.g., cooperation, diverse representation of game characters). The verification process focused on evaluating the designed game (e.g., user testing and empirical evaluation) with the intended audience. Rankin and Irish similarly advocate that the intended audience should play a more significant role during the design and development of educational games and emphasizes the need for underrepresented groups (e.g., Black or African American women) to participate in the process of game design [121]. The authors highlight how they applied a framework grounded in critical theory ("Black Feminist Thought") to guide the development of a Spanish learning game wherein both the designers and the participants were Black or African American women. As such, this framework points to the importance of involving relevant stakeholders—game designers and participants—during the game-making process so that the game reflects the identity of the intended audience.

While the frameworks provide valuable guidelines to design inclusive games, we argue that an empirical evaluation of educational games can provide further clarity on the player experience. This is especially important for *educational games* because the frameworks mentioned above do not explicitly give recommendations on how to design computer programming learning activities in the game. By evaluating the player experience, game designers can understand the degree to which the experience of the player matched the intended player experience. In this sense, it helps game designers improve on the game's initial design and better structure learning activities in the game.

## 2.3 Games User Research and Inclusivity

In industry, GUR is commonly used to understand the usability and experience of participants for many aspects of games: gameplay, controls, UI, audio, etc. [38]. Games user researchers employ a broad range of qualitative (e.g., interviews, focus groups) and quantitative techniques (e.g., game analytics) to understand how players interact with the game, which subsequently informs game design [48, 104]. Making educational games (like games) is an iterative process. This means that the learning activity situated in the game needs to go through frequent revisions before a polished version is available for the target audience to play. Previously, many interventions only collected summative measures of learning (e.g., a pre-post test design) [17]. While measuring overall learning outcomes is essential, the summative evaluation fails to provide meaningful insights into the game's design or the learning activity. This suggests a need to involve GUR when designing and evaluating games.

Employing GUR techniques for educational games is a relatively new phenomenon. In a study investigating player behavior in a computational learning game (*GrACE*), researchers visualized participants' progression through levels that increased in difficulty [59]. The study found that some intermediate levels of the game were more challenging to players compared to later levels. Similarly, researchers compared player actions of participants differing in gameplay experience for a popular puzzle game (Portal) to the optimal number of player actions needed to complete a given level [146]. The study found that novice players (but not expert players) performed significantly more actions in some game levels, suggesting a need to redesign levels for novice players. Both studies employed qualitative (e.g., observation, think-aloud, and interview protocols) and quantitative methods (game analytics), allowing researchers to make robust inferences about player behavior as well as suggest improvements in the game design. It isn't easy to employ qualitative research techniques with larger sample sizes. While this limits inference making on player behavior, the studies have suggested improvements to the level design [54, 115]. In a math-learning game aimed to facilitate learning fractions for middle school students, researchers visualized player activity (collected by game analytics) to reflect how level design—that embedded a series of increasingly difficult mathematical concepts relating to fractions—affected students' interaction with the game [115]. The study found several levels that did not support student learning and transfer to different problem representations. Similarly, modeling player data from a game designed to teach physics concepts (such as balance) to young children (ages 5-8) discovered that some levels facilitated simplistic learning strategy (rote learning) that failed to promote a deeper understanding of physics concepts [54].

A common trend across the studies is that GUR techniques were used to evaluate the game design. In evaluation, the researchers found instances where the game supported learning and made recommendations to re-design learning activities and gameplay. While the design process for the studies mentioned above did not explicitly focus on making inclusive games, participants across various studies differed in socio-demographic indicators [59], level of expertise [146], age [54], and gender [146]. This suggests that GUR has the potential to make inclusive games. In this paper, we similarly describe the design of *Mazzy* and evaluate the degree to which the game facilitated an engaging learning experience for participants of diverse backgrounds.

# 3 MAZZY

*Mazzy* is a computer programming game where players program their game character to reach a goal [76, 80].[1] Players use typical programming constructs such as loops (e.g., "for loops") and conditional statements (e.g., "if block") to navigate their game character through increasingly complex levels. *Mazzy* was developed to facilitate computational learning among under-represented students (e.g., female and Black or African American learners). *Mazzy* was designed through a process of iteration wherein user evaluations for the game were conducted through in-person and online methods. As such, the designers of *Mazzy* leaned heavily on GUR. We note that the use of inclusive game design frameworks mentioned above could result in a different look and feel of the game. However, many of the game design frameworks have only been recently proposed [61, 121]. Moreover, the focus of the game was to understand how a specific aspect of game—avatar creation and customization—affects learning and engagement of participants who are under-represented in STEM. *Mazzy* included two key inclusive ideas to facilitate a positive, engaging player experience: provide space for learners to construct their virtual identity (avatar creation and customization) and inclusive learning design. We detail each of these concepts below.

## 3.1 Avatar Creation and Customization

Avatars are often referred to as "digital selves" because they represent a user in a virtual world [37]. Avatar customization is the process of changing avatars' physical resemblance (e.g., changing body shape, age, race, gender, name, and clothes) to create an authentic representation of the user in virtual environments [12, 144]. The process of creating and customizing an avatar is known to

[1]See gameplay video: http://youtu.be/n2rR1CtVal8


**Figure 1: Mii Avatar Creator**

engender avatar identification that has shown to improve intrinsic motivation [12], performance [13], and enjoyment [143] in the virtual environment.

Avatar customization and identification are highlighted as one of the key features that help players, especially the players belonging to an ethnic minority, feel more comfortable playing video games [61, 139]. When the players can identify with their game characters, they enjoy the games more because they experience a merging of identities [24]. Ibrahim mentions avatar representation as an essential feature in developing inclusive educational games for female learners [61]. Rankin and Irish highlighted the emphasis African American women place on having their avatar look authentic in a Spanish language learning game [121]. Virtually all users desire to create a representation of themselves in digital spaces. However, previous studies have noted a lesser number of customization options available for players of other races (e.g., Black or African American, Latinx) and gender expression (e.g., female, transgender) [56, 68, 102, 106]. Only a handful of studies have explored how appearance of avatar influences how students learn in digital environments [93, 136]. A key recommendation to create inclusive games is to adequately support character creation for under-represented players by offering them a multitude of choices [68, 101, 139]. Some authors have also suggested randomizing the character creation process to create diverse game characters [139]. For instance, the game Heroes Wanted [142] randomly assigned aspects of a character that do not conform to stereotypical representation of race and gender [139]. In this sense, the game created diverse game characters that players can embody and interact with in the game world.

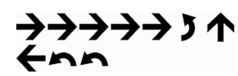
In this paper, the users customized their avatars before playing the computer programming game *Mazzy*. Players were offered a wide range of avatar choices[2] [77, 79, 81], such as creating avatars that looked like the player (i.e., self-avatar) [77], role-models (e.g., scientists, athletes) as avatars [75], and avatars that changed based on game contexts (e.g., dynamic avatars) [77]. See Figure 1.

## 3.2 Learning Activity Design

In digital learning environments, learner's experiences depend not only on the content of the subject [128] but also on how it is presented to the learner [87]. In order to develop computational thinking for students coming from diverse backgrounds, *Mazzy* incorporated use of pseudocode and learning curves.

*3.2.1 Pseudocode.* While computer code must strictly adhere to formal representations of logic, pseudocode is a representation of a program written in a natural language [9, 44]. Writing pseudocode as a precursor to formal programming adheres to various guidelines and best practices found in books [34] and pedagogy [109] that stress programmers not to "rush to keyboard" but to first reflect on the problem to come up with a general solution [30]. The advantage of working with pseudocode is that it reduces the cognitive load experienced by the learner. The learner can first focus on writing a program in a natural language (e.g., English). This way, the learner fully understands the steps they need to include to write a given program successfully. Learners can then convert the pseudocode into programming language-specific syntax. Studies show that students learning with pseudocode developed positive perceptions regarding programming [30], promoted development of formal programming skills [30, 108, 124], experienced lower perceived difficulty [108], and increased confidence and interest in computer science and programming [108], especially for women [30].

**Table 1: Levels 1 through 6**

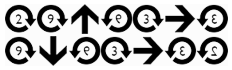
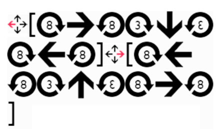
| Level | Screenshot | A Sample Solution |
|---|---|---|
| 1 |  |  |
| 2 |  |  |
| 3 |  |  |
| 4 |  |  |
| 5 |  |  |
| 6 |  |  |

In *Mazzy*, learners write pseudocode in symbols instead of language. Users input arrow keys to move the player in a particular direction (↑→↓←). If the user wants to move their character forward twice, then they can use either input forward arrow keys (↑↑) or use a loop symbol with a number embedded inside in conjunction with arrow keys to navigate the player (see Table 1, 2).

*3.2.2 Learning Curves.* Learning curves refers to the "structure and the pace through which challenges are introduced to a player"

[2]See [69] (Chapter 5) for a complete list of experiments.

**Table 2: Levels 7 through 12**

| Level | Screenshot | A Sample Solution |
|---|---|---|
| 7 |  |  |
| 8 |  |  |
| 9 |  |  |
| 10 |  |  |
| 11 |  |  |
| 12 |  |  |

[95]. Challenge in educational video games can take the form of game difficulty and inherent difficulty of the educational content. Game difficulty is the difficulty that is perceived by the player as they learn how to play the game [127]. Difficulty of the educational content depends on the inherent difficulty that learners face when learning educational content [21]. For example, many scholars agree that computational concepts such as using loops and branching statements is a more difficult concept to learn [19, 52].

The shape of the learning curve is known to influence engagement and performance of the learners [54, 115, 127]. A multitude of factors, such as game elements, and the type of problem solving activity determine the shape of the learning curve [115]. In other words, there isn't a generalized prescribed curve applicable for all educational video games. A general design guideline is to incrementally increase the complexity of learning activity [54] and provide space for the player to practice a newly learned skill [95]. In *Mazzy*, the players are first introduced to basic commands (level 1-5), after which they are introduced to usage of loops (level 6-9) and loops with conditionals (level 10-12). These learning activities are spread across 12 different levels, wherein subsequent levels allow the player the space to practice the newly learned skill (see Tables 1, 2).


**Figure 2: Level 1 (left) and Level 6 (right) in *Mazzy*.**

# 4 METHODS

## 4.1 Secondary Data

Secondary data is broadly defined as "inquiry based on the re-analysis of previously analyzed research data" [63]. This extends to re-analyzing one's data with a different purpose [1, 133]. Secondary data is useful because it allows other researchers to generate new insights from existing data without additional costs [91]. Its use is widespread in academic community in the form of literature reviews [39], meta-analysis [43] and has contributed to developing a greater understanding of video-games [29, 67, 122] and educational games [17, 26, 103]. However, secondary data may have methodological challenges. The reused dataset may contain errors in measurement and sampling errors [138] which may limit the validity of it. Even if the secondary data may not contain errors, secondary data presents problems when it is used to make meaningful comparisons within groups [133]. We address these concerns here before moving on to the analysis.

The dataset we use in this paper comes from experiments run by the 3rd and last authors during the period 2014-2017, where the purpose was to investigate the role played by the avatars in helping students learn computational concepts[3]. The project had a particular emphasis on improving learning outcomes of under-represented groups in STEM fields (e.g., Black or African American and women) through the use of avatars in a game-based learning environment. Participants were recruited from Amazon Mechanical Turk (AMT)—a reliable research tool to recruit participants and is commonly used in research (e.g., computer science, psychology, economics) to conduct online studies [11, 18, 99]. The platform provides the functionality to ensure that every participant recruited for research purposes can be unique. We used AMT's worker qualification feature, which prevents workers from participating in subsequent studies. The dataset contains multiple experiments with slight differences in experimental protocols[4]. For instance, participants answered different inventories or played with a different avatar. This means that all participants in the dataset—aside from avatar design and selection—played the same identical game. All the experiments in the dataset employed a between-subjects design and randomly assigned the participants to different (and varied) avatars.

In this paper, we only focus on the data that was collected in-game through the use of game analytics. Game analytics are sophisticated tools that capture in-game player data (e.g., playtime) with minimal risk of measurement error. We used the in-game data

[3] Raw data available on OSF: https://osf.io/fx24g/?view_only=30eaa00cbdae4fa8935f708c85887cf8
[4] For additional details, see [69].

to investigate the relationship between playtime and progression of players and demographic variables. Our aim in this paper is not to compare demographic groups so as to make inferences about the populations but to investigate the degree to which the game, *Mazzy* promoted an engaging play experience for various specific demographic groups. We argue that the use of a reliable research platform, use of robust protocols, and robust analysis minimizes the measurement errors and validity concerns expressed by various scholars [133, 138].

## 4.2 Dataset

*Mazzy* was developed iteratively over a period of 3 years and contains five game versions that had slight variations in aesthetics and level design. In this paper, we focus on analyzing the final game version (Version #4) as it contains polished graphics and updated user interface compared to its previous versions. Moreover, the data from early game versions did not collect all demographic information from the participants (e.g., age, gender) which are present in the final game version. It also contains almost half of the total participants recruited (Participant distribution: Version #0—23.4%, Version #1—4.0%, Version #2—11.9%, Version #3—17.0%, and Version #4—43.8%). Game version #4 is the final version described in this paper and shown in Figure 2.

This version contains a total of N=3919 participants, recruited through Amazon Mechanical Turk (AMT). On AMT, requesters use the platform to post Human Intelligence Tasks (HITs), such as games, which participants complete for monetary compensation. The majority of participants are based in the U.S., while a small number (~5%) are from outside of the U.S. Participants were allowed to play for any length of time before exiting the game and were paid $3.00 USD (~$9.68 USD / hour based on average playtime).

## 4.3 Measures

Here, we describe the measures used in this study. Playtime, progression, and age were numerical variables, while gender, income, education, and race were categorical variables. For each variable, we describe the data cleaning process (e.g., removing entries that did not meet the minimum threshold, invalid entries, etc.). The final dataset for analysis contained N=3574 participants, after the data cleaning process.

*4.3.1 Gender.* Gender data was collected using questions from the U.S. census[5]. Participants self-identified as either male or female. 2084 participants self-identified as male (~58%) and 1490 participants self-identified as female (~42%).

*4.3.2 Age.* Participants entered an exact numerical value for their age. We removed impossible or blank entries (N=10, ~0.2% of the raw data). The mean age was 30.5 (SD=8.83), with the majority of participants in the age bracket of 25-44 years (N=2303, 64.4%) followed by 18-24 years (N=981, 27.4%), 45-64 years (N=276, 7.72%) and lastly, 65 and above years (N=14, 0.39%).

*4.3.3 Race.* Race data was collected in the same manner as the U.S. census. A sample size less than 25 limits inference making because data patterns may not be accurately determined [65]. As such, scholars have recommended a minimum sample size of 25 when conducting regression analysis to clearly determine patterns [65]. We drop all categories where the sample size was less than 25. This included dropping a level (N=215) called 'Other' that combined numerous ethnicities (for instance, White-Black or African American) into a single category. In total, we drop (N=272, ~7% of the raw data). The majority of the data set self-identified as White (N=3094, proportion=86.5%). Table 3 provides a breakdown of participants by self-identified race.

**Table 3: Self-reported race**

| | Race | Count | Proportion |
|---|---|---|---|
| 1 | White | 3094 | 86.57 |
| 2 | Black or African American | 219 | 6.13 |
| 3 | Asian Indian | 104 | 2.91 |
| 4 | Chinese | 64 | 1.79 |
| 5 | Filipino | 32 | 0.90 |
| 6 | Korean | 31 | 0.87 |
| 7 | Other Asian | 30 | 0.84 |

**Table 4: Self-reported household income of participants.**

| | Income | Count | Proportion |
|---|---|---|---|
| 1 | Less than $12,500 | 430 | 12.03 |
| 2 | $12,500 - $24,999 | 559 | 15.64 |
| 3 | $25,000 - $37,499 | 615 | 17.21 |
| 4 | $37,500 - $49,999 | 464 | 12.98 |
| 5 | $50,000 - $62,499 | 459 | 12.84 |
| 6 | $62,500 - $74,999 | 240 | 6.72 |
| 7 | $75,000 - $87,499 | 211 | 5.90 |
| 8 | $87,500 - $99,999 | 203 | 5.68 |
| 9 | $100,000 or more | 393 | 11.00 |

*4.3.4 Education.* The majority of participants had some college experience but not degree (N=1260, 35.3%), followed by bachelor's degree (N=1225, 34.3%), followed by master's or doctorate degree (N=518, 14.5%), followed by associate's degree (N=417, 11.7%) and lastly, some high school (N=154, 4.31%). We removed invalid entries (N=1 ~0.02% of the raw data).

*4.3.5 Income.* Participants were asked to select their household income from a dropdown box containing 9 different income ranges. We removed invalid entries (N=5, ~0.13% of the raw data). See Table 4.

*4.3.6 Progression.* Levels in *Mazzy* were designed to have incremental difficulty and introduced new concepts through level progression (see Section 3). Participants can only move to the next level when they demonstrate competency with a particular programming concept. Playing a greater number of levels means that the participants interacted with a greater variety of computational concepts which consequently signals greater opportunity to learn [33, 87].

[5] At the time these studies were run, the U.S. census regrettably did not have non-binary options for gender. We acknowledge that not having non-binary options is a limitation of this work and discuss this further in the limitations section.

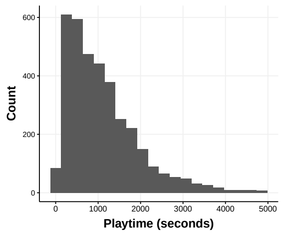



**Figure 3: Playtime frequency distributions.**

We operationalized progression as number of levels completed: Progression = $\frac{100}{12}$(Levels Completed). Progression takes numerical values ranging from 0 to 100 and is calculated based on the last level completed. That is, if a participant completes 4 levels in the game, their progression is ~33%. A score of 50 roughly meant that the participants completed half of the game while a score of 100 means that participants completed all twelve levels in the game.

*4.3.7 Playtime.* We interpret playtime as an implicit measure of engagement (e.g., motivated behavior) [31, 36, 118, 132]. Playing for a greater amount of time, especially in volitional context such as the one in this study, suggests that the participants were more engaged with the game [8, 132]. However, in this context, we interpret playtime in conjunction with progression. Low playtime may be a sign of disengagement when the participant simultaneously has a low progression score (i.e., the participant played briefly and made little progress). A high progression score and a short playtime, on the other hand, does not necessarily signal disengagement, but could simply represent quicker game completion.

Playtime was measured in seconds. The playtime graph has a right skew (see Figure 3) and so we removed participants that played longer than 82.5 mins (mean + 3*SD). We considered these as extreme outliers that would affect the validity of our conclusions (N=57, ~1.4% for the raw data) [129]. The mean playtime was 1103 seconds (18.4 mins), SD=840 (~14 mins).

## 4.4 Reflection on Participant Sample

We note that our dataset contains substantially more participants identifying as female (~42%) compared to 18.7% female students in Computer Sciences who were awarded bachelors degree in 2016 [3]. However, our dataset contains slightly lower proportion of participants (~6%) identifying as Black or African American compared to 8.68% of students who earned a bachelors degree in 2016 for Computer Sciences [2]. In recent years, scholars have observed a trend of decreasing enrollment for under-represented groups in Computer Sciences [105, 150]. For instance, 27% of bachelors degrees in Computer Sciences were awarded to women in 1997 [3] and 9.58% for Black or African American students during the same time period [2]. As such, there is a crucial need to support education of under-represented groups in computer science education.

## 4.5 Analysis

Our goal was to determine if different demographic factors (age, gender, race, income, and education level) influenced our measured gameplay variables (playtime and progression). Data was cleaned according to the criteria in Section 4.3 (Measures). Considering the study's exploratory nature, we first created a regression model consisting of all the explanatory variables and interactions. The explanatory variables included were age (numerical), gender (dichotomous categorical), household income (categorical with nine levels), race (categorical with seven levels), highest education attained (categorical with five levels). The inclusion of interaction terms depended on the sample size. As small sample sizes would limit inference-making, we only include interaction terms with large enough samples in each cell (>25) [65]. In the "full-model," all two-factor interactions were specified except for interactions between education–income, gender–race, income–race. The three-factor interactions—age–gender–education and age–gender–income were included. We created a second model consisting of main effects only. We then compared the two models for goodness-of-fit and proceeded with the simpler model when possible. After specifying the model, we first checked for the assumptions related to the regression and then proceeded to interpret the findings.

We conducted the analysis in R [137]. To avoid overparameterization, R automatically deletes the reference level of the categorical variable from the output tables when fitting the regression equation [27]. As a result, "White" (race), "male" (gender), "Less than $12,500" (household income), and "Some High School" (education) are not present in the regression output. The coefficients for the levels of categorical variables are interpreted in relation to their reference level.

# 5 RESULTS

## 5.1 Playtime

The playtime variable was first log transformed due to its right skew (see Figure 3) before specifying the initial linear model. We first conducted an F-test for model comparison between the main-effects model and the full-model. We found no significant difference between the two models: $F(1,43) = 1.04$, $p = 0.38$. This means that the "full-model" that contained two and three factor interaction terms did not significantly differ from the "main-effects model" containing the independent variables. As such, we proceeded with the main-effects regression model. We found that the model violated the constant variance assumption, $\chi^2$=5.42, df=1, $p = 0.01$. To address this violation, we used a heteroscedasticity consistent covariance matrix (HCCM) [97] to construct standard errors using sandwich package in R [149].

The results show that 3.86% of the variance can be accounted for by the 20 predictors collectively, $F(20,3553)$=8.17, $p < 0.001$. We found a positive effect of age ($\beta$=0.01, $p < 0.001$) on playtime. Participants self-identifying as female ($\beta$=0.08, $p = 0.003$) played for a longer time compared to their male counterparts. We observed a general trend that participants with a higher household income played for less time compared to the participants with self-reported household income of less than $12,500. Education and race of participants did not significantly explain differences in playtime. See Table 5 for complete results.

**Table 5: Playtime Regression Results**

| | *Dependent variable:* |
|---|---|
| | log(Playtime) |
| Age | $0.018^{***}$ (0.015, 0.021) |
| Race:Black or African American | 0.108 (−0.007, 0.223) |
| Race:Korean | −0.032 (−0.293, 0.229) |
| Race:Chinese | −0.166 (−0.374, 0.042) |
| Race:Asian Indian | −0.079 (−0.265, 0.108) |
| Race:Filipino | 0.098 (−0.217, 0.413) |
| Race:Other Asian | 0.140 (−0.132, 0.413) |
| Gender:Female | $0.083^{**}$ (0.029, 0.137) |
| Education:Some college, no degree | −0.085 (−0.219, 0.049) |
| Education:Associates degree | −0.112 (−0.263, 0.039) |
| Education:Bachelors degree | −0.106 (−0.241, 0.030) |
| Education:Graduate degree | −0.065 (−0.212, 0.082) |
| Income:$12,500 - $24,999 | $-0.100^{*}$ (−0.199, −0.0002) |
| Income:$25,000 - $37,499 | $-0.109^{*}$ (−0.208, −0.011) |
| Income:$37,500 - $49,999 | $-0.156^{**}$ (−0.265, −0.048) |
| Income:$50,000 - $62,499 | $-0.159^{**}$ (−0.266, −0.052) |
| Income:$62,500 - $74,999 | −0.110 (−0.236, 0.016) |
| Income:$75,000 - $87,499 | $-0.153^{*}$ (−0.282, −0.024) |
| Income:$87,500 - $99,999 | $-0.243^{**}$ (−0.395, −0.090) |
| Income:$100,000 or more | $-0.136^{*}$ (−0.244, −0.028) |
| Constant | $6.327^{***}$ (6.159, 6.495) |
| Observations | 3,574 |
| $R^2$ | 0.044 |
| Adjusted $R^2$ | 0.0386 |
| Residual Std. Error | 0.826 (df = 3553) |
| F Statistic | $8.172^{***}$ (df = 20; 3553) |
| *Note:* | $^{*}$p<0.05; $^{**}$p<0.01; $^{***}$p<0.001 |

**Table 6: Progression Regression Results**

| | *Dependent variable:* |
|---|---|
| | Progression |
| Age | $-0.170^{***}$ (−0.265, −0.074) |
| Race:Black or African American | $-8.421^{***}$ (−11.858, −4.983) |
| Race:Korean | −0.372 (−9.184, 8.440) |
| Race:Chinese | $-6.681^{*}$ (−12.859, −0.503) |
| Race:Asian Indian | $-17.278^{***}$ (−22.220, −12.337) |
| Race:Filipino | 2.735 (−5.921, 11.391) |
| Race:Other Asian | 0.434 (−8.506, 9.375) |
| Gender:Female | $-4.482^{***}$ (−6.152, −2.812) |
| Education:Some college, no degree | −2.041 (−6.209, 2.128) |
| Education:Associates degree | −4.031 (−8.640, 0.578) |
| Education:Bachelors degree | −0.890 (−5.097, 3.317) |
| Education:Graduate degree | 0.373 (−4.182, 4.927) |
| Income:$12,500 - $24,999 | −1.851 (−4.980, 1.278) |
| Income:$25,000 - $37,499 | $-3.749^{*}$ (−6.826, −0.672) |
| Income:$37,500 - $49,999 | $-5.504^{**}$ (−8.792, −2.217) |
| Income:$50,000 - $62,499 | $-4.424^{**}$ (−7.732, −1.117) |
| Income:$62,500 - $74,999 | −3.097 (−7.056, 0.863) |
| Income:$75,000 - $87,499 | −2.529 (−6.655, 1.598) |
| Income:$87,500 - $99,999 | $-4.892^{*}$ (−9.076, −0.708) |
| Income:$100,000 or more | −1.772 (−5.253, 1.708) |
| Constant | $77.345^{***}$ (72.074, 82.616) |
| Observations | 3,574 |
| $R^2$ | 0.036 |
| Adjusted $R^2$ | 0.031 |
| Residual Std. Error | 24.809 (df = 3553) |
| F Statistic | $6.717^{***}$ (df = 20; 3553) |
| *Note:* | $^{*}$p<0.05; $^{**}$p<0.01; $^{***}$p<0.001 |

## 5.2 Progression

We first conducted an F-test for model comparison between the main-effects model and the full-model. We found no significant difference between the two models, $F(1,43) = 1.06$, $p = 0.35$. This means that the "full-model" that contained two and three factor interaction terms did not significantly differ from the "main-effects model" containing the independent variables. As such, we proceeded with the main-effects regression model. We found that the model satisfied the constant variance assumption, $\chi^2$=0.91, df=1, $p = 0.33$ but violated the normality assumption. However, the ordinary least squares (OLS) is considered to be robust against violations of normality, especially if the sample sizes are sufficiently large [89].

The results show that 3.1% of the variance can be accounted for by the 20 predictors collectively: $F(20,3553)$=6.71, $p < 0.001$. We found a negative effect of age ($\beta$=-0.16, $p < 0.001$) on progression. Participants identifying as Asian Indian ($\beta$=-17.27, $p < 0.001$), Black or African American ($\beta$=-8.42, $p < 0.001$), and Chinese ($\beta$=-6.68, $p = 0.03$) had significantly lower progression compared to participants self-identifying as White. Female participants had a significantly lower progression compared to male participants ($\beta$=-4.4, $p < 0.001$). The education of the participants did not explain progression. Some levels of income had a negative effect on progression. See Table 6 for complete results.

*5.2.1 Progression Across Levels.* We visualized level progression data to understand better why we found group differences in level progression. Our aim was to investigate specific levels that may have been too difficult for the participants. Figure 4 shows cumulative distribution for the entire data set. The figure shows a step pattern where participants' cumulative proportion gradually increases as we move from level 0 (before level 1 is completed) to level 12. When the participants stop playing at any level, their progression score is calculated based on the last level completed (see Subsection 4.3). For instance, if a participant completes level 4 but is unable to complete the next level (level 5), then the participant's progression reflects the last level that was completed (i.e., level 4). At each level, some participants stop playing, which increases the cumulative proportion for a given level (i.e., a step pattern graph). In other words, we can interpret the cumulative proportion figure as a graph showing the attrition of participants. Participants have

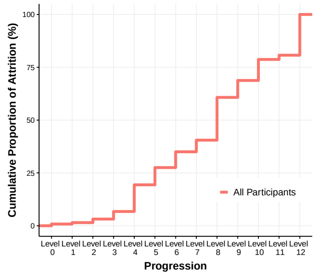



**Figure 4: Cumulative proportion of attrition for all participants**

~0% drop-off at Level 0[6] and 100% drop-off at level 12 because all participants stop playing the game. Of interest is the size of the step because it suggests that more participants quit after completing a given level. From Figure 4 we see three points of interest: Level 4, Level 8, and Level 12. We only focus on understanding why participants quit after completing levels 4 and 8 because all participants quit after completing level 12 (end of the game). 12.6% of participants stopped playing after completing level 4, while 20.2% of students stopped playing after completing level 8.

We now focused on two aspects of participants' backgrounds: race and gender. As mentioned previously, the game's goal was to facilitate learning of computational concepts for under-represented groups, specifically female and Black or African American participants. Additionally, the regression results suggest that participants self-identifying as female, Black or African American, Chinese, and Asian Indian had difficulty progressing through the game. We see a similar trend for the Black or African American participants' progression: a sharp increase in attrition at two points: Level 4 and Level 8 (see Figure 5 top-left). 21.9% of the participants identifying as Black or African American stopped playing after completing level 4 (vs. 11.6% of the participants identifying as White who stopped playing after level 4). We see a similar pattern for Asian Indian participants (23.1% drop-off after completing level 4) but not for Chinese participants (12.5% drop-off after completing level 4). See Figure 5 (top-right). Similarly, 23.3% of participants self-identifying as Black or African American stopped playing after completing level 8 (vs. 20.2% of White participants who stopped playing after completing level 8). Similarly, 26.6% of Chinese participants stopped playing after completing level 8. However, a small proportion (6.7%) of Asian Indian participants stopped playing after completing level eight. Participants self-identifying as Filipino, Korean and Other Asian showed a similar pattern of attrition as participants White participants (see Figure 5 bottom-left).

Lastly, we compared progression between male and female participants (see Figure 5 bottom-right). Female participants had a higher drop-off rate after completing levels 4 (15.2%) and 8 (20.1%). Male participants had a similar drop-off rate after completing level 8 (20.3%) but had considerably less drop-off than female participants after completing level 4 (10.7%) .

[6]We see a tiny proportion of participants that stopped playing the game before completing the first level of the game.

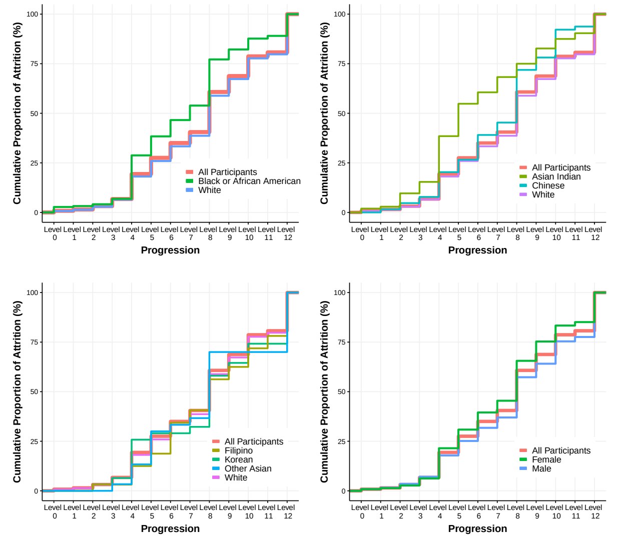



**Figure 5: Cumulative proportion of attrition for self-identifying Black or African American, White and all participants (top-left); self-identifying Chinese, Asian Indian, White and all participants (top-right); self-identifying Filipino, Korean, Other Asian, White and all participants (bottom-left); self-identifying male and female participants (bottom-right).**

# 6 DISCUSSION

As educational game designers, we want the player experience to be primarily determined by the game's design. Through its design, the game should facilitate an engaging experience for the learner where they can leave behind the preconceived attitudes and beliefs they may have regarding the learning activity and fully immerse themselves into the game world. We see some evidence of this happening with the game *Mazzy*. The low R-squared value for playtime (Adjusted R-squared=0.0386) and progression (Adjusted R-squared=0.031) regression models suggest that the demographic factors had little impact on the player experience (playtime and progression). To ensure that the low R-squared value was not due to model specification, we compared the R-squared value of the "full-model" that contained interaction terms, as opposed to "main-effects model" containing only the independent variables of progression (Adjusted R-squared=0.0317) and playtime (Adjusted R-squared=0.0391) regression models. The small difference between the R-squared values for "main-effects" and "full-model" indicates that the participants' background, in general, did not greatly explain the variability in progression and playtime. It was only when we visualized player progression that we could understand why the regression models have a low R-squared value: we see that participants, regardless of their group membership, stopped playing the game in substantially higher proportions after completing levels 4 and 8. In other words, participants' behavior in the game was governed less by their group membership and more by game elements (e.g., level design).

Participants quit in significantly greater proportions after completing levels 4 and 8 than after other levels. It is difficult to comment on the "acceptable" level of attrition. In educational contexts, we want *all* players to complete *all levels* in the game. However, other scholars have argued that attrition is "natural" for computer-mediated treatments such as eHealth [41]. Attrition is also a common phenomenon in contexts of volitional play, with many games and educational games reporting high levels of attrition [5, 8, 32, 98, 131]. A consensus among scholars is that it is important to understand the reason why participants quit [5, 131]. For this, researchers have modeled the playtime of players in games [8, 131] and concluded that playtime frequency distributions are useful to predict departure. However, other scholars have suggested that player departure can be attributed to the success rate in games [32]. That is, players stop playing the game when they are demotivated or discouraged by failure. The results of this study strongly favor the "success-rate" explanation. Tables 1 and 2 show specific learning challenges and puzzles for the twelve levels in the game. While participants still used basic commands to navigate their game character to the goal in levels 1-5, level 5 introduced a more complex map for navigation than previous maps and contained significantly more commands than levels 1-4.

Similarly, participants attempting level 9 used nested loops and faced a significantly more complex map than in previous levels. The preceding levels (e.g., levels 7 and 8) adhered to the recommendations outlined to design learning curves [95]. For instance, level 7 introduced a new concept (loops) and participants needed to use loops *once* to navigate their game character through a relatively short maze. The next level (level 8) introduced complexity in the learning activity and the level design. The players needed to use loops *thrice* to navigate their game character through a maze that was longer but of less complexity. However, the use of loops in the previous level may not have conveyed an understanding to the participants to use nested loops in level 9. Overall, 60.7% of the participants did not play more than 8 levels in the game. This suggests that the difficulty faced by participants in the subsequent levels (levels 5 and 9) may have been a cause of attrition for the participants.

We find further support for the "success-rate" explanation when observing the playtime differences across various demographic groups. The participants had similar playtime across race and education. Moreover, female and older participants had significantly higher playtime compared to male and younger participants, respectively. The findings demonstrate that participants, especially under-represented groups, were more engaged with the game. We understand the cause of disengagement when we look at the regression model for progression and cumulative proportion of attrition graphs. The drop-off rate was significantly higher for participants self-identifying as female, Black or African American, Chinese, and Asian Indian after completing levels 4 and 8. The similar (or higher) playtime but a lower progression suggests that participants became disengaged after repeated unsuccessful attempts in the subsequent levels (levels 5 and 9).

A clear recommendation from this study is to redesign levels 5 and 9. Literature is largely in agreement that loops are a difficult concept [20, 40, 47, 52, 120] for students to learn. However, participants playing *Mazzy* had a significant problem in the level that introduced the concept of nested loops (Level 9) compared to the level that introduced the loops (level 7). Only a handful of studies have investigated how students understand an advanced concept such as nested-loops [20, 47]. These studies show a need for a gentler introduction to advanced computational concepts. The level redesign can be especially beneficial for under-represented learners and make the game (*Mazzy*) more inclusive. Our results are in line with other studies that have suggested that under-represented learners stand to benefit substantially more from incorporating greater support in STEM learning environments [57].

We observed that the prior education of the participants did not explain playtime or progression. Moreover, participants in the lowest income group ('Less than $12,500') had a significantly higher playtime and progression score compared to other income groups. These results are encouraging because it suggests that *Mazzy's* design supported learning and engagement across a broad range of education groups as well for the participants who would stand to benefit the most from learning computational concepts. However, we do see a lower playtime and a lower progression score that suggests disengagement with the game for participants in higher income groups. A redesign of levels can therefore further increase efficacy of *Mazzy* in facilitating learning of computational concepts for participants of higher income backgrounds.

## 6.1 Contribution to Broader Literature

Our study has implications for the design of learning curves for computer programming games. The Additive Factors Model (AFM) is a learning model that assumes that the success of the learner depends on the sum total of individual factors such as students' characteristics, prior attempts, and skill difficulty [54, 115]. Apart from considering the difficulty of the learning content, the skill difficulty in educational games also comes from the difficulty presented by game elements—which is sometimes referred to as difficulty curves [127]. Players playing level 5 of the game *Mazzy* needed to navigate the character through a substantially more complex maze design using simple navigational commands that they learned in the prior levels. This suggests that the players may have perceived a greater difficulty due to level design. Level 9 of the game may have introduced a higher difficulty in the learning material (in the form of nested loops) and through a more complex level design. Other studies have also found that students perceive difficulty with game elements as well as the difficulty of the learning material [96]. As such, learning curves for educational games should consider the difficulty of the learning material as well as the difficulty introduced by game elements.

Our study has implications for designing computational concepts in educational video games. In *Mazzy*, learners used pseudocode to navigate their game character in the game. New commands to navigate the player were introduced first, and the subsequent levels facilitated practice and mastery of the previously introduced commands. However, this approach did not facilitate students intuitively using nested loops despite prior levels introducing the idea of loops (level 7). We recommend a gentler approach when introducing computational concepts embedded inside loops, such as conditional statements and loops (nested loops). This finding is in line with other studies that have advocated deconstructing loops further [20] wherein the learners are introduced to a specific aspect of loops (e.g., introducing the idea of the loop, followed by syntax, and so on).

Research has begun investigating video games at a granular level by looking at their building blocks: game elements [140, 141]. *Mazzy* included role-play and puzzle game elements. Participants selected their avatars to role-play while being engaged in a problem-solving activity. While previous research has sought to understand how we can make inclusive games designed to take into account the values and preferences of particular demographic groups (e.g., female learners [61]), the results of this study take a promising step towards developing and designing games that are inclusive to broader demographics. The playtime of participants did not differ according to their education or race. Moreover, female participants, participants in the lowest income group, and older participants had significantly greater playtime than male participants in the higher income groups and younger participants. This means that participants found the game broadly engaging and were motivated to complete the game. The progression of participants did not differ for participants of different educational backgrounds. The participants who self-identified as 'Korean', 'Other Asian', and 'Filipino' had similar progression scores as self-identifying 'White' participants. In cases of lower progression, redesigning the levels may facilitate a more inclusive learning experience for all participants.

As educational game designers think more critically about how to design inclusive game experiences, the results of the study highlight that game elements such as role-playing (e.g., creating an avatar) and puzzle game elements have the potential to provide an inclusive learning experience for broader demographic groups. While research has connected game elements to the playstyles of gamers [140], our study advances the understanding of how combinations of game elements (e.g., role-playing and puzzle) relate to designing inclusive educational video games. However, it is important to note that other game elements, such as presence of audio in the game [70, 72, 85, 86], design of tutorials [73, 83], social elements [74], and narratives (storylines) [14, 119], have also been shown to influence player experience. Future studies should investigate the degree to which various game elements and their combinations promote inclusivity.

## 7 LIMITATIONS

Because our data set is drawn from experiments that involved a random assignment to different experimental conditions (e.g., different player avatar types), there will inevitably be a degree of noise in the data. Moreover, all of our data comes from the same game (*Mazzy*), and therefore, our results may not necessarily generalize to all educational games. Similarly, because our participant pool comes from Amazon Mechanical Turk, where workers accept work in exchange for payment, this may differ from contexts of completely volitional play. Given the scope of our analysis, our data also does not capture cultural differences between individuals with the same self-reported race/ethnicity background. Participants were almost entirely from the U.S., and adapting U.S. census questions directly meant that some demographic groups were not adequately accounted for (e.g., non-binary genders). Additionally, we did not analyze participants of mixed ethnicity due to an inadequate sample size. We stress that it is crucial to consider these groups and to be more inclusive in our future work, we aim to follow the best-practice recommended in the literature [94, 134] such as providing the participants a greater variety of gender options, including a "prefer to self-describe" option. We also plan to conduct population-specific studies so that a greater number of participants may be encouraged to participate.

While the progression data highlights reliable patterns of player behavior, an explanation of why players quit can have other explanations (e.g., disinterest). Several researchers have suggested that employing methods of triangulation can help game designers make robust inferences about player behavior [59]. Because the data set does not contain this information, it is difficult to make inferences about player behavior. Future research should consider adopting more fine-grained analyses of in-game actions such as retry attempts and pre/post-test to better measure learning gains. This can also help explain player behavior in educational games [55]. Lastly, the study's cross-sectional nature limits causal inferences that can be made.

## 8 CONCLUSION

Inclusivity needs to be valued not only at the initial stages of game design but also at a latter stage after participants play a polished version of the game. Our study highlights a large scale evaluation of participants' game play experience can reveal additional insights on how to make games that are inclusive. The analysis of player experience and learning outcomes can inform game design processes such as re-designing levels and learning activities [54, 115] which can facilitate an improved learning experience for all participants. Evaluating games can also help game designers better design feedback systems [64, 112], or including other engaging features such as verbal encouragement [78]. As a study on demographic factors in an educational game contexts, we recommend further work in this domain to create effective and engaging game-based learning experiences for all learners.

## ACKNOWLEDGMENTS

Authors were supported by NSERC, the NSF (#1542970), and QCRI-CSAIL.

## REFERENCES

[1] 2008. Blackwell Encyclopedia of Sociology. *Reference Reviews* 22, 3 (Jan. 2008), 18–18. https://doi.org/10.1108/09504120810859657
[2] 2019. Field of degree: Minorities. https://ncses.nsf.gov/pubs/nsf19304/digest/field-of-degree-minorities#blacks-or-african-americans
[3] 2019. Field of degree: Women. https://ncses.nsf.gov/pubs/nsf19304/digest/field-of-degree-women#computer-sciences
[4] 2019. Women, Minorities, and Persons with Disabilities in Science and Engineering: 2019 | NSF - National Science Foundation. https://ncses.nsf.gov/pubs/nsf19304/digest/field-of-degree-women#computer-sciences.
[5] Dmitry Alexandrovsky, Maximilian Achim Friehs, Jendrik Grittner, Susanne Putze, Max V. Birk, Rainer Malaka, and Regan L Mandryk. 2021. Serious Snacking: A Survival Analysis of How Snacking Mechanics Affect Attrition in a Mobile Serious Game. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–18. https://doi.org/10.1145/3411764.3445689
[6] Janice L Anderson and Mike Barnett. 2013. Learning physics with digital game simulations in middle school science. *Journal of science education and technology* 22, 6 (2013), 914–926.
[7] Leonard A Annetta. 2010. The "I's" Have It: A Framework for Serious Educational Game Design. *Review of general psychology* 14, 2 (2010), 105–113.

[8] Christian Bauckhage, Kristian Kersting, Rafet Sifa, Christian Thurau, Anders Drachen, and Alessandro Canossa. 2012. How Players Lose Interest in Playing a Game: An Empirical Study Based on Distributions of Total Playing Times. In *2012 IEEE Conference on Computational Intelligence and Games (CIG)*. IEEE, Granada, Spain, 139–146. https://doi.org/10.1109/CIG.2012.6374148

[9] Rachel Bellamy. 1994. What Does Pseudo-Code Do? A Psychological Analysis of the Use of Pseudo-Code by Experienced Programmers. *Human-Computer Interaction* 9, 2 (June 1994), 225–246. https://doi.org/10.1207/s15327051hci0902_3

[10] Björn Berg Marklund. 2015. *Unpacking Digital Game-Based Learning: The Complexities of Developing and Using Educational Games.* Ph. D. Dissertation. University of Skövde, Skövde.

[11] Adam J Berinsky, Gregory A Huber, and Gabriel S Lenz. 2012. Evaluating online labor markets for experimental research: Amazon. com's Mechanical Turk. *Political analysis* 20, 3 (2012), 351–368.

[12] Max V. Birk, Cheralyn Atkins, Jason T. Bowey, and Regan L. Mandryk. 2016. Fostering Intrinsic Motivation through Avatar Identification in Digital Games. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, San Jose California USA, 2982–2995. https://doi.org/10.1145/2858036.2858062

[13] Max V. Birk and Regan L. Mandryk. 2018. Combating Attrition in Digital Self-Improvement Programs Using Avatar Customization. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, Montreal QC Canada, 1–15. https://doi.org/10.1145/3173574.3174234

[14] Jim Bizzocchi and Theresa Jean Tanenbaum. 2012. *Mass Effect 2*: A Case Study in the Design of Game Narrative. *Bulletin of Science, Technology & Society* 32, 5 (Oct. 2012), 393–404. https://doi.org/10.1177/0270467612463796

[15] Jim Bizzochi. 2007. Games and Narrative: An Analytical Framework. *Loading-The Journal of the Canadian Games Studies Association* 1, 1 (2007).

[16] K.A. Blocker, T.J. Wright, and W.R. Boot. 2014. Gaming Preferences of Aging Generations. *Gerontechnology* 12, 3 (June 2014), 174–184. https://doi.org/10.4017/gt.2014.12.3.008.00

[17] Elizabeth A. Boyle, Thomas Hainey, Thomas M. Connolly, Grant Gray, Jeffrey Earp, Michela Ott, Theodore Lim, Manuel Ninaus, Claudia Ribeiro, and João Pereira. 2016. An Update to the Systematic Literature Review of Empirical Evidence of the Impacts and Outcomes of Computer Games and Serious Games. *Computers & Education* 94 (March 2016), 178–192. https://doi.org/10.1016/j.compedu.2015.11.003

[18] Michael Buhrmester, Tracy Kwang, and Samuel D Gosling. 2016. Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality data? (2016).

[19] Ricardo Caceffo, Pablo Frank-Bolton, Renan Souza, and Rodolfo Azevedo. 2019. Identifying and Validating Java Misconceptions toward a CS1 Concept Inventory. In *Proceedings of the 2019 ACM Conference on Innovation and Technology in Computer Science Education*. 23–29.

[20] Ibrahim Cetin. 2015. Students' Understanding of Loops and Nested Loops in Computer Programming: An APOS Theory Perspective. *Canadian Journal of Science, Mathematics and Technology Education* 15, 2 (April 2015), 155–170. https://doi.org/10.1080/14926156.2015.1014075

[21] Paul Chandler and John Sweller. 1991. Cognitive Load Theory and the Format of Instruction. *Cognition and Instruction* 8, 4 (Dec. 1991), 293–332. https://doi.org/10.1207/s1532690xci0804_2

[22] Sapna Cheryan, Victoria C Plaut, Paul G Davies, and Claude M Steele. 2009. Ambient Belonging: How Stereotypical Cues Impact Gender Participation in Computer Science. *Journal of personality and social psychology* 97, 6 (2009), 1045.

[23] Hongmei Chi, Edward Agama, and Zornitza Genova Prodanoff. 2017. Developing Serious Games to Promote Cognitive Abilities for the Elderly. In *2017 IEEE 5th International Conference on Serious Games and Applications for Health (SeGAH)*. IEEE, 1–8.

[24] Klimmt Christoph, Hefner Dorothée, and Vorderer Peter. 2009. The Video Game Experience as "True" Identification: A Theory of Enjoyable Alterations of Players' Self-Perception. *Communication Theory* 19, 4 (Nov. 2009), 351–373. https://doi.org/10.1111/j.1468-2885.2009.01347.x

[25] Kaitlan Chunhui Chu. 2004. *Gender Reactions to Games for Learning among Fifth and Eighth Graders.* Citeseer.

[26] Thomas M. Connolly, Elizabeth A. Boyle, Ewan MacArthur, Thomas Hainey, and James M. Boyle. 2012. A Systematic Literature Review of Empirical Evidence on Computer Games and Serious Games. *Computers & Education* 59, 2 (Sept. 2012), 661–686. https://doi.org/10.1016/j.compedu.2012.03.004

[27] R Dennis Cook and Sanford Weisberg. 2009. *Applied regression including computing and graphics*. Vol. 488. John Wiley & Sons.

[28] Brent Cowan and Bill Kapralos. 2014. A Survey of Frameworks and Game Engines for Serious Game Development. In *2014 IEEE 14th International Conference on Advanced Learning Technologies*. IEEE, Athens, Greece, 662–664. https://doi.org/10.1109/ICALT.2014.194

[29] James J. Cummings and Jeremy N. Bailenson. 2016. How Immersive Is Enough? A Meta-Analysis of the Effect of Immersive Technology on User Presence. *Media Psychology* 19, 2 (April 2016), 272–309. https://doi.org/10.1080/15213269.2015.1015740

[30] Stephen Davies. 2008. The Effects of Emphasizing Computational Thinking in an Introductory Programming Course. In *2008 38th Annual Frontiers in Education Conference*. IEEE, Saratoga Springs, NY, USA, T2C–3–T2C–8. https://doi.org/10.1109/FIE.2008.4720362

[31] Ana Fernández del Río, Pei Pei Chen, and África Periáñez. 2019. Profiling Players with Engagement Predictions. *arXiv* (2019), 2–5.

[32] Simon Demediuk, Alexandra Murrin, David Bulger, Michael Hitchens, Anders Drachen, William L Raffe, and Marco Tamassia. 2018. Player Retention In League of Legends: A Study Using Survival Analysis. (2018), 9.

[33] Guillaume Denis and Pierre Jouvelot. 2005. Motivation-Driven Educational Game Design: Applying Best Practices to Music Education. In *Proceedings of the 2005 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology - ACE '05*. ACM Press, Valencia, Spain, 462–465. https://doi.org/10.1145/1178477.1178581

[34] Peter J. Denning. 2004. The Field of Programmers Myth. *Commun. ACM* 47, 7 (2004), 15–20.

[35] Betsy DiSalvo and Amy Bruckman. 2010. Race and Gender in Play Practices: Young African American Males. In *Proceedings of the Fifth International Conference on the Foundations of Digital Games*. 56–63.

[36] Kevin Doherty and Gavin Doherty. 2019. Engagement in HCI: Conception, Theory and Measurement. *Comput. Surveys* 51, 5 (Jan. 2019), 1–39. https://doi.org/10.1145/3234149

[37] Edward Downs, Nicholas D. Bowman, and Jaime Banks. 2019. A Polythetic Model of Player-Avatar Identification: Synthesizing Multiple Mechanisms. *Psychology of Popular Media Culture* 8, 3 (July 2019), 269–279. https://doi.org/10.1037/ppm0000170

[38] Anders Drachen, Pejman Mirza-Babaei, and Lennart E. Nacke (Eds.). 2018. *Games User Research* (first edition ed.). Oxford University Press, Oxford, UK ; New York.

[39] Tobias Drey, Fabian Fischbach, Pascal Jansen, Julian Frommel, Michael Rietzler, and Enrico Rukzio. 2021. To Be or Not to Be Stuck, or Is It a Continuum? A Systematic Literature Review on the Concept of Being Stuck in Games. *Proceedings of the ACM on Human-Computer Interaction* 5, CHI PLAY (2021), 1–35.

[40] Benedict Du Boulay. 1986. Some Difficulties of Learning to Program. *Journal of Educational Computing Research* 2, 1 (1986), 57–73.

[41] Gunther Eysenbach. 2005. The Law of Attrition. *Journal of Medical Internet Research* 7, 1 (March 2005), e11. https://doi.org/10.2196/jmir.7.1.e11

[42] Mary Flanagan and Helen Nissenbaum. 2007. A Game Design Methodology to Incorporate Social Activist Themes. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, San Jose California USA, 181–190. https://doi.org/10.1145/1240624.1240654

[43] Jesse Fox, Sun Joo (Grace) Ahn, Joris H. Janssen, Leo Yeykelis, Kathryn Y. Segovia, and Jeremy N. Bailenson. 2015. Avatars Versus Agents: A Meta-Analysis Quantifying the Effect of Agency on Social Influence. *Human–Computer Interaction* 30, 5 (Sept. 2015), 401–432. https://doi.org/10.1080/07370024.2014.921494

[44] Stuart Garner. 2007. A Program Design Tool to Help Novices Learn Programming. *ICT: Providing choices for learners and learning* 321 (2007), 324.

[45] James Paul Gee. 2003. What Video Games Have to Teach Us about Learning and Literacy. *Computers in Entertainment (CIE)* 1, 1 (2003), 20–20.

[46] James Paul Gee. 2007. *Good Video Games+ Good Learning: Collected Essays on Video Games, Learning, and Literacy.* Peter Lang.

[47] David Ginat. 2004. On Novice Loop Boundaries and Range Conceptions. *Computer Science Education* 14, 3 (Sept. 2004), 165–181. https://doi.org/10.1080/0899340042000302709

[48] Marcello A. Gómez-Maureira, Michelle Westerlaken, Dirk P. Janssen, Stefano Gualeni, and Licia Calvi. 2014. Improving Level Design through Game User Research: A Comparison of Methodologies. *Entertainment Computing* 5, 4 (Dec. 2014), 463–473. https://doi.org/10.1016/j.entcom.2014.08.008

[49] Kishonna L. Gray. 2012. INTERSECTING OPPRESSIONS AND ONLINE COMMUNITIES: Examining the Experiences of Women of Color in Xbox Live. *Information, Communication & Society* 15, 3 (April 2012), 411–428. https://doi.org/10.1080/1369118X.2011.642401

[50] Bradley S. Greenberg, John Sherry, Kenneth Lachlan, Kristen Lucas, and Amanda Holmstrom. 2010. Orientations to Video Games Among Gender and Age Groups. *Simulation & Gaming* 41, 2 (April 2010), 238–259. https://doi.org/10.1177/1046878108319930

[51] Mark D. Griffiths, Mark N.O. Davies, and Darren Chappell. 2004. Demographic Factors and Playing Variables in Online Computer Gaming. *CyberPsychology & Behavior* 7, 4 (Aug. 2004), 479–487. https://doi.org/10.1089/cpb.2004.7.479

[52] Shuchi Grover and Satabdi Basu. 2017. Measuring Student Learning in Introductory Block-Based Programming: Examining Misconceptions of Loops, Variables, and Boolean Logic. In *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education*. ACM, Seattle Washington USA, 267–272. https://doi.org/10.1145/3017680.3017723

[53] Juho Hamari and Veikko Eranti. 2011. Framework for Designing and Evaluating Game Achievements.. In *Digra Conference*, Vol. 10. Citeseer, 9966.

[54] Erik Harpstead and Vincent Aleven. 2015. Using Empirical Learning Curve Analysis to Inform Design in an Educational Game. (Oct. 2015), 197–207. https://doi.org/10.1145/2793107.2793128

[55] Erik Harpstead, Brad A. Myers, and Vincent Aleven. 2013. In search of learning: Facilitating data analysis in educational games. In *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/2470654.2470667

[56] D Fox Harrell. 2009. Toward a theory of critical computing: The case of social identity representation in digital media applications. (2009).

[57] Casper Harteveld, Nithesh Javvaji, Tiago Machado, Yevgeniya V. Zastavker, Victoria Bennett, and Tarek Abdoun. 2020. Gaming4All: Reflecting on Diversity, Equity, and Inclusion for Game-Based Engineering Education. In *2020 IEEE Frontiers in Education Conference (FIE)*. IEEE, Uppsala, Sweden, 1–9. https://doi.org/10.1109/FIE44824.2020.9274176

[58] Bruce D. Homer, Elizabeth O. Hayward, Jonathan Frye, and Jan L. Plass. 2012. Gender and Player Characteristics in Video Game Play of Preadolescents. *Computers in Human Behavior* 28, 5 (Sept. 2012), 1782–1789. https://doi.org/10.1016/j.chb.2012.04.018

[59] Britton Horn, Amy K. Hoover, Jackie Barnes, Yetunde Folajimi, Gillian Smith, and Casper Harteveld. 2016. Opening the Black Box of Play: Strategy Analysis of an Educational Game. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play*. ACM, Austin Texas USA, 142–153. https://doi.org/10.1145/2967934.2968109

[60] Robin Hunicke, Marc LeBlanc, and Robert Zubek. 2004. MDA: A Formal Approach to Game Design and Game Research. In *Proceedings of the AAAI Workshop on Challenges in Game AI*, Vol. 4. San Jose, CA, 1722.

[61] Roziana Ibrahim. 2011. A Conceptual Framework for Supporting Gender Inclusivity in Games. (2011).

[62] Jeroen Jansz, Corinne Avis, and Mirjam Vosmeer. 2010. Playing The Sims2: An Exploration of Gender Differences in Players' Motivations and Patterns of Play. *New Media & Society* 12, 2 (March 2010), 235–251. https://doi.org/10.1177/1461444809342267

[63] D. Jary and J. Jary. 2000. *Collins Dictionary of Sociology*. HarperCollins.

[64] Sandeep M Jayaprakash, Erik W Moody, Eitel JM Lauría, James R Regan, and Joshua D Baron. 2014. Early alert of academically at-risk students: An open source analytics initiative. *Journal of Learning Analytics* 1, 1 (2014), 6–47.

[65] David G. Jenkins and Pedro F. Quintana-Ascencio. 2020. A Solution to Minimum Sample Size for Regressions. *PLOS ONE* 15, 2 (Feb. 2020), e0229345. https://doi.org/10.1371/journal.pone.0229345

[66] Xina Jiang, Casper Harteveld, Xinyuan Huang, and Anthony Y. H. Fung. 2019. The Computational Puzzle Design Framework: A Design Guide for Games Teaching Computational Thinking. In *Proceedings of the 14th International Conference on the Foundations of Digital Games*. ACM, San Luis Obispo California USA, 1–11. https://doi.org/10.1145/3337722.3337768

[67] Juho Hamari and Lauri Keronen. 2017. Why Do People Play Games? A Meta-Analysis. *International Journal of Information Management* (2017). https://doi.org/10.1016/j.ijinfomgt.2017.01.006

[68] Yasmin B. Kafai, Melissa S. Cook, and Deborah A. Fields. 2010. "Blacks Deserve Bodies Too!": Design and Discussion About Diversity and Race in a Tween Virtual World. *Games and Culture* 5, 1 (Jan. 2010), 43–63. https://doi.org/10.1177/1555412009351261

[69] Dominic Kao. 2018. *Researching and developing the impacts of virtual identity on computational learning environments*. Ph. D. Dissertation. Massachusetts Institute of Technology.

[70] Dominic Kao. 2019. Infinite Loot Box: A Platform for Simulating Video Game Loot Boxes. *IEEE Transactions on Games* 12, 2 (2019), 219–224.

[71] Dominic Kao. 2019. JavaStrike: A Java Programming Engine Embedded in Virtual Worlds. In *Proceedings of the 14th International Conference on the Foundations of Digital Games* (San Luis Obispo California USA). ACM, 1–5. https://doi.org/10.1145/3337722.3341828

[72] Dominic Kao. 2020. The effects of juiciness in an action RPG. *Entertainment Computing* 34 (2020), 100359.

[73] Dominic Kao. 2020. Exploring Help Facilities in Game-Making Software. (2020), 14.

[74] Dominic Kao. 2021. The Effects of Observation in Video Games: How Remote Observation Influences Player Experience, Motivation, and Behaviour. *Behaviour & Information Technology* (April 2021), 1–23. https://doi.org/10.1080/0144929X.2021.1906321

[75] Dominic Kao and D Fox Harrell. 2015. Exploring the Impact of Role Model Avatars on Game Experience in Educational Games. 571–576.

[76] Dominic Kao and D Fox Harrell. 2015. Mazzy: A STEM Learning Game.. In *FDG*.

[77] Dominic Kao and D Fox Harrell. 2016. Exploring the Effects of Dynamic Avatar on Performance and Engagement in Educational Games. (2016).

[78] Dominic Kao and D Fox Harrell. 2016. Exploring the effects of encouragement in educational games. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. 1906–1914.

[79] Dominic Kao and D Fox Harrell. 2016. Exploring the impact of avatar color on game experience in educational games. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. 1896–1905.

[80] Dominic Kao and D Fox Harrell. 2017. MazeStar: A platform for studying virtual identity and computer science education. In *proceedings of the 12th international conference on the foundations of digital games*. 1–6.

[81] Dominic Kao and D Fox Harrell. 2018. The effects of badges and avatar identification on play and making in educational games. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–19.

[82] Dominic Kao, Amogh Joshi, Christos Mousas, Abhigna Peddireddy, Arjun Kramadhati Gopi, Jianyao Li, John Springer, Bethany S. McGowan, and Jason B. Reed. 2022-05-06. Fighting COVID-19 at Purdue University: Design and Evaluation of a Game for Teaching COVID-19 Hygienic Best Practices. In *The 16th International Conference on the Foundations of Digital Games (FDG) 2021* (Montreal QC Canada, 2021-08-03). ACM, 1–23. https://doi.org/10.1145/3472538.3472552

[83] Dominic Kao, Alejandra J. Magana, and Christos Mousas. 2021. Evaluating Tutorial-Based Instructions for Controllers in Virtual Reality Games. *Proceedings of the ACM on Human-Computer Interaction* 5, CHI PLAY (Oct. 2021), 1–28. https://doi.org/10.1145/3474661

[84] Dominic Kao, Christos Mousas, Alejandra J Magana, D Fox Harrell, Rabindra Ratan, Edward F Melcer, Brett Sherrick, Paul Parsons, and Dmitri A Gusev. 2020. Hack.VR: A Programming Game in Virtual Reality. (2020), 6.

[85] Dominic Kao, Rabindra Ratan, Christos Mousas, Amogh Joshi, and Edward F Melcer. 2022. Audio Matters Too: How Audial Avatar Customization Enhances Visual Avatar Customization. In *CHI Conference on Human Factors in Computing Systems*. 1–27.

[86] Dominic Kao, Rabindra Ratan, Christos Mousas, and Alejandra J Magana. 2021. The Effects of a Self-Similar Avatar Voice in Educational Games. *Proceedings of the ACM on Human-Computer Interaction* 5, CHI PLAY (2021), 1–28.

[87] Fengfeng Ke, Valerie Shute, Kathleen M. Clark, and Gordon Erlebacher. 2019. *Interdisciplinary Design of Game-based Learning Platforms: A Phenomenological Examination of the Integrative Design of Game, Learning, and Assessment*. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-030-04339-1

[88] Mable B. Kinzie and Dolly R. D. Joseph. 2008. Gender Differences in Game Activity Preferences of Middle School Children: Implications for Educational Game Design. *Educational Technology Research and Development* 56, 5-6 (Dec. 2008), 643–663. https://doi.org/10.1007/s11423-007-9076-z

[89] Ulrich Knief and Wolfgang Forstmeier. 2021. Violating the Normality Assumption May Be the Lesser of Two Evils. *Behavior Research Methods* 53, 6 (Dec. 2021), 2576–2590. https://doi.org/10.3758/s13428-021-01587-5

[90] Sabine C. Koch, Stephanie M. Müller, and Monika Sieverding. 2008. Women and Computers. Effects of Stereotype Threat on Attribution of Failure. *Computers & Education* 51, 4 (Dec. 2008), 1795–1803. https://doi.org/10.1016/j.compedu.2008.05.007

[91] Margaret Law. 2006. Reduce, Reuse, Recycle: Issues in the Secondary Use of Research Data. *IASSIST Quarterly* 29, 1 (Feb. 2006), 5. https://doi.org/10.29173/iq599

[92] Debra A Lieberman, Maria Chesley Fisk, and Erica Biely. 2009. Digital Games for Young Children Ages Three to Six: From Research to Design. *Computers in the Schools* 26, 4 (2009), 299–313.

[93] Lorraine Lin, Dhaval Parmar, Sabarish V. Babu, Alison E. Leonard, Shaundra B. Daily, and Sophie Jörg. 2017. How Character Customization Affects Learning in Computational Thinking. In *Proceedings of the ACM Symposium on Applied Perception*. ACM, Cottbus Germany, 1–8. https://doi.org/10.1145/3119881.3119884

[94] Anna Lindqvist, Marie Gustafsson Sendén, and Emma A. Renström. 2021. What Is Gender, Anyway: A Review of the Options for Operationalising Gender. *Psychology & Sexuality* 12, 4 (Oct. 2021), 332–344. https://doi.org/10.1080/19419899.2020.1729844

[95] Conor Linehan, George Bellord, Ben Kirman, Zachary H. Morford, and Bryan Roche. 2014. Learning Curves: Analysing Pace and Challenge in Four Successful Puzzle Games. In *Proceedings of the First ACM SIGCHI Annual Symposium on Computer-human Interaction in Play*. ACM, Toronto Ontario Canada, 181–190. https://doi.org/10.1145/2658537.2658695

[96] Derek Lomas, Kishan Patel, Jodi L. Forlizzi, Jodi L. Forlizzi, Jodi Forlizzi, and Kenneth R. Koedinger. 2013. Optimizing Challenge in an Educational Game Using Large-Scale Design Experiments. (April 2013), 89–98. https://doi.org/10.1145/2470654.2470668

[97] J. Scott Long and Laurie H. Ervin. 2000. Using Heteroscedasticity Consistent Standard Errors in the Linear Regression Model. *The American Statistician* 54, 3 (Aug. 2000), 217–224. https://doi.org/10.1080/00031305.2000.10474549

[98] Enrica Loria and Annapaola Marconi. 2021. Exploiting Limited Players' Behavioral Data to Predict Churn in Gamification. *Electronic Commerce Research and Applications* 47 (2021), 101057.

[99] Winter Mason and Siddharth Suri. 2012. Conducting behavioral research on Amazon's Mechanical Turk. *Behavior research methods* 44, 1 (2012), 1–23.

[100] Allison Master, Sapna Cheryan, and Andrew N. Meltzoff. 2016. Computing Whether She Belongs: Stereotypes Undermine Girls' Interest and Sense of Belonging in Computer Science. *Journal of Educational Psychology* 108, 3 (April 2016), 424–437. https://doi.org/10.1037/edu0000061

[101] Victoria McArthur and Jennifer Jenson. 2014. E Is for Everyone? Best Practices for the Socially Inclusive Design of Avatar Creation Interfaces. In *Proceedings of the 2014 Conference on Interactive Entertainment*. ACM, Newcastle NSW Australia, 1–8. https://doi.org/10.1145/2677758.2677783

[102] Victoria McArthur, Robert John Teather, and Jennifer Jenson. 2015. The Avatar Affordances Framework: Mapping Affordances and Design Trends in Character Creation Interfaces. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play*. ACM, London United Kingdom, 231–240. https://doi.org/10.1145/2793107.2793121

[103] Simon McCallum and Costas Boletsis. 2013. Dementia Games: A Literature Review of Dementia-Related Serious Games. In *Serious Games Development and Applications*, David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Madhu Sudan, Demetri Terzopoulos, Doug Tygar, Moshe Y. Vardi, Gerhard Weikum, Minhua Ma, Manuel Fradinho Oliveira, Sobah Petersen, and Jannicke Baalsrud Hauge (Eds.). Vol. 8101. Springer Berlin Heidelberg, Berlin, Heidelberg, 15–27. https://doi.org/10.1007/978-3-642-40790-1_2

[104] Michael C Medlock. 2018. An overview of gur methods. *Games User Research* (2018).

[105] Kathryn J Moland. 2011. Decline of US Student Enrollment in Computer Science Programs. IEEE, 297–299.

[106] Lisa Nakamura. 2013. *Cybertypes: Race, Ethnicity, and Identity on the Internet*. Routledge.

[107] Nintendo and Niantic. 2016. *Pokemon Go*. Game [Mobile]. Nintendo, Kyoto, Japan. Last played December 2020.

[108] Ogen Odisho, Mark Aziz, and Nasser Giacaman. 2016. Teaching and Learning Data Structure Concepts via Visual Kinesthetic Pseudocode with the Aid of a Constructively Aligned App. *Computer Applications in Engineering Education* 24, 6 (Nov. 2016), 926–933. https://doi.org/10.1002/cae.21768

[109] Anne L Olsen. 2005. USING PSEUDOCODE TO TEACH PROBLEM SOLVING. (2005), 6.

[110] Eleanor O'Rourke, Eric Butler, Yun-En Liu, Christy Ballweber, and Zoran Popovic. 2013. The Effects of Age on Player Behavior in Educational Games.. In *FDG*. 158–165.

[111] Stamatios Papadakis. 2018. The Use of Computer Games in Classroom Environment. *International Journal of Teaching and Case Studies* 9, 1 (2018), 1–25.

[112] Abelardo Pardo, Jelena Jovanovic, Shane Dawson, Dragan Gašević, and Negin Mirriahi. 2019. Using learning analytics to scale the provision of personalised feedback. *British Journal of Educational Technology* 50, 1 (2019), 128–138.

[113] Cale J. Passmore, Max V. Birk, and Regan L. Mandryk. 2018. The Privilege of Immersion: Racial and Ethnic Experiences, Perceptions, and Beliefs in Digital Gaming. (April 2018), 383. https://doi.org/10.1145/3173574.3173957

[114] Celia Pearce. 2008. The Truth About Baby Boomer Gamers: A Study of Over-Forty Computer Game Players. *Games and Culture* 3, 2 (April 2008), 142–174. https://doi.org/10.1177/1555412008314132

[115] Zhongxiu Peddycord-Liu, Rachel Harred, Sarah M. Karamarkovich, Tiffany Barnes, Collin F. Lynch, and Teomara Rutherford. 2018. Learning Curve Analysis in a Large-Scale, Drill-and-Practice Serious Math Game: Where Is Learning Support Needed? (June 2018), 436–449. https://doi.org/10.1007/978-3-319-93843-1_32

[116] Anton Petrov. 2014. Using Minecraft in Education: A Qualitative Study on Benefits and Challenges of Game-Based Education. (2014).

[117] Martin Pichlmair and Mads Johansen. 2020. Designing Game Feel. A Survey. *arXiv:2011.09201 [cs]* (Nov. 2020). arXiv:2011.09201 [cs]

[118] Johanna Pirker, André Rattinger, Anders Drachen, and Rafet Sifa. 2018. Analyzing Player Networks in Destiny. *Entertainment Computing* 25 (March 2018), 71–83. https://doi.org/10.1016/j.entcom.2017.12.001

[119] Donald E Polkinghorne. 1988. *Narrative Knowing and the Human Sciences*. Suny Press.

[120] Ralph T Putnam, Derek Sleeman, Juliet A Baxter, and Laiani K Kuspa. 1986. A Summary of Misconceptions of High School Basic Programmers. *Journal of Educational Computing Research* 2, 4 (1986), 459–472.

[121] Yolanda A. Rankin and India Irish. 2020. A Seat at the Table: Black Feminist Thought as a Critical Framework for Inclusive Game Design. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (Oct. 2020), 1–26. https://doi.org/10.1145/3415188

[122] Rabindra Ratan, David Beyea, Benjamin J. Li, and Luis Graciano. 2020. Avatar Characteristics Induce Users' Behavioral Conformity with Small-to-Medium Effect Sizes: A Meta-Analysis of the Proteus Effect. *Media Psychology* 23, 5 (Sept. 2020), 651–675. https://doi.org/10.1080/15213269.2019.1623698

[123] Gabriela T Richard. 2013. Designing Games That Foster Equity and Inclusion: Encouraging Equitable Social Experiences Across Gender and Ethnicity in Online Games. (2013), 6.

[124] Geoffrey G Roy. 2006. Designing and Explaining Programs with a Literate Pseudocode. *Journal on Educational Resources in Computing* 6, 1 (March 2006), 1. https://doi.org/10.1145/1217862.1217863

[125] Isabel Ruthotto, Quintin Kreth, Jillian Stevens, Clare Trively, and Julia Melkers. 2020. Lurking and Participation in the Virtual Classroom: The Effects of Gender, Race, and Age among Graduate Students in Computer Science. *Computers & Education* 151 (July 2020), 103854. https://doi.org/10.1016/j.compedu.2020.103854

[126] Robin Samuel, Manfred Max Bergman, and Sandra Hupka-Brunner. 2013. The Interplay between Educational Achievement, Occupational Success, and Well-Being. *Social Indicators Research* 111, 1 (2013), 75–96.

[127] Anurag Sarkar and Seth Cooper. 2019. Transforming Game Difficulty Curves Using Function Composition. (May 2019), 551. https://doi.org/10.1145/3290605.3300781

[128] Carsten Schulte and Maria Knobelsdorf. 2007. Attitudes towards Computer Science-Computing Experiences as a Starting Point and Barrier to Computer Science. In *Proceedings of the Third International Workshop on Computing Education Research - ICER '07*. ACM Press, Atlanta, Georgia, USA, 27. https://doi.org/10.1145/1288580.1288585

[129] Songwon Seo. 2006. A Review and Comparison of Methods for Detecting Outliers in Univariate Data Sets. (2006).

[130] David Williamson Shaffer, Kurt R. Squire, Richard Halverson, and James P. Gee. 2005. Video Games and the Future of Learning. *Phi Delta Kappan* 87, 2 (Oct. 2005), 105–111. https://doi.org/10.1177/003172170508700205

[131] Rafet Sifa, Christian Bauckhage, and Anders Drachen. 2014. The Playtime Principle: Large-scale Cross-Games Interest Modeling. In *2014 IEEE Conference on Computational Intelligence and Games*. IEEE, Dortmund, Germany, 1–8. https://doi.org/10.1109/CIG.2014.6932906

[132] Rafet Sifa, Christian Bauckhage, and Anders Drachen. 2014. The Playtime Principle: Large-scale cross-games interest modeling. *IEEE Conference on Computatonal Intelligence and Games, CIG* (2014). https://doi.org/10.1109/CIG.2014.6932906

[133] Emma Smith. 2008. *Using Secondary Data in Educational and Social Research*. Open University Press ; McGraw-Hill, Maidenhead, England; New York.

[134] Katta Spiel, Oliver L. Haimson, and Danielle Lottridge. 2019. How to Do Better with Gender on Surveys: A Guide for HCI Researchers. *Interactions* 26, 4 (June 2019), 62–65. https://doi.org/10.1145/3338283

[135] Kurt Squire. 2003. Video Games in Education. *Int. J. Intell. Games & Simulation* 2, 1 (2003), 49–62.

[136] Frankie Tam and Shashank Pawar. 2020. Emerging Design Factors in Game-based Learning: Incentives, Social Presence, and Identity Design. (2020), 367.

[137] R Core Team et al. 2013. R: A language and environment for statistical computing.

[138] Scott L Thomas and Ronald H Heck. 2001. Analysis of Large-Scale Secondary Data in Higher Education Research: Potential Perils Associated with Complex Sampling Designs. *Research in higher education* 42, 5 (2001), 517–540.

[139] Alexandra To, Joselyn McDonald, Jarrek Holmes, Geoff Kaufman, and Jessica Hammer. 2018. Character Diversity in Digital and Non-Digital Games. *Transactions of the Digital Games Research Association* 4, 1 (Oct. 2018). https://doi.org/10.26503/todigra.v4i1.84

[140] Gustavo F. Tondello and Lennart E. Nacke. 2019. Player Characteristics and Video Game Preferences. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. ACM, Barcelona Spain, 365–378. https://doi.org/10.1145/3311350.3347185

[141] Gustavo F. Tondello, Rina R. Wehbe, Rita Orji, Giovanni Ribeiro, and Lennart E. Nacke. 2017. A Framework and Taxonomy of Videogame Playing Preferences. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. ACM, Amsterdam The Netherlands, 329–340. https://doi.org/10.1145/3116595.3116629

[142] Travis R. Chance, Nick Little. 2014. *Heroes Wanted*. Game [Board Game]. Action Phase Games, Indiana, USA..

[143] Sabine Trepte and Leonard Reinecke. 2010. Avatar Creation and Video Game Enjoyment: Effects of Life-Satisfaction, Game Competitiveness, and Identification with the Avatar. *Journal of Media Psychology* 22, 4 (Jan. 2010), 171–184. https://doi.org/10.1027/1864-1105/a000022

[144] Selen Turkay and Charles K. Kinzer. 2014. The Effects of Avatar-Based Customization on Player Identification:. *International Journal of Gaming and Computer-Mediated Simulations* 6, 1 (Jan. 2014), 1–25. https://doi.org/10.4018/ijgcms.2014010101

[145] Sylke Vandercruysse, Judith ter Vrugte, Ton de Jong, Pieter Wouters, Herre van Oostendorp, Lieven Verschaffel, and Jan Elen. 2017. Content Integration as a Factor in Math-Game Effectiveness. *Educational Technology Research and Development* 65, 5 (Oct. 2017), 1345–1368. https://doi.org/10.1007/s11423-017-9530-5

[146] Monica Visani Scozzi, Ioanna Iacovides, and Conor Linehan. 2017. A Mixed Method Approach for Evaluating and Improving the Design of Learning in Puzzle Games. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. ACM, Amsterdam The Netherlands, 217–228. https://doi.org/10.1145/3116595.3116628

[147] William R. Watson and Jun Fang. 2012. PBL as a Framework for Implementing Video Games in the Classroom:. *International Journal of Game-Based Learning*

2, 1 (Jan. 2012), 77–89. https://doi.org/10.4018/ijgbl.2012010105
[148] Nick Yee. 2006. The Demographics, Motivations, and Derived Experiences of Users of Massively Multi-User Online Graphical Environments. *Presence: Teleoperators and virtual environments* 15, 3 (2006), 309–329.
[149] Achim Zeileis. 2004. Econometric Computing with HC and HAC Covariance Matrix Estimators. *Journal of Statistical Software* 11, 10 (2004). https://doi.org/10.18637/jss.v011.i10
[150] Stuart Zweben. 2019. Enrollment and Retention in US Computer Science Bachelor's Programs in 2016-17. 10, 4 (2019), 47–59.