A DATA-DRIVEN, PLAYER-CENTRIC APPROACH TO DESIGNING SPATIAL SKILL TRAINING VIDEO GAMES

BY

HELEN WAUCK

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Computer Science
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2020

Urbana, Illinois

Doctoral Committee:

Professor Brian Bailey, Chair
Associate Professor Craig Zilles
Teaching Associate Professor Geoffrey Herman
Associate Professor H. Chad Lane
Research Assistant Professor Gale Lucas, Institute for Creative Technologies

Abstract

Spatial skills are a subset of cognitive skills essential to success in many different STEM fields, including engineering, chemistry, geology, and computer science [1, 2]. Those with low spatial skill often struggle in introductory college-level STEM coursework and drop out of STEM majors. Fortunately, spatial skills are quite malleable. Video games present a particularly promising way of training spatial skills; certain commercial video games, such as *Medal of Honor*, *Portal 2*, and *Tetris*, have been empirically shown to improve players' spatial skills after just a few hours of training [3, 4, 5], and video games provide a motivational advantage over other forms of spatial skill training interventions since they are designed to be fun. However, other commercial games, such as the "brain-training" game *Lumosity*, seem to have no effect on players' spatial skills [5]. It is not clear what makes some games effective and others ineffective at spatial skill training, which makes it difficult to design game-based spatial skill training interventions to improve students' STEM proficiency and retention.

Prior work studying the effectiveness of game-based spatial skill training interventions is also limited by the fact that it does not take the their motivational appeal to their target audience into account - the main advantage games have over other kinds of training interventions. Low spatial skill students, who are disproportionately female, stand the most to gain from spatial skill training interventions through improved proficiency in STEM coursework, but are not targeted in the design or evaluation of spatial skill training games.

In this dissertation, I present a data-driven, player-centric approach to designing spatial skill training video games that contributes to our understanding of what game features may contribute to a game's effectiveness at training spatial skills and its motivational appeal to its critical target audience of low spatial skill students. First, I explain the design of *Homeworld Bound*, a game I designed as a testbed for evaluating the ability of different spatial game features to tap into players' spatial skills and demonstrate its effectiveness as a training intervention for children. In my first study, I demonstrate my data-driven approach to evaluating the effectiveness of specific game features at tapping into players' spatial skills by analyzing the relationship between player performance and spatial skill in *Homeworld Bound*. My results reveal that most of the levels in *Homeworld Bound* successfully tap into players' spatial skills and provide insights about how to fix the levels that do not.

In my second study, I take a player-centric approach to designing the player experience of spatial skill training games, investigating how demographic factors and gaming habits, preferences, and motivations predict spatial skill in young adults. I use my findings to develop a set of recommendations for designing spatial skill training games that appeal to low spatial skill young adults specifically. I then present a revised and improved version of *Homeworld Bound*, *Homeworld Bound: Redux*, and demonstrate how I incorporated my findings from my data-driven and player-centered research studies to improve the game's ability to tap into players' spatial skills and its motivational appeal for low spatial skill young adults. Finally, I present the results of a controlled training study I conducted in a large introductory STEM course for non-majors to evaluate the training effectiveness and motivational appeal of *Homeworld Bound: Redux* for low spatial skill college students. While I found no training effects of *Homeworld Bound: Redux* compared to alternative training programs after 70 minutes of training, I found that performance on certain levels of the game was correlated with spatial skill and that low spatial skill students were more intrinsically motivated to play the game than to complete a non-game spatial skill training intervention.

My findings in this dissertation contribute a deeper understanding of what features influence a game's effectiveness at training spatial skills and motivating low spatial skill students to play it. These findings can be used to inform the design of more effective and motivating spatial skill training interventions in the future to promote the development of skills critical for pursuing STEM majors and careers.

To Kristen, for your friendship and mentorship from the very beginning of my PhD journey to the very end.

Acknowledgments

I owe the completion of this dissertation to the many people who have provided advice, feedback, moral support, and material assistance to my dissertation work over the years.

First, I would like to thank my advisor, Professor Brian Bailey. You had my back even before I became your student, you helped me write my first conference paper, and you always made time to offer me feedback, suggestions, and advice that shaped my research over the past 6 years of my Ph.D. Thank you for taking me on as a student, supporting me, and providing the guidance I needed to finish strong.

To the rest of my committee members: thank you for your helpful suggestions, which improved the quality of my dissertation work and provided me with diverse perspectives from educational psychology, computer science education, and social psychology to shape this interdisciplinary dissertation project. Professor Chad Lane, thank you for introducing me to the field of academic games research in my first year through your amazing class and getting me started on the project that has become my dissertation. I am grateful to have been in touch with you and the rest of the educational psychology folks these past years. Professor Alex Kirlik, thank you for your insights on the cognitive psychology aspect of my project and providing the skeptical sanalysis that helped me strengthen my arguments and methods. Professor Craig Zilles, thank you for collaborating with me on the very last project in my dissertation and giving me the opportunity to conduct research in the context of your programming course. Who knows when I would have graduated if not for you! Professor Geoffrey Herman, thank you for bringing your engineering and computer science education expertise to bear in advising me on my dissertation project and helping to shape the direction of my final project. Professor Gale Lucas, thank you for taking me on as a summer research assistant 3 years ago and giving me the opportunity to conduct some really unique games research. It was a pleasure working with you and the rest of the folks at the Institute for Creative Technologies.

To all of the many undergraduate and graduate students who helped me work on my games and assisted with my research studies: Rebecca Teasdale, Jackie Huey, Xiaoyi Chen, Hanzhao Deng, Qixin Wang, Luke Lu, Robin Sheong, Alexander Dzurick, EJ Lee, Jiexin Lu, Yuqi Yao, Jamie Lee, India Owens, Sebastian Saraceno, James Yang, Jinyuan Li, Qingqing Yang, Zhengqi Fang, and Tiffany Li. Thanks especially to Ziang Xiao and Po-Tsung Chiu, who ran an entire study for me when I had to travel, and Nick Olenz, who singlehandedly built Exploration Mode and several crucial pieces of my games' infrastructure. This project would not have been possible without you all.

Thank you to Professor Brian Woodard, Angie Wolters, Ziang Xiao, Tiffany Li, and the rest of the SIIP spatial visualization team for collaborating with me on various spatial skill training projects and helping me expand the scope of my work to engineering students. I hope to collaborate on more spatial training programs with you someday!

Professor Elisa Mekler, thank you and the rest of the folks in the MMI group at the University of Basel for welcoming me to Switzerland for the summer and for your valuable advice on my research. It was a pleasure collaborating with you and I would love to do more in the future!

Special thanks to Dr. So Yoona Yoon, Professor Mary Hegarty, and Professor Michael Peters for allowing me to use their psychometric tests for my research, and for their advice about how best to use them.

I am also grateful for the support of the NSF Graduate Research Fellowship that supported my dissertation work and gave me the chance to collaborate with international researchers in Switzerland as part of the NSF GROW program.

Thank you, Professor Wai-Tat Fu, for believing in my games research, encouraging me to pursue it from the start, and your valuable feedback on my work.

To Professor Marion Scheepers, Professor Liljana Babinkostova, and my fellow Boise State REU cohort, thank you for that amazing summer in 2012 that convinced me to apply to graduate school and get my Ph.D.

To my undergraduate advisors, Professor Max Hailperin and Professor Mike Hvidsten, who supported me every step of the way towards graduate school.

To my wonderful labmates and friends in the Human-Computer Interaction (HCI) research area at the University of Illinois: Kristen Vaccaro, Amy Oetting, Grace and Eric Yen, Sanorita Dey, Hidy Kong, Robert Deloatch, Jennifer Kim, Motahhare Eslami, Wayne Wu, John Lee, Po-Tsung Chiu, India Owens, Sneha Krishna Kumaran, Patrick Crain, Mingkun Gao, Ziang Xiao, Emily Hastings, Gina Do, Sebastian Rodriguez, Farnaz Jahanbakhsh, Tiffany Li, Rick Barber, Joon Park, and Silas Hsu. Thank you for the constant emotional and moral support over these past 6 years.

To my non-HCI friends I met in the graduate program at the University of Illinois: Cecilia Mauceri, Daphne Tsatsoulis, Shane Rife, Amanda Bienz, Carl Pearson, Cassandra Jacobs, Jason Rock, Nate Bowman, Erin Carrier, Imani Palmer, Andrew Reisner, Paul Eller, Jonathan Ligo, Megan Emigh, Sean Thetallone, Doris Xin, Alli Nilles, Everett Hildenbrandt, Ally Kaminsky, Adam Stewart, Shant Boodaghians, Margaret Lawson, Sherry Yi, Ben Moy, Zane Ma, Deepak Kumar, Josh Leveillee, and Riccardo Paccagnella. Thank you for getting me to take frequent breaks from my research and socialize. And, of course, for the moral and emotional support that kept me going over the past 6 years.

To Monica Ste. Marie and Emily Hamberg - your long distance friendship means the world to me! Thank you for your support and for providing perspective from outside of academia to my life! Hopefully we can hang out more in person now that I'm graduating.

And finally, to my family, who are always there to listen, give advice, and help me through the ups and downs of graduate school. And for providing me a place to flop down in your backyard and lie in the sunshine with our cats whenever I needed a break.

Table of Contents

Chapter 1: Introduction

There is a huge demand for STEM expertise in today's increasingly technology-driven economy; in the United States, STEM jobs are currently growing at double the rate of non-STEM jobs [6]. Unfortunately, there are simply not enough students graduating from college with the necessary skill set to fill the need for scientists, engineers, computer programmers, mathematicians, and other STEM occupations, particularly in industry and government [7], and the gender gap in STEM disciplines remains a problem [8]. And this is despite countless interventions have been proposed and implemented with the goal of attracting more students, especially women and girls, to STEM fields [9]. While there are many factors that affect students' decisions to pursue STEM careers, recent research suggests that these interventions may be missing a key piece of the puzzle: a set of cognitive skills called *spatial skills*.

Spatial skills are a specific set of cognitive skills that we use every day to perceive and comprehend the spatial relationships between and within different objects in our environment, enabling us to use tools and navigate [10]. They are also one of the strongest predictors of future achievement in STEM coursework and STEM careers, independent of math and verbal ability [11, 2]. Spatial skills come into play in a variety of different tasks required in STEM coursework: imagining cross-sectional structures in geology [12, 13], translating between 2D and 3D representations of 3D objects in computer aided design for engineering [14, 15, 16, 17], reading x-ray and MRI images [18], and comprehension of large programs in computer science [19, 20].

Students with underdeveloped spatial skills often struggle in their STEM classes and become discouraged and drop out, deciding that a STEM major is too hard and not for them. Therefore, spatial skills may act as a gatekeeper of sorts for STEM disciplines, causing low spatial skill students to encounter too much difficulty in their coursework and drop out [10]. Furthermore, there is a consistently demonstrated gender gap in spatial skill mirroring the gender gap in STEM participation that begins as early as elementary school [21, 22, 23, 8]. Thus, it is important to train students' spatial skills from an early age in order to prevent low spatial skill from becoming a barrier to success in STEM coursework, majors, and future careers for otherwise interested, motivated students.

The research literature to date suggests that spatial skills are quite malleable. The two primary methods of training spatial skills in the research literature are workbook exercises and video games. A meta-analysis of research studies utilizing these two methods revealed that they tend to be equally effective with moderate effect sizes [10]. However, video games are a particularly good medium for spatial skill training interventions since they naturally bring the player into a visual environment where learning can occur in a self-directed manner. More importantly, though, video games are designed to harness the power of intrinsic motivation to keep players engaged without any extrinsic incentives like the cash payment or course credit typically used in spatial skill training interventions [24, 25]. Thus, they are likely to be more appealing to students than more traditional types of spatial training that utilize multiple choice questions and sketching exercises; currently, 79% of people under age 18 play video games of some type in their free time [26].

Several commercial video games have already demonstrated empirical effectiveness in spatial skill training, including the first person shooters *Medal of Honor: Pacific Assault* [3] and *Unreal Tournament 2004* [27], the first person puzzle game *Portal 2* [5], the platformer *Super Mario 64* [28], the ball rolling arcade platformer game *Marble Madness* [29], the puzzle game *Tetris* [4], and the driving game *Crazy Taxi* [30] (see Figure 1.1). However, other commercial games, such as the 3D ball rolling game *Ballance* [3] and even so-called "cognitive training" games designed specifically to train these skills like *Lumosity*, produce no training effects at all [5, 31] (see Figure 1.2). The fact that games developed specifically to train cognitive skills have failed to produce training effects, while the only games shown to be effective thus far are commercial games that were not even intended to train any specific cognitive ability suggests that the research community and video game developers do not truly understand what factors contribute to an effective training game for spatial skills. How are we to design effective spatial skill training games that help students gain the skills they need to be successful in STEM if we don't know why some games are effective and others are not?

One barrier to discovering this knowledge is the near exclusive use of commercial games in the spatial skill training literature. Using a pre-built, non-open source commercial game makes it very difficult or impossible to isolate and test specific features of interest, which in turn makes it very difficult to determine what features to include to ensure a game is effective at training. Another concern with the overwhelming use of commercial games in the spatial skill training game literature is that they do not scale well; they must be purchased individually and may not be affordable for public school districts on tight budgets trying to prepare their students for success in future STEM coursework and majors. A third and final problem with using only pre-built commercial games is that they are not designed to appeal specifically to those with low spatial skill, who stand the most to benefit from spatial skill training interventions in terms of persistence in STEM coursework and majors. Since both gender and pre-existing spatial skill have been shown to influence preferences for different kinds of games [32, 33, 34], a commercial game that trains spatial skills may not appeal at

Figure 1.1: Examples of commercial games successful at training spatial skills in empirical studies. Clockwise from upper left: Tetris, Portal 2, Crazy Taxi, Medal of Honor: Pacific Assault, Super Mario 64, and Marble Madness. All screenshots are shown under fair use doctrine.

Figure 1.2: Examples of commercial games unsuccessful at training spatial skills in empirical studies. Left: the Rotation Matrix minigame in the brain training game Lumosity, where players must track a pattern as it rotates. Right: the ball rolling game Ballance. All screenshots are shown under fair use doctrine.

all to the demographic that could benefit the most from spatial skill training: those with low spatial skill, who tend to be disproportionately women and girls [8]. Less motivation to play means a lower likelihood of persisting in playing long enough to see a significant and durable training effect, and thus the main advantage of game training is lost.

1.1 MY APPROACH

To address the limitations in the existing research literature on spatial skill training games, I use a combination of *data-driven* and *player-centric* design approaches inspired by *design-based research* in the field of education and drawn from my own background in human-computer interaction (HCI). Design-based research combines theory and practice in developing and evaluating designed learning interventions, using theory to inform the design of interventions, evaluating them quantitatively and qualitatively with feedback from real users in naturalistic settings, and using the evaluation results to inform theory and improve the intervention's design simultaneously [35, 36, 37]. Using this approach, I designed my own game as an intervention to train spatial skills, *Homeworld Bound*, and its later revision, *Homeworld Bound: Redux* (see Figure 1.3), with modular level structure to isolate different combinations of theoretically grounded spatial game features across different levels (Chapters 3 and 6). Both design-based research and user interface design in HCI use an iterative design approach [38, 37], which changes and refines the end product through repeated cycles of design and testing with users. The design process for *Homeworld Bound*, as described in Chapter 3, involved iterative cycles of design and evaluation as well. I conducted each evaluation in a naturalistic context (either as a school classroom activity or in a more informal children's museum setting) in line with the principles of design-based research [39, 37].

Figure 1.3: Screenshots of *Homeworld Bound: Redux*, the final version of my spatial skill training game. The game's premise is that the player is stranded on an alien planet and must collect parts from the environment with which to rebuild their spaceship. In Exploration Mode levels (upper and lower left), players navigate through a 3D world in first person, collecting parts from the environment. In Construction Mode levels (upper and lower right), players use the parts they collected in Exploration Mode to build useful items in 3D by rotating and aligning parts.

As part of this iterative, design-based approach to developing *Homeworld Bound*, I implemented automatic collection of detailed player behavior data, allowing me to take a data-driven approach to evaluating which combinations of spatial features may be tapping into players' spatial skills by analyzing the relationship between spatial skill and player performance in each level (Chapters 4 and 7) with finer granularity than has been done in any previous work. Seeing which levels tap into players' spatial skills and which do not can provide new and more detailed insights about what combinations of features may be more likely to produce a game effective at training spatial skills, simultaneously advancing the theoretical understanding of spatial cognition in video games and informing future implementations of game-based spatial skill training interventions.

However, designing a game that trains players' spatial skills is not enough. It must also be fun for its target audience, otherwise the motivational power of a game-based intervention is lost. In the later chapters of this dissertation (from Chapter 5 onwards), I focus on designing for a target audience of low spatial skill young adults. This focus led me to conduct an online survey study to understand what low spatial skill young adults look for in a game (Chapter 5), bringing my target population in as consultants and active participants in shaping the game's future design as advocated for in design-based research [35, 37]. I then made a new version of my game, *Homeworld Bound: Redux*, redesigned based on what I learned about the preferences and motivations of this target audience (Chapter 6). My final evaluation study of *Homeworld Bound: Redux* as a training intervention continued this design-based, player-centric approach; I evaluated not just the game's training effectiveness but also its intrinsic appeal to low spatial skill students and to demographic groups likely to have lower spatial skill: women and those with less prior gaming experience, in a naturalistic large classroom setting (Chapter 7).

1.2 CONTRIBUTIONS

This dissertation advances knowledge about how to design spatial skill training video games for both training effectiveness and desirability in their target population. In particular, this dissertation makes four main contributions:

- **A theoretical framework for mapping spatial game features to spatial skills.** In Chapter 3, I present a theoretically grounded approach to developing games and game features that tap into players' spatial skills. This approach is grounded in a theoretical framework I developed to map different dimensions of spatial skill to different game features derived from both the existing research literature on spatial skill typology and existing empirical evidence regarding which video games are effective at training players' spatial skills. I demonstrate the implementation of my theoretical framework in a functional, engaging children's computer game: *Homeworld Bound*. My small scale training study of the game finds that *Homeworld Bound* improves children's scores on a test of spatial skill over the course of four 45 minute sessions, indicating that a game designed using this framework can be an effective spatial skill training intervention. In addition, my analysis of the relationship between spatial skill and level-by-level performance for *Homeworld Bound* in Chapter 5 and for its revised version, *Homeworld Bound: Redux* in Chapter 7 provides evidence that each game feature in my theoretical framework, as implemented in the game, taps into players' spatial skills in different ways.

- **A data-driven approach to designing and evaluating spatial game features.** In Chapter 4, I take a data-driven approach to studying what game features may contribute to the effectiveness of a game at training spatial skills, using my spatial skill training game *Homeworld Bound* as a testbed. I present the first empirical study analyzing the potential of different game features to tap into the spatial skills of children ages 7-12, how this relationship differs by gender, and give practical recommendations for implementing these features in games to assess or train players' spatial skills based on the complex relationship between gender, spatial skill, and in-game behavior I found. My data-driven, level-by-level player behavior analysis reveals that many but not all levels in *Homeworld Bound* tap into children's spatial skills via some aspect of player performance. These findings allow me to hypothesize potential reasons for the lack of relationship between spatial skills and performance in certain levels, which in turn allow me to make the revisions to the game described in Chapter 6 to improve its ability to tap into players' spatial skill. In Chapter 7, my similar analysis of the relationship between player performance and spatial skill with adults playing *Homeworld Bound: Redux*, the revised version of *Homeworld Bound*, shows that no Exploration Mode levels and only about half of Construction Mode levels are effective at tapping into players' spatial skills. As in my study with children, getting these insights allows me to hypothesize potential reasons for the lack of association between player performance and spatial skill in various levels of the game and suggest potential fixes that might be implemented and tested in future work.

- **A player-centric approach to designing and evaluating spatial skill training games.** In Chapter 5, I present a player-centric approach to designing spatial skill training games. I conduct a survey study across three diverse populations of high school students and college-age adults to gain a deeper and more complex understanding of how the following demographic and prior gaming experience variables predicted pre-existing spatial skill: gender, socioeconomic status (SES), gaming habits, gaming preferences, and gaming motivations. I find that the only predictors of spatial skill were gender and population. While this finding is consistent with prior work, the lack of relationship between preference for action video games and spatial skill is not and suggests that this relationship may be absent for certain populations. I use the findings of this study to develop a set of design recommendations for spatial skill training games that align with the preferences of the low spatial skill populations I identified: *facilitate short gameplay sessions*, *promote simple fun and thrill*, and *focus on adventure and puzzle genres*.

I then continue this player-centric approach by implementing these recommendations in my own spatial skill training game, *Homeworld Bound*, to produce *Homeworld Bound: Redux*, a new version designed to be more appealing to low spatial skill young adults (Chapter 6). I evaluate *Homeworld Bound: Redux* in a controlled study conducted in the context of an introductory programming course in Chapter 7. My analysis of both the game's training effects and its intrinsic appeal relative to other interventions focus on low spatial skill students since they are the target audience of the game. I also look at how other demographic factors such as gender and prior gaming experience influence training effects and intrinsic appeal. The results of the study show that gender and prior gaming experience has no effect on training effects or intrinsic appeal. However, I find that low spatial skill students rated their intrinsic motivation to play *Homeworld Bound: Redux* higher than that to complete a non-game spatial training activity, indicating that the game succeeds in its goal of being more intrinsically motivating for low spatial skill students than a non-game intervention.

- ***Homeworld Bound: Redux* as a research and design tool for spatial skill training games.** In Chapters 3 and 4, I describe the design and evaluation of *Homeworld Bound*, a game I have designed as a testbed for analyzing the effectiveness of different game features in tapping into the spatial skills of children ages 7-12. In Chapter 6, I describe the design process I used for *Homeworld Bound: Redux*, a new version of the game I built for a target audience of high school students and college-age adults. The design of *Homeworld Bound: Redux* combines the theoretically grounded, data-driven, and player-centric approaches to game design I used in Chapters 3, 4, and 5. A playable version of the game and its source code are both easily accessible online so that it can be modified, used for future research studies, and set up in classrooms large or small for training interventions by the larger community of cognitive training game designers, researchers, and educators.

The findings underlying the four main contributions of my dissertation have been published in several ACM SIGCHI conference papers. My findings in Chapter 4 were published in the IUI 2017 conference paper *Untangling the Relationship Between Spatial Skills, Game Features, and Gender in a Video Game* [40]. My Chapter 5 findings were published in the CHI 2019 paper *A Player-Centric Approach to Designing Spatial Skill Training Games* [41].

The findings of Chapter 6 were published in the CHI PLAY 2019 poster paper *A Testbed for Fun and Effective Features in Spatial Skill Training Games* [42]. My findings in Chapter 7 have been submitted to the CHI PLAY 2020 conference as the paper *Evaluating a Game-Based Spatial Skill Training Intervention for Low Spatial Skill Students* and are currently under review.

The four contributions of my dissertation outlined in this chapter address the limitations of the current state of the research on spatial skill training by providing 1) the first theoretical framework for spatial features in video games, 2) a data-driven approach to selecting and identifying spatial game features successful at tapping into players' spatial skills, 3) a player-centric approach to designing spatial skill training games focusing on the player experience of their target audience, and 4) the first publicly accessible noncommercial spatial skill training game for use in training interventions and as a testbed for evaluating the effectiveness of various spatial game features at tapping into players' spatial skill and appealing to low spatial skill players, who stand the most to benefit from spatial skill training interventions.

The use of noncommercial spatial skill training interventions underlies all four of these contributions. Selecting and implementing specific theory-driven features in a modular fashion to isolate them in different game levels is not possible in a pre-built commercial game. In addition, collecting sufficiently detailed player behavior data on a level-by-level basis to detect which game features are successful at tapping into player's spatial skills is infeasible with commercial games, which do not have built-in automatic player data collection accessible to researchers. Commercial games also are not designed with low spatial skill players as the specific target audience, and so they may not appeal to this critical target population for spatial training interventions. In addition, commercial games are difficult to modify for future investigations of different combinations of game features and incur a financial cost on any school or research group wishing to use them as a training intervention. Finally, companies selling their own cognitive training or brain training games have a financial incentive to make their customers believe their game works, whether it actually does or not, whereas noncommercial cognitive training games reduce the incentive to exaggerate a game's training capabilities and the potential for customers to be scammed out of money. It is for these reasons that I believe research with noncommercial games is the future of not just game-based spatial skill training research, but game-based cognitive training more broadly.

Chapter 2: Related Work

This chapter gives a broad overview of the existing research literature related to spatial skills, their connection to STEM efficacy, their ability to be trained, and why video games in particular have the potential to be very effective spatial skill training tools. I also discuss the limitations of the existing research in the realm of video game-based spatial skills training that provide the motivation for the research described in this dissertation. This chapter contains a small amount of material from my previously published work on the topic of spatial skill training games [40, 41, 42].

2.1 DEFINING SPATIAL SKILLS: A TAXONOMY

The research literature has consistently demonstrated that a particular subset of cognitive skills called spatial reasoning skills are important for success in many different STEM disciplines. Defining exactly what spatial skills are has been driven mostly by the use and exploratory factor analysis of psychometric tests that require a variety of different mental visualization tasks (see [43] for a summary). Thus, spatial skills are often defined in terms of different subskills derived from these psychometric tests. While there is no strong consensus on what these subskills are, the most commonly agreed upon subskills are *spatial visualization*, which involves mentally manipulating or transforming visual stimuli (e.g. mental rotation, mental paper folding), and *spatial orientation*, which involves imagining a pattern or scene and how it might change when viewed from different egocentric perspectives [44].

Chatterjee proposed that spatial skills can be broken down into a 2x2 taxonomy along two axes (see Figure 2.1): *intrinsic-extrinsic* (information about a specific object versus information about the relations between a group of objects) and *static-dynamic* (fixed information versus information about how something is changing over time) [45]. In this taxonomy, *spatial visualization* (and thus mental rotation and mental paper folding) corresponds to the *intrinsic-dynamic* dimension, while perspective-taking most closely aligns with the *extrinsic-static* dimension. This taxonomy has gained gradual support in the research literature, with its endorsement by Newcombe and Shipley [46] and adoption by Uttal et al. in their meta-analysis of spatial skill training studies [10]. Empirical support for this categorization comes from numerous research studies on the distinction between extrinsic and intrinsic spatial skills [47, 48, 49, 50] as well as the distinction between static and dynamic spatial skills [51, 52]. Therefore, I adopt this 2x2 taxonomy as my working definition of spatial skills as well in this dissertation, summarizing spatial skills as a whole as *the ability to perceive* *and comprehend the spatial relationships both within and between different objects in one's environment.*

2.2 THE RELATIONSHIP BETWEEN SPATIAL SKILLS AND STEM

Spatial skill is correlated with performance on a variety of different tasks spanning many different STEM disciplines, such as three dimensional biology problems [53], practical anatomy classes [54], reading x-ray and MRI images [18], imagining cross-sectional structures in geology [12, 13], program comprehension in computer science [19, 20], and a variety of visualization skills necessary for computer aided design (CAD) in engineering, such as translating between 2D orthographic projections and corresponding 3D representations of objects [14, 15, 16, 17]. More broadly, over 50 years of longitudinal research involving over 400,000 participants has consistently shown across multiple datasets that spatial skills are one of the strongest predictors of future achievement in STEM coursework and STEM careers, independent of math and verbal ability [11, 2].

2.2.1 The Relationship Between Spatial Subskills and STEM

The vast majority of studies on the relationship between spatial skills and STEM use tasks from the *intrinsic-dynamic* quadrant of the spatial skills taxonomy described in Section 2.1 as measures of spatial skill. Thus, for most of the studies described in this dissertation, I used standardized tests of mental rotation (an *intrinsic-dynamic* spatial skill) as my measure of spatial skills. However, other quadrants of the 2x2 taxonomy may also be relevant to STEM performance. For instance, in a study by Hegarty et al., a measure of *extrinsic-static* spatial skill predicted the performance of first year dentistry students in restorative dentistry classes [55]. More generally, Uttal et al's meta-analysis results show that training in one quadrant of the 2x2 taxonomy often produces transfer effects to other quadrants [10]. In most of the studies I discuss in this dissertation, the risk of inducing mental fatigue in child participants (Chapter 4) as well as the time constraints produced by conducting my other studies online (Chapter 5) or in a classroom (Chapter 7) made adding a test of *extrinsic-static* spatial skills unfeasible. However, in my pilot study in Chapter 3, I was able to use measures of both *intrinsic-dynamic* and *extrinsic-static* spatial skill to achieve better coverage of spatial subskills likely to be relevant to STEM performance.

When developing the theoretical mapping between spatial subskills and spatial game features that informed the design of my spatial skill training game *Homeworld Bound* (Chapter 3), I included three out of the four spatial subskills in the 2x2 taxonomy (*intrinsic-*

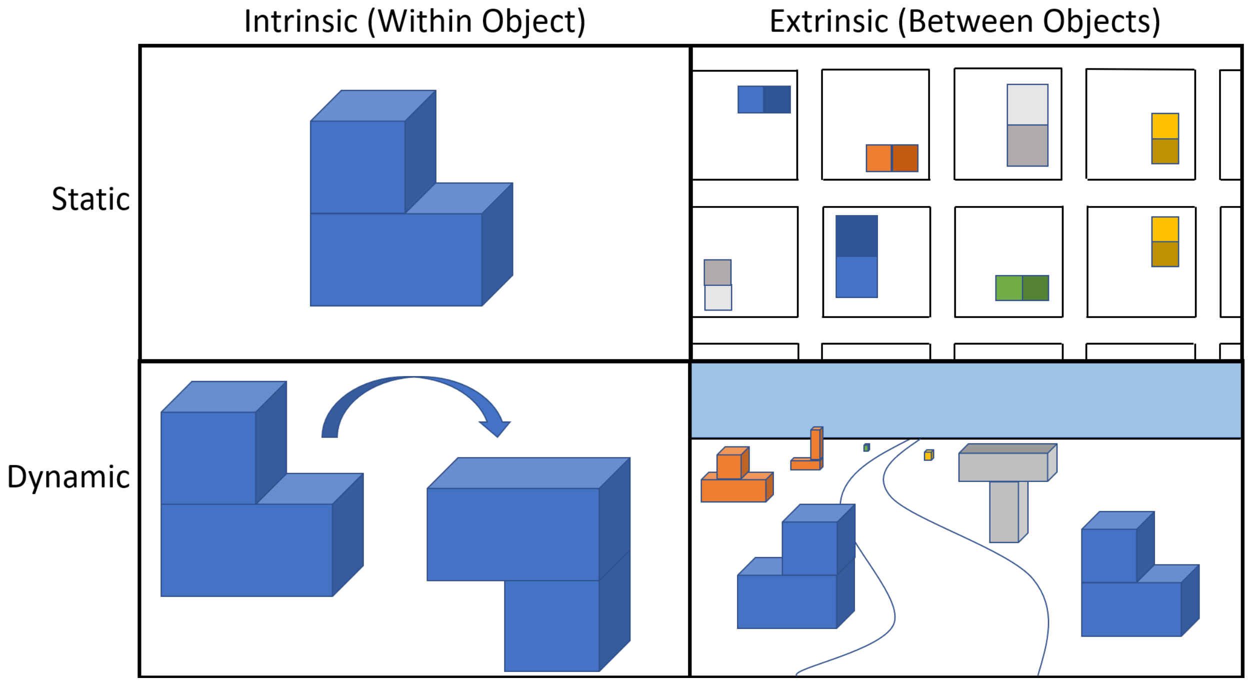


Figure 2.1: The 2 x 2 taxonomy of spatial skills, as proposed by Chatterjee [45].

dynamic, *extrinsic-static*, and *extrinsic-dynamic*). I excluded *intrinsic-static* quadrant from consideration in my theoretical framework due to the seeming inverse relationship it exhibited with STEM proficiency in one study by Kozhevnikov et al., which found that scientists and engineers performed much better on tests of *intrinsic-dynamic* spatial skill than *intrinsic-static* spatial skill, whereas the opposite pattern was true for artists [52].

2.2.2 Low Spatial Skills as Barrier to Entry to STEM Disciplines

Despite the strong association between spatial skills and STEM performance, spatial skills become less and less important in specific STEM disciplines such as geology, physics, and chemistry as students complete their STEM coursework, become more specialized, and move from novice to expert in their fields. This is likely due to the acquisition of domain-specific semantic knowledge that allows them to visualize 3D structures and spatial operations without actually performing mental operations [1], a more specific version of Hambrick et al.'s "circumvention-of-limits" hypothesis about the relationship between domain-specific knowledge and cognitive abilities [56, 57]. This hypothesis is supported by the results of several studies. For instance, Stieff found in a series of 3 experiments that students taking an organic chemistry class used mental rotation strategies to determine whether two molecular diagrams represented identical molecules or mirror images, whereas expert chemists used analytical strategies instead [58]. In the field of geology, Hambrick et al. found that spatial skill predicted performance in a bedrock mapping task only for those with low levels of geology knowledge [57].

Low spatial skills may therefore serve as a barrier to entry to STEM disciplines, causing students to avoid STEM majors or drop out of them before they have had a chance to develop domain-specific skills that render spatial skills less important [2, 59]. Attaining a certain level of spatial skill may be enough to help low spatial skill students persist and succeed in spatially-demanding introductory courses. Thus, as I and others have argued, low spatial skill students may benefit the most from spatial skill training interventions [1, 60, 41]. It is for this reason that my player-centric approach to designing the player experience of spatial skill training games in Chapter 5, my resulting revisions to *Homeworld Bound* to create an improved version, *Homeworld Bound: Redux*, and my evaluation of *Homeworld Bound: Redux* as a training intervention in Chapter 7 all focus on low spatial skill populations.

2.3 THE MOTIVATIONAL POWER OF VIDEO GAMES AS TRAINING TOOLS

Video games have been widely studied as interventions in the cognitive training literature, and this is also true of the spatial skill training literature more specifically. But why should game be the means of training spatial skill and not some other kind of training program? The appeal of video games is particularly strong for children and adolescents, who stand the most to gain from spatial skill training; 79% of them report playing video games of some kind in their free time [26]. Deci and Ryan's *self-determination theory* of motivation provides a good grounding theory for explaining the motivational appeal of video games.

Self-determination Theory (SDT) was developed around the concept that humans have an inherent tendency towards growth and have innate psychological needs which motivate their behavior: competence, autonomy, and relatedness. Of particular relevance to game-based training is the subtheory of SDT Deci and Ryan developed in 1985: *cognitive evaluation theory* (CET), which focuses on *competence* and *autonomy* as the primary factors influencing *intrinsic motivation*, motivation that is driven by inherent interest rather than *extrinsic* incentives such as a monetary reward [61, 62]. According to CET, people are intrinsically motivated to seek out optimally challenging activities by their need for competence, but feelings of competence and self-efficacy alone are not sufficient to promote intrinsic motivation. A person must also feel that it was their own decision to engage in a task - they must feel a sense of *autonomy* or *self-determination.* Feelings of competence are increased by an optimal level of challenge and constructive feedback, but diminished by negative feedback [63, 64].

Feelings of autonomy, meanwhile, are fostered via opportunities for choice and self-direction in an activity [61], and diminished by offering extrinsic incentives: tangible rewards like money or punishments like a bad grade in school [65].

Good video games naturally incorporate competence and autonomy-promoting experiences, which have been shown to promote motivation, enjoyment and feelings of well-being in players [25, 66]. Game designers strive to promote players' sense of competence and autonomy in order to keep players engaged and enjoying themselves. To do this, they *scaffold* the player's experience of progressing through a game. Scaffolding is a method rooted in the learning sciences literature and has been shown to be empirically effective at enhancing learning [67]; to scaffold a learning experience is to provide a high level of initial support, guidance, and feedback on the learner's actions, and then gradually remove this support as the learner progresses [68]. The scaffold helps the player progress initially, increasing their sense of competence, while its gradual removal increases their sense of autonomy. Specific examples of scaffolds in video games include tutorials and hints, which can be made adaptive to further tailor the scaffolding experience to the learner [69, 70, 71, 72]. Thus, video games promote intrinsic motivation in players through the use of competence and autonomy-promoting experiences.

Unfortunately, game-based spatial skill training studies lose some of this intrinsic motivation-boosting autonomy because they use prebuilt video games that are not designed with the target audience of low spatial skill populations in mind. This is a problem because as previously mentioned, those with low spatial skill also tend to spend less time playing with "spatial" toys and games during childhood [73]. Low spatial skill players may therefore not enjoy the kinds of out-of-the-box games they are asked to play during training studies, and this in turn may undermine their motivation to play their assigned game, decrease the effort they put into it, and limit the effectiveness of the training intervention.

Therefore, I argue for a player-centric approach [74, 75] to designing games for low spatial skill populations by asking directly for their input about what they like in a gaming experience. I apply this player-centric approach in Chapter 5 in an online study to identify the gaming habits, preferences, and motivations of low spatial skill populations, and then demonstrate how to apply this knowledge to the design of a spatial skill training game, *Homeworld Bound: Redux*, to improve the player experience and motivational appeal of the game for its target audience of low spatial skill young adults in Chapter 6. Finally, in Chapter 7, I evaluate not only the training effects of *Homeworld Bound: Redux*, but also its intrinsic appeal relative to other training interventions, with low spatial skill young adults.

2.4 SPATIAL SKILL AND DEMOGRAPHIC FACTORS

Prior research has found that certain demographic factors are associated with spatial skill level: gender, socioeconomic status, and frequency of action video gameplay. Each of these factors has the potential to affect both enjoyment of a game-based intervention and the game's training effectiveness.

2.4.1 Gender

One of the most well-known and consistent findings in the research literature on spatial skills is the gender gap. Multiple meta-analyses have found that women and girls score lower than men and boys on various standard tests of dynamic spatial skills such as mental rotation and spatial perception, with effect sizes ranging from 0.44 to 0.90 [21, 76, 77, 22]. These differences have been found to be consistent across age groups [21] and cultures [76], and emerge as early as elementary school (around age 9) [21, 23] or even in some cases as early as age 4 [78]. Gender differences have also been found to widen with age [76].

Several reasons have been proposed for this gender gap (for a good review, see [77]). Some research suggests that biological differences such as hormone levels or brain structure may cause gender differences in spatial skill. However, studies of the effect of sex hormones such as testosterone or estrogen on spatial skills in humans have found mixed effects [79, 80, 81, 82, 83, 84, 85, 86, 87], and studies on brain structure seem to indicate that gender differences are very task specific and do not generalize more broadly to spatial skills as a whole [88, 89, 90, 91, 92, 93, 94, 95]. Some evolutionary explanations have been proposed as well [96, 97], but cannot be empirically tested.

Another explanation for gender differences in the research literature with more consensus is that environmental factors, such as the frequency of participation in spatial activities during childhood, explains the development of boys' and girls' spatial skills. This hypothesis is supported by findings that boys are more likely to report having spatial experiences growing up than girls [98, 99] and that among men and women with high amounts of prior spatial experience, gender differences in spatial skill are small or nonsignificant [100, 101, 102, 73].

A third proposed reason for the gender gap in spatial skills is stereotype threat due to high spatial skill being perceived as a masculine trait. There are several examples in the research literature that support this hypothesis by studying priming effects. For instance, Sharps et al. showed in two studies that emphasizing the spatial nature of a spatial skills test, such as an object location memory task or a mental rotation task, increased the gender gap in spatial test scores, while there was no gender difference in test scores when the spatial nature of the task was not mentioned in the instructions [103, 104]. In another study, McGlone and Aronson found that when male and female undergraduates were primed to think of their gender identity, male students' performance increased, while female students' performance decreased [105]. Other research by Ariel et al. has found that women tend to report less confidence in their spatial skills than men even when no actual differences in performance are present on tests of spatial skills, and that women have less confidence in their self-assessments of spatial skills than men [106]. Even just telling study participants that women tend to perform better on a mental rotation task than men results in a narrower gender gap in performance on the mental rotation task [107].

In summary, the consensus in the research literature seems to be that prior spatial experience acquired while growing up is likely a primary cause of gender differences in spatial skill, but that stereotype threat exacerbates this difference. Inherent biological differences may contribute as well, but the evidence in this case appears rather inconsistent and difficult to empirically test. Thus, giving all children, regardless of gender, the opportunity to participate in spatial activities early on in life is likely to plant the seeds of achieving more gender equity in spatial skill performance later in life, in turn potentially reducing the influence of stereotype threat and helping to encourage more women to pursue and persist in STEM majors and careers.

Thus, I focus on analyzing gender effects throughout this dissertation with the goal of helping the research community, game designers, and educators design spatial skill training interventions that serve this important target audience of women and girls. Children were my target audience for *Homeworld Bound* and my initial evaluations of the game in Chapters 3 and 4 since this would enable training programs using the game to intervene and reduce the gender gap in spatial skills from an early age, before students start taking STEM electives and deciding on a major in college. In both studies, I focused part of my analysis on gender differences in training effects and in the extent to which each level of *Homeworld Bound* tapped in to children's spatial skills. This focus allowed me to discover important gender differences that suggested that girls may not have been served as well by my training intervention as boys, which in turn led me to conduct an online player research survey in Chapter 5 to better understand the preferences of my game's target audience of low spatial skill students (who are disproportionately female). My final evaluation of *Homeworld Bound: Redux* in Chapter 7 as a training intervention likewise involved the analysis of how gender influenced training effects and enjoyment of *Homeworld Bound: Redux* relative to other interventions.

2.4.2 Socioeconomic Status (and Gender)

Socioeconomic status (SES) is also related to spatial skill. For instance, Levine et al. found that children ages 7-9 with higher SES performed better on a spatial skills test than those with lower SES. Furthermore, SES and gender interacted in their study such that for low SES students, there was no gender difference in scores [108]. Another study by Noble et al. found that among first children ages 6-7, low SES populations had lower spatial skill than high SES populations [109]. Carr et al 2018 found in a longitudinal study of 304 elementary children that both gender and low SES predict low spatial skill and suggested that children with low SES or who are female and have poor visual working memory are good candidates for spatial skill interventions [110]. Verdine et al. found that this low SES disadvantage in spatial skills is present even as early as age 3 [111].

However, there is not yet any body of research I am aware of studying the relationship between SES, spatial skill, and gender in adults. To address this gap in the research literature, I investigate the relationship between demographic characteristics related to SES, gender, and spatial skills in Chapter 5 and develop design recommendations for improving the design of spatial skill training games to appeal to women and girls from low SES populations, the demographics most likely to have low spatial skill and thus the demographics with the most to gain from spatial skill training interventions. In Chapter 6, I implement these player-centric design recommendations in my spatial skill training game *Homeworld Bound: Redux* to improve the game's appeal to low SES women and girls.

2.4.3 Action Gaming Experience (and Gender)

There is also a substantial body of work investigating the relationship between prior gaming experience and spatial skills. Several studies have shown that action video game players (categorized based on their self-reported frequency of playing action games such as *Grand Theft Auto 3*, *Half-Life*, *Counter-Strike*, *Crazy Taxi*, *Team Fortress Classic*, *Spider-Man*, *Halo*, *Marvel vs Capcom*, *Roguespeare* and *Super Mario Kart*) tend to have higher spatial skill than those who do not play action games, performing better on various measures of spatial attention and visual search [112, 113, 114, 115, 116, 27, 117].

In addition, prior work has shown that men tend to play more action video games, such as shooting and sports games, which are are disproportionately represented in the spatial training literature (see Section 2.5), than women [118, 119, 34, 120]. Most video games are designed with a male target audience in mind and therefore the game content may not appeal as much to women and girls [118, 121]. Adding to this the well-documented gender gap in spatial skills in favor of men, it seems that those who stand the most to benefit from spatial skill training interventions - those with low spatial skill, who are disproportionately women and girls - will not be as motivated to play the kinds of action games that have been shown to be effective at training spatial skills in the research literature, such as *Unreal Tournament* and *Medal of Honor*.

To date, there exists only a single study that has investigated the relationship between spatial skills, video game play, and demographics together. Quaiser-Pohl et al. looked at the relationship between gender, game genre preferences, and spatial skill among secondary school students in Germany (ages 10-20). In addition to completing a paper test of spatial skill, students rated how frequently they played each of 8 different video game genres. Students were then grouped into three latent classes based on their stated genre frequencies: "non-players", "action-and-simulation game players", and "logic-and-skill-training players." Quaiser-Pohl et al. found that male action-and-simulation players had higher spatial skill test scores than male non-players, but there was no difference in spatial skill between different player classes for females [32]. More research of this kind, which examines multiple predictors of spatial skill, is needed in order to develop a more complete, up-to-date picture of the gaming preferences of low spatial skill populations that enables us to design game-based spatial skill training interventions that appeal to this critical target audience.

Gender and prior gaming experience may impact not just motivational appeal, but training effectiveness as well. Heterogeneity of treatment effects is common phenomenon in the educational research literature [122], and in the spatial skill training literature as well. The effect of gender is mixed; some studies have found that women and men improve equally with a game-based spatial skill training intervention [123, 124, 125], while others have found that women improve more than men [3, 126, 30]. There is currently little in the research literature about how prior gaming experience affects training effectiveness, although Terlecki et al. found that women who reported low levels of spatial activities (including computer/video game play) improved their spatial skills with a game intervention slower than men and women with high levels of spatial activities [4].

2.4.4 My Contribution

Given that demographic factors such as gender and action gaming experience are related to both spatial skill and gaming habits, it is critical to ensure that a game-based spatial skill training intervention is effective and motivating for demographic groups that may be less interested in playing the existing commercial action games that have been shown to be effective for a general population, especially demographic groups with lower spatial skill (women, girls, those with low SES, and those who do not play action video games) who stand the most to benefit from spatial skill training as a means of enabling future success in STEM coursework, majors, and careers. Therefore, in Chapter 5 of this dissertation, I investigate the demographic predictors of spatial skill using a large scale online survey that analyzes the relationship between gender, belonging to a low SES population, and gaming habits, preferences, and spatial skill simultaneously. Combining the predictors studied in different prior studies allows me to develop a more complex and specific model of low spatial skill populations and their gaming preferences, habits, and motivations, as does my sampling from three distinct and diverse populations. Understanding low spatial skill populations' gaming preferences, habits, and motivations, in turn, allows me to make design recommendations for game-based spatial skill training interventions targeting this critical target audience so that such interventions are more motivating and fun for them - the main advantage of using a game-based intervention. Without considering the appeal of a game to its target audience, the motivational advantage of game-based interventions is lost.

Another way this dissertation research builds on prior work on the relationship between demographic factors and spatial skill is by investigating the different appeal specific spatial skill training interventions may have for demographics that tend to have lower spatial skill, such as women and those who do not play action video games. Since those with low spatial skill stand the most to benefit from spatial skill training interventions, it is important that a spatial skill training intervention designed for them is evaluated with them as well. To investigate the extent to which different types of spatial skill training interventions produce training effects in and are appealing to low spatial skill demographics, I evaluate how gender and prior gaming experience influence the training effects and intrinsic appeal of my spatial skill training game *Homeworld Bound: Redux* compared to a more traditional workbook-based spatial skills training intervention in Chapter 7.

2.5 SPATIAL SKILL TRAINING

While the effectiveness of cognitive skill training in general has been questioned frequently in the last few years [127, 31], evidence for the effectiveness of spatial skill training in particular remains consistent across multiple meta-analyses. Uttal et al.'s meta-analysis of spatial skill training intervention studies analyzed all types of spatial skill training, the vast majority of which were either workbook exercises (often in the context of a course) or video games, and cited an average effect size of d=0.47 (moderate). This effect size was consistent across both workbook exercise training and game training. The results also demonstrated a moderate transfer effect of $d = 0.48$, indicating that training on one type of spatial task often transfers to other spatial tasks [10]. Sala et al.'s more recent meta-analysis of cognitive skill training in general focused exclusively on video game training and cited a more conservative effect size range of g=0.14-0.22 (small) for spatial skill training studies, with the specific effect size dependent on the type of game used (action, non-action, *Tetris*-like, or any kind of game) [31]. Below, I review the current state of the spatial skill training literature for both non-game interventions and game-based interventions.

2.5.1 Non-Game Spatial Skill Training Interventions

Non-game spatial skill training interventions have primarily been used in the context of STEM courses - particularly in engineering, where researchers have sought to develop training programs aimed at reducing the failure rate of students in introductory engineering courses, which tend to heavily emphasize inherently spatial exercises such as CAD modeling, visualizing cross sections, and isometric sketching. Hsi et al. designed one of the earliest interventions, a weekend spatial strategy tutorial, to supplement one such engineering course. They invited students who had received a low score on a spatial skills pre-course assessment to take part in the intervention. Students participating in the intervention improved their spatial skills, and no student failed the course that semester, compared to the 25% failure rate of previous semesters [15].

Sorby and Baartmans implemented a more comprehensive 10 week spatial skill training course for first year engineering students with low spatial skill involving many different types of computerized and paper and pencil spatial exercises, such as cross section visualization, revolving solids about an axis, mental rotation, and translating between 2D orthographic views and 3D views of an object. The results of multiple training studies using these exercises showed substantial spatial skill learning gains for both engineering and non-engineering students as well as lower dropout rates in early engineering courses [128, 129, 130, 131]. Miller and Halpern conducted a longitudinal study in which undergraduate STEM students were given 12 hours of spatial training using Sorby's workbook exercises. Compared to a randomized control group, the students who received the spatial training not only improved their spatial skills, but also had improved performance in an introductory physics course. However, these benefits did not persist when a follow-up analysis was conducted 8 months later [132].

Mostly digital training materials have been successful as well. For instance, Roca-González et al. found that 16 hours of augmented reality (AR) training on translating between orthographic 2D and 3D object views and virtual orienteering improved engineering students' spatial skills [133]. In addition, Onyancha et al. found that low spatial skill students improved their spatial skills relative to a control group after 4 hours of training with digital CAD, rotation, and drawing exercises over a 4 week period [134].

Non-game spatial skill training has also been implemented for children with the goal of increasing math performance. Cheng and Mix showed that practicing a mental rotation task improved first and second grade children's scores on mental rotation test and on fill-in-the-blank mathematical equation solving [135].

2.5.2 Spatial Skill Training with Video Games

Studies of game-based spatial skill training have focused almost exclusively on pre-built commercial games. Action video games in particular have been distinguished in the research literature as one of the more effective genres of games at training spatial skills. Green and Bavelier hypothesize that this is due to the combination of high speed gameplay and the requirement to quickly attend to multiple moving objects on the screen [136] typically present in action games. However, there is substantial disagreement in the research literature about what defines an action game, especially given the extent to which the games industry has changed in the last couple decades to create many new game genres [137, 138, 139, 140]. To date, the most consistently agreed upon and most widely studied action game genre in both academia [3, 114, 116, 141, 142, 143, 31] and industry [144, 145, 146] seems to be first person shooters.

First Person Shooters

One of the most commonly studied action games in the spatial skill training literature is the first person shooter *Medal of Honor*. Green and Bavelier had non-action game players play either *Medal of Honor: Allied Assault* or *Tetris* for 10 hours in 1 hour sessions across 10 consecutive days and found that the group playing *Medal of Honor* improved their visual attention skills, whereas the group playing *Tetris* did not [112]. Feng et al. conducted a training study using a newer game in the same *Medal of Honor* series: *Medal of Honor: Pacific Assault* and the 3D ball rolling puzzle game *Ballance* as an active control group. Both groups played their assigned game for 10 hours over four weeks. Participants in the *Medal of Honor* group improved on measures of both spatial attention and mental rotation, whereas participants in the *Ballance* group did not [3]. Boot et al. conducted a similar study where participants trained on *Medal of Honor: Allied Assault*, *Tetris*, or *Rise of Nations* (a real-time strategy game) for a total of 21.5 hours over the course of 4-5 weeks, while a passive control group did not complete any kind of training. All four groups were tested at the beginning of the study, in the middle, and at the end. In contrast to the findings of Green and Bavelier, playing *Medal of Honor* did not improve performance on any measure of visual attention, but the group playing *Tetris* did improve relative to other groups on a mental rotation task [141].

Other first person shooters have been studied as well. Green and Bavelier conducted another study in which participants played either the first person shooter *Unreal Tournament* or *Tetris* for a total of 30 hours over 4-6 weeks. Both groups improved the spatial resolution of visual processing, but the group playing *Unreal Tournament* improved more [27]. Choi and Lane used a similar study design (a total of 30 hours of training, with *Tetris* as the control group's game), but used a third person and a first person version of the shooter *BeGone* as the experimental groups. Only participants in the first person shooter group improved on measures of visual attention, and no groups improved on measures of navigational spatial skill [147]. In a study with 10-15 year old children, McClurg and Chaillé found that participants playing the first person spacecraft shooter *Stellar 7* for 9 hours over the course of six weeks improved on a measure of mental rotation compared to a passive control group [148]. *Battlezone*, a first person tank combat game very similar to *Stellar 7*, was also found to improve spatial orientation and mental rotation in undergraduate and graduate students after 5 hours total of training over a week compared to a no-contact control group [149].

Other Genres of Video Games

However, spatial skill training effects are not limited to first person shooters. Several other games from a wide variety of genres, some of which may fall into the somewhat ill-defined category of "action" game and some of which may not (see Chapter 5 and [137, 138, 139, 140] for a more thorough discussion) have also produced training effects in controlled studies. In the aforementioned study with *Stellar 7*, another experimental group played the puzzle game *The Factory*, which requires complex object manipulation, and found that players in this group also improved their mental rotation skills [148]. Similarly, in the aforementioned study with *Battlezone*, a different experimental group that played a top down arcade shooter, *Targ*, improved on measures of spatial orientation and mental rotation as well [149].

Platformer games, in which the player's primary concern is navigating safely to the end of a level without falling down into pits, also show potential as training interventions. A study conducted by Kühn et al. found significant gray matter increases in brain areas essential for spatial navigation in participants who played the platformer *Super Mario 64* for at least 30 hours over the course of 2 months relative to a passive control group [28]. Another platformer, the arcade game *Marble Madness*, has also been shown to produce training effects in 10-11 year olds on multiple measures of dynamic spatial skill after 135 minutes of practice, whereas no such improvement was present in another group of participants assigned to play the word game *Conjecture* instead [29].

Several racing and driving games have been shown to produce training effects as well. In a study by Cherney, participants trained on *Antz Extreme Racing* for 4 hours improved on a measure of mental rotation, while a control group competing crossword puzzles and logic games such as sudoku did not [126]. In a later study by Cherney et al. [30], female participants who played either the *Segway Circuit* minigame from *Wii Fit* or *Crazy Taxi*, both driving games requiring navigation, for just one hour improved their mental rotation skills relative to a control group competing the same sort of puzzle activities as in [126].

Other games' training effects present a more complex picture. The real time strategy (RTS) game *Rise of Nations* produced no spatial attention or mental rotation training effects in a mostly undergraduate participant pool in a study by Boot et al. after 21.5 hours of training [141]. However, in another study by Basak et al., after 23.5 hours of training with *Rise of Nations*, older adults improved on a measure of mental rotation relative to a no-contact control group [150].

The puzzle game *Tetris*, in which players must rotate and fit together falling blocks in completely filled rows to score points, presents a somewhat contradictory picture of training effects as well. In a study by Okagaki and Frensch, participants who played *Tetris* for 6 hours total across 12 sessions improved their reaction time on mental rotation and spatial visualization tasks relative to a no-contact control group [151]. Similarly, Cherney found that participants assigned to play *Tetrus* (a version of *Tetris* designed for the PC) for 4 hours improved a measure of mental rotation relative to a control group completing crossword puzzles and logic games [126]. On the other hand, Terlecki et al. found that 12 hours of training with *Tetris* over 12 weeks produced similar gains in mental rotation skill as in a control group that played *Solitaire*, which lasted several months after both interventions ceased. However, training effects for those who played *Tetris* transferred to other spatial tasks, and these transfer effects persisted several months later for the *Tetris* group as well.

In contrast to Terlecki et al.'s study, Sims and Mayer found that after 12 hours of training with *Tetris*, participants did not improve any more than a passive control group on any of several different mental rotation tests and other spatial skill tests. One reason for the stark contrast in results between this last study and the others may be the difference in sample size; Terlecki et al. had a total of 72 participants in the training condition and 108 in the control condition, Cherney had 20 per condition, Okagaki had around 25 per condition, but Sims and Mayer had less than 10 per condition. Thus, the effect of *Tetris* training may be detectable only with sample sizes larger than those in Sims and Mayer's study.

One last study worth mentioning was conducted by Shute et al., who assigned participants to play either the first person navigation puzzle game *Portal 2* or the commercial brain training game *Lumosity* for a total of 8 hours each, administering a battery of various cognitive tests before and after the intervention. While players of *Portal 2* improved on nearly every cognitive test (including tests of mental rotation and navigation), players in the *Lumosity* condition improved on none of these measures [5].

Some of these results may not be too surprising; after all, games shown to be effective at training spatial skills like *Medal of Honor: Pacific Assault*, *Portal 2*, *Super Mario 64*, *Crazy Taxi*, *Battlezone*, *Stellar 7*, and *Marble Madness* are all 3D and require the player to navigate the environment in some way, and the block rotation tasks required of *Tetris* players closely resemble tasks on standardized tests of mental rotation. Likewise, games like *Solitaire*, *Conjecture*, *Sudoku*, word searches, and crossword puzzles that did not show spatial skill training effects do not seem to incorporate much of any spatial elements, so this at least makes sense. But *Ballance*, a game where the player must navigate a rolling ball carefully along narrow paths and avoid rolling off ledges, seems to have a lot in common with the games that were effective at training spatial skills: like them, it involves navigation through a 3D environment, and like *Marble Madness* in particular, players navigate the environment by controlling a rolling a ball. It is not clear why this game is ineffective at training spatial skills but *Marble Madness* is.

Furthermore, all the games shown to be effective at training spatial skills vary widely across different dimensions such as game genre (shooter, puzzle, platformer, racing), player perspective (first person versus third person), and speed of play (fast-paced versus slow), so it is not quite clear exactly what features of these games may be relevant to training spatial skills. It is particularly embarrassing that *Lumosity*, a game designed specifically to train spatial skills (among other cognitive skills), should have no training effects, but a bunch of commercial games not even designed for the purpose of training should be effective training tools. Clearly, we do not yet understand what features determine a game's effectiveness at training spatial skills, and it is very difficult to determine this with commercial games since they make it very difficult to manipulate and test specific game features.

2.5.3 Noncommercial Games

To investigate the potential of specific game features for tapping into players' spatial skills, researchers have turned to non-commercial, custom-built games, although research in this area is still in its infancy. Only two studies thus far outside of this dissertation have investigated training effects of noncommercial games on spatial skill. Mazalek et al.

found that using embodied, gesture-based controls for a teapot-touching game resulted in improvements in a mental rotation task after just 13 minutes of training relative to two other non-embodied experimental conditions playing the same game with either an Xbox controller or a keyboard [152]. Similarly, Chang et al. found that those who played a virtual reality perspective-taking game with a tangible and embodied control scheme improved their performance on a test of perspective-taking relative to those who a low embodiment version of the game or completed a questionnaire of algebra and grammar problems instead after just 15-20 minutes of training [153].

However, these studies use low sample sizes and conduct post-tests of spatial skills immediately after the training intervention. Low sample size could adversely affect the reliability of their results, and the very short term training effects they found may not generalize to studies of longer term training effects. In addition, Chang et al. used a VR game and Mazalek et al. used a custom-built puppet control interface, approaches which are currently not easily scalable and deployable across a large population of students and school districts.

2.5.4 My Contribution

In this dissertation, I contribute to this sparse literature on non-commercial spatial skill training games with my novel methods for the development and evaluation of my modular spatial skill training game *Homeworld Bound* and its successor, *Homeworld Bound: Redux* (discussed in Chapters 3 and 6, respectively). The newest version of the game, *Homeworld Bound: Redux*, contains spatial game features inspired not only by theory but also by those present in commercial games that have been successful at training spatial skills, such as *Portal 2* and *Tetris*. However, unlike many of the commercial action games successful at training spatial skills like *Unreal Tournament 2004* and *Medal of Honor*, *Homeworld Bound: Redux* does not contain any violence and thus presents no problem of inappropriate content for implementation in school classrooms.

Homeworld Bound: Redux offers an estimated 2+ hours' worth of gameplay, the longest gameplay experience yet for a noncommercial spatial skill training game. More gameplay means longer training periods are possible, which are more likely to produce training effects. In addition, it is the first spatial skill training game to be open source and easily accessible to other researchers and educators for use in research studies and training interventions. *Homeworld Bound: Redux* is also the first noncommercial spatial skill training game scalable to large classrooms and remote play thanks to its implementation as an online game playable on the average personal computer or laptop to which a student or public school might have access.

My main contribution to the state of the art in the research literature on spatial skill training studies in this dissertation is the use of *Homeworld Bound: Redux* in a large scale classroom training study in Chapter 7 - the first study of spatial skill training effects in a noncommercial game with more than 20 minutes of training time (70 minutes for my study) and more than 16 participants per training condition (65 or more per condition for my study). This study is also the first game-based spatial skill training study conducted as part of students' regular coursework in an introductory STEM course, which lends it more ecological validity in the context of classroom spatial skill interventions. Furthermore, the training study described in Chapter 7 is the first game-based spatial skill training study to focus on training effects for a low spatial skill population specifically. Focusing on the low spatial skill population allows me to focus my evaluation of training effects on the subset of students for whom training interventions are most critical for improving the chance of success in STEM coursework and majors.

As evidenced by my review of the spatial skill training game literature above, game-based spatial skill training studies with children are overall quite rare and none have been done with noncommercial games. However, children are an essential target population for spatial skill training games given that training in childhood gives students more time to develop the spatial skills necessary for success in STEM electives and introductory STEM coursework in college and that the gender gap in spatial skills favoring men begins to emerge in childhood [21, 23, 8]. My controlled study of the training effects of *Homeworld Bound* discussed in Chapter 3 makes an important contribution to the research literature on game-based spatial skill training with children. While only a pilot study with low sample size, this was the first spatial skill training study of a noncommercial game conducted with children, with the longest training period to date of any spatial skill training study of a noncommercial game (a total of 3 hours distributed over 2 weeks), and showed that children's spatial skills could be trained with a game designed around my theoretical mapping between spatial subskills and game features.

Finally, the modular level design of *Homeworld Bound* and *Homeworld Bound: Redux*, coupled with its automatic player behavior logging, allow me to do what no game-based spatial skill training study has been able to do as of yet - identify specific game features that may be tapping more or less into players' spatial skills on a level-by-level basis to understand not just the what, but the why behind a game's training effectiveness (or lack thereof). In turn, identifying levels with specific combinations of game features that are ineffective at tapping into players' spatial skills allows me to identify ways of improving the game's effectiveness at tapping into players' spatial skills on a level-by-level basis and therefore its potential as a training intervention. In the next section, I discuss existing work analyzing the relationship between player performance, spatial skill, and game features and how the studies I conducted with *Homeworld Bound* and *Homeworld Bound: Redux* in this dissertation build upon and extend this work.

2.6 GAME FEATURES RELATED TO SPATIAL SKILLS

To understand why some games are effective at training spatial skills and others are not, it is essential to understand what specific game features may be contributing to a game's effectiveness. As mentioned in Section 2.4, numerous studies have found that those who play action video games frequently tend to have higher spatial skills than those who do not. In these studies, researchers recruited participants that were either action video game players (AVGPs) or non-action video game players (NVGPs) based on their self-reported frequency of playing action or first person shooter games for a series of experiments. In these experiments, AVGPs consistently showed higher performance on various measures of spatial attention and visual search [112, 113, 114, 115, 116, 27, 117].

However, very few researchers have examined features in specific action games that may be tapping into players' spatial skill. Only two studies I am aware of looked at the relationship between player behavior and spatial skill in a specific action game. One study by Greenfield et al. found that expertise in the 3D arcade game *The Empire Strikes Back* was related to players' mental paper folding skill (an *intrinsic-dynamic* spatial skill) [124]. Another study by Adams and Mayer found that overall performance in the action game *Unreal Tournament 2004* in terms of total number of enemy kills was more related to players' spatial skill than their performance (high score) in the non-action game Tetris [154].

Only two studies outside of this dissertation conducted a more fine-grained analysis of specific game features related to spatial skills. In a study where participants completed a simple object location and navigation game utilizing the Virtual Morris Water Maze, de Castell et al. found that higher spatial skill was associated with lower search time and longer dwell time in the correct location. However, when proximal navigation cues were introduced to the environment, the association between performance on the task and spatial skill vanished [155]. Another study was conducted by Xiao et al., who developed *Cubicle*, a gamified version of paper-based spatial skill training exercises adapted from Sorby et al.'s spatial skill training workbook for engineering students [128, 129]. Each level of the game was designed to train a different combination of the following spatial skills relevant to engineering: 3D object visualization and manipulation, perspective taking, mental rotation, 2D to 3D transformation, and spatial memory. In a pilot study, Xiao et al. found that performance in each level correlated highly with players' mental rotation test scores [156].

This dissertation aims to fill this large gap in the research literature by presenting the first two studies conducting a fine-grained analysis of the relationship between player performance and spatial skill across multiple sets of features and across levels with two very different types of gameplay: first person exploration and 3D object construction (Chapters 4 and 7). These studies were made possible by the intentional modular design of levels in both *Homeworld Bound* and *Homeworld Bound: Redux* as well as the detailed automatic player behavior logging capabilities built in to both versions of the game, features that to my knowledge are not present in any other game studied in the spatial skill training literature. The modular level design of *Homeworld Bound* and *Homeworld Bound: Redux*, combined with this detailed information about player behavior and performance, enables researchers to study the effect of different combinations of theoretically grounded spatial game features on the relationship between player performance and spatial skill on a level-by-level basis, identify specific levels that are not tapping into players' spatial skills sufficiently, and apply fixes to these problematic levels before investing in a time-consuming and potentially expensive controlled training study with the game. In addition, my study in Chapter 4 is the first study of the relationship between player performance and spatial skill conducted with children, an especially important target population for spatial skill training interventions given that the gender gap in spatial skills favoring boys begins to emerge in childhood [21, 78, 23] and low spatial skills present a barrier to success in STEM electives in high school and beyond [1, 2].

Chapter 3: Homeworld Bound, A Game for Training Children's Spatial Skills

In this chapter, I describe the design process for and preliminary evaluation of *Homeworld Bound*, a game I designed from the ground up not only to train players' spatial skills, but also as a testbed for evaluating the effectiveness of different combinations of game features at tapping into players' existing spatial skills. The ultimate goal of developing *Homeworld Bound* is to present a non-commercial game-based training intervention that is free and easy for educators and students to use for spatial skill training interventions, and at the same time easily modifiable and extendable by researchers and game designers for the continual study of how different game features contribute to the training effectiveness of game-based spatial skill training interventions.

3.1 OVERVIEW OF THE GAME AND DESIGN PROCESS

Homeworld Bound has a relatively simple premise: the player has crash-landed on a seemingly uninhabited alien planet, their spaceship destroyed. The player must navigate through several different alien landscapes to find the scrap parts they need to rebuild their spaceship and return home. Along the way, the player collects parts that can be used to build useful items that allow them to access new areas of the environment. Gameplay is divided between Exploration Mode levels and Construction Mode levels. In Exploration Mode levels, the player navigates canyons, fields, and city ruins, collecting parts until they have enough to build something, at which point they can enter Construction Mode to actually build an item out of the parts they have collected. In Construction Mode, the player builds items such as the Rocket Boots (allows the player to jump higher) and Sledgehammer (allows the player to destroy debris blocking their path) by correctly matching up the faces of two parts that can be fused together, rotating one of the parts until the areas to be fused are aligned correctly, and then fusing the parts together. This process is repeated for each part, until the construction is finished and the player is returned to Exploration Mode to use their newly built item to access new areas of the game world.

The design process for *Homeworld Bound* was driven by an in-depth analysis combining theory of spatial skill typologies and evidence from empirical studies about what kinds of games have been effective at training spatial skills in the past. Sections 3.2, 3.3, 3.4, and 3.5 describe the theoretical basis for the selection of game features and mechanics in the final implementation of *Homeworld Bound*. Section 3.6 explains the iterative design and playtest process that led up to the game's final version. Section 3.7 discusses the implications of this work in the larger context of spatial skill training research. This chapter's main contributions to the research literature are 1) a theoretically grounded approach to developing games and game features that tap into players' spatial skills and 2) the implementation of these features in a functional, engaging children's computer game that has demonstrated potential to train spatial skills.

3.2 TARGET AUDIENCE AND TECHNICAL SPECIFICATIONS

I designed *Homeworld Bound* for a target audience of late elementary school students (ages 8-11) since this is the age when gender differences in spatial skill first clearly emerge [21, 76]. Since addressing the well-established gender gap in spatial skills [21] as early as possible is essential for making STEM fields more accessible to girls and women (and to anyone with low spatial skills more broadly), an intervention at this stage would prevent girls' spatial skills from falling too far behind boys' and becoming a barrier to participation in STEM courses. More broadly, successful spatial skill training interventions at this young age give all low spatial skill students a chance to develop their spatial skills well before they encounter elective STEM courses in middle school, high school, and college, preventing low spatial skill from being a barrier to success in STEM coursework, majors, and careers.

I decided to build *Homeworld Bound* using the Unity game engine for three reasons. First, Unity is free for noncommercial use, so it would be more accessible to researchers and educators wanting to modify the game in some way in the future for their purposes. Second, Unity is designed for 3D games, making it easier for me to incorporate specific spatial features from 3D video games that have been shown to be successful at training spatial skills like *Portal 2*, *Medal of Honor*, and *Unreal Tournament*. Third, the Unity engine is platform-independent and makes it easy to port games across different platforms, meaning that I could target different platforms, such as PCs, Macs, mobile devices, various game consoles, or online (WebGL) depending on the needs of the game's target audience.

Homeworld Bound was initially developed for standalone PCs and Macs due to their common use in schools; school districts on tight budgets, whether public or private, generally possess at least a few computers for students to use. Textures for the game were created primarily in GIMP, and 3D models were built mainly with Blender and 3DS Max. To reduce development time, keep costs low, and ensure that the game would run even on older desktop computers with limited computational resources, I used low fidelity stylized textures and low polygon 3D models.

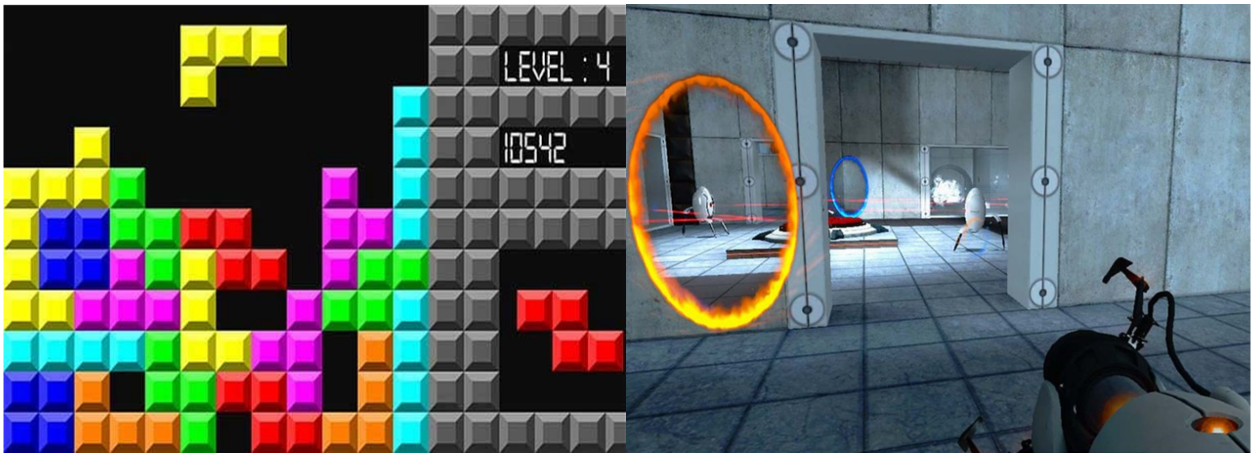


Figure 3.1: Screenshots of *Tetris* and *Portal 2*. In *Tetris*, players must rotate falling blocks to fit them together at the bottom of the screen. In *Portal 2*, the player must navigate through the environment and complete puzzles by placing and traveling through pairs of portals. In this example, the player has placed an orange portal and a blue portal. They will enter the orange portal in front of them and come out of the blue portal on the far wall in order to avoid crossing the red laser and activating the turret gun in the center of the room. Once placed, a portal shows the player a view of what they will see once they step through the portal, allowing the player to determine in advance if they have placed it in the correct location via *landmark orientation.* Screenshots are shown under fair use doctrine.

3.3 GAME FEATURES TO SUPPORT SPATIAL SKILL DEVELOPMENT

In order to design a game that incorporated evidence-based content, I looked to two popular commercial games shown to be effective in the spatial training literature as inspiration: *Tetris* and *Portal 2* (Figure 3.1). Although both *Tetris* and *Portal 2* have been shown to be effective at training certain spatial subskills [151, 4, 5], each game differs in the degree of transfer its training affords. *Tetris* has been shown to improve *mental rotation* skills, an *intrinsic-dynamic* spatial subskill that measures the ability to imagine how the appearance of an object would change as it is rotated in various ways [157, 10]. However, this performance improvement is generally limited to mental rotation tasks where the objects to be rotated look very similar to the blocks the player must manipulate in *Tetris* [126, 158]. *Portal 2*, in contrast, is a physics-based first person puzzle game requiring the player to decide how to walk, climb, and jump their way through the environment using a special gun that allows them to create portals on floors, walls, and ceilings and travel through them. Despite the substantially different gameplay, training with *Portal 2* has also been shown to improve

Game Feature	*Spatial Operation*
Object Rotation	Intrinsic-Dynamic
Object Alignment	Intrinsic-Dynamic
Landmark Orientation	Extrinsic-Static
Navigation Visualization	Extrinsic-Dynamic

Table 3.1: Theoretical framework for mappings between game features and spatial operations used to inform the design of *Homeworld Bound.* Note: The mapping shown here reflects the most current version; an older and less theoretically grounded version of this mapping was used initially at the time of publication of this work [40] and then revised later to be more in line with the research literature on spatial subskills [42].

players' mental rotation ability and performance on an additional spatial subskill in the *extrinsic-dynamic* category called as *environmental spatial skill* [159, 160, 5], demonstrating the far transfer potential of training with this game.

While *Tetris* does not seem to afford the same far transfer benefits as *Portal 2*, it is a much simpler game, and therefore it is easier to determine which of its features might play a role in improving players' spatial skills. In *Tetris*, players have two tasks: rotate falling blocks quickly, and move those blocks into the most dense configuration possible. I will refer to these two features as *object rotation* and *object alignment.* It is more difficult to determine which features of *Portal 2* gameplay might be contributing to its effectiveness at training spatial skills. However, the game's emphasis on its central mechanic of *portals* (see Figures 3.1 and 3.2 for explanations) suggests that players may be using landmark cues to decide where to place a portal and then visualizing how their character would move from one end of the portal through the other in order to evaluate whether their portal was placed in the correct location. I will refer to these two features of the gameplay as *landmark orientation* and *navigation visualization.*

Each of these four game features, *object rotation*, *object alignment*, *landmark orientation*, and *navigation visualization*, correspond to specific spatial subskills. The *object rotation* and *object alignment* operations required in *Tetris* correspond to the *intrinsic-dynamic* quadrant in Uttal's 2x2 taxonomy of spatial skills, while the *landmark orientation* and *navigation visualization* required of *Portal 2* players correspond to the *extrinsic-static* and *extrinsic-dynamic* quadrants in the taxonomy, respectively (Table 3.1) [10]. I focused on incorporating these features and their corresponding spatial operations into my design of *Homeworld Bound* to make sure players' spatial skills would be taxed by playing the game.

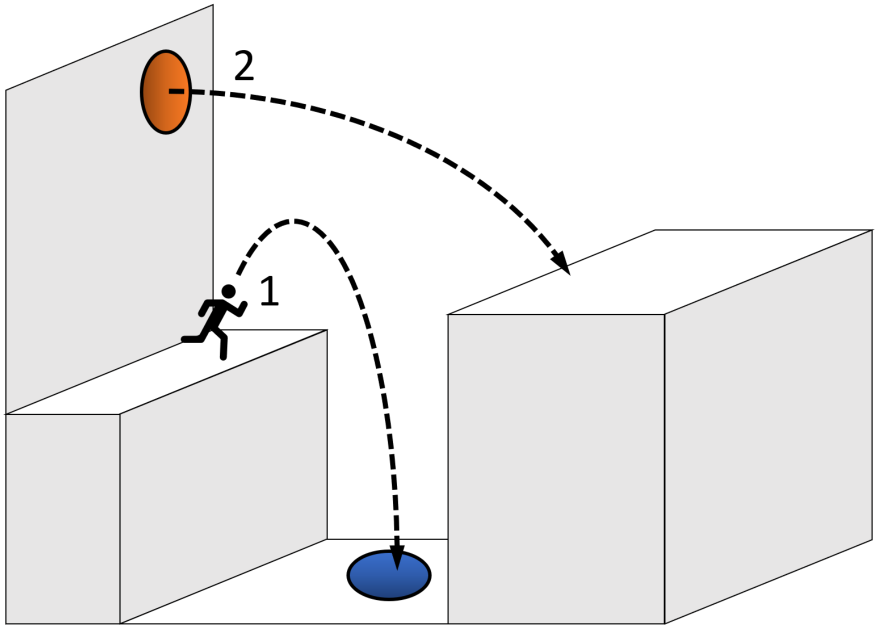


Figure 3.2: Diagram of a commonly used method for traversing a room with portals in *Portal 2*. Since the player's momentum but not direction is conserved while traveling through portals, falling through a portal (1) can generate the momentum necessary to fling the player across a horizontal gap once they emerge from the second portal opening (2). In order to place portals correctly to achieve this outcome, players must employ *navigation visualization* to mentally visualize how their character would move in space as a result of entering and exiting each portal.

The well-documented relationship between play with LEGO block construction sets and spatial skill [161, 162, 163] provided the motivation for including LEGO-like construction tasks in *Homeworld Bound* that incorporated *object rotation* and *object alignment* into gameplay. My idea was to have players construction a 3D object, one part at a time. Players would be given a part to start with and would then determine how each subsequent part should be placed and rotated to fit together with the rest and achieve the desired structure. This task maps well to the kind of object rotation and alignment required in *Tetris*, except in three dimensions rather than two. Since the other two spatial game features I proposed, *landmark orientation* and *navigation visualization*, require the player to orient themselves in the environment relative to specific landmarks and mentally plan a route through the environment, respectively, tasks involving navigation were a natural fit for the game as well. Therefore, I decided to design *Homeworld Bound* with construction and navigation as the two central game elements.

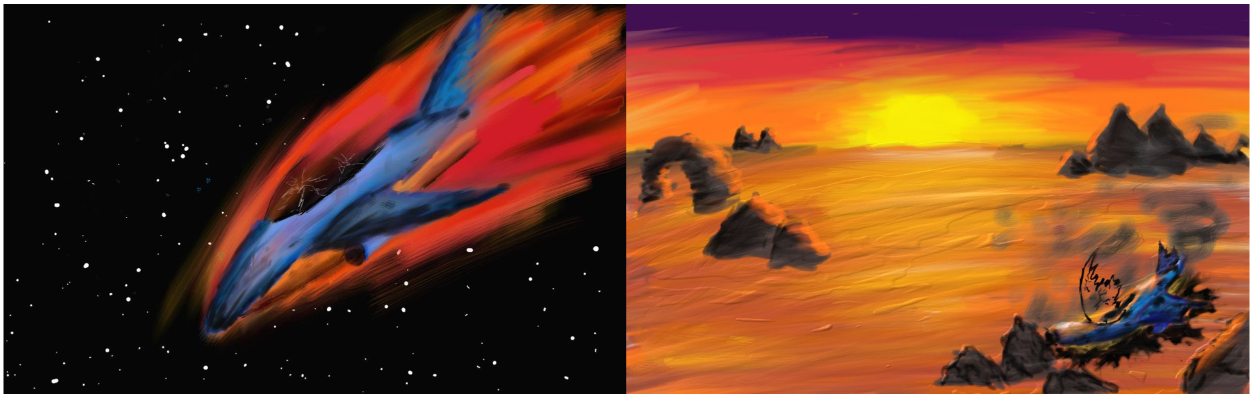


Figure 3.3: Concept art for *Homeworld Bound* showing the general premise of crash-landing a spaceship on an alien world and exploring it.

3.4 GAME FEATURES TO SUPPORT ENGAGEMENT

Designing with engagement in mind, I had to decide how to integrate both construction and navigation game elements into one seamless game experience that would be appealing for the maximum number of children. The commercial game *Minecraft* caught my attention because it incorporates both construction and navigation into a highly entertaining gaming experience that has made it extremely popular with children [164]. In *Minecraft*, players search the environment for and collect materials that they can use to build useful items, which in turn help them to collect more kinds of materials, survive, and explore new areas of the game world.

The appeal of *Minecraft*'s gameplay can be explained by the psychological need satisfaction of *competence* and *autonomy* it offers according to Deci and Ryan's *self-determination theory* (SDT) [61], as well as its incorporation of *curiosity* and *control* elements from Malone and Lepper's taxonomy of intrinsic motivations for learning [24]. *Minecraft* gameplay is very self-directed, allowing players the freedom to explore wherever they want to satisfy their *curiosity*, collecting items or building things as they choose. This self-direction likely promotes feelings of *autonomy* and *control* in players. *Minecraft* gameplay can also build players' sense of *competence* as they collect more items and are gradually able to build more impressive, useful items. These feelings of autonomy and competence, in turn, increase players' intrinsic motivation to play [62]. Inspired by *Minecraft*'s method of integrating construction and navigation into a continuous feedback loop and promoting autonomy and competence in players, I designed similar mechanics for *Homeworld Bound*.

To further increase engagement and motivation to play, I added a simple narrative to the game. This narrative served the dual goals of tying together the construction and exploration elements of *Homeworld Bound* into one coherent game experience and incorporating an element of *fantasy* from Malone and Lepper's taxonomy of intrinsic motivations for learning [24]. The premise of the game, and the reason it is titled *Homeworld Bound*, is that the player has crash-landed their spaceship on an alien world and must rebuild the spaceship to escape the planet and return home (see Figure 3.3). In order to do so, the player must scavenge materials from the environment, which they can use either to rebuild parts of their spaceship or to build equipment that will allow them to reach more areas of the game world and collect more parts.

3.5 GAME FEATURES TO SUPPORT MODULARITY

I designed *Homeworld Bound* to have a modular structure, which allowed me to analyze the relationship between players' spatial skills and each specific game feature individually. The modular structure of *Homeworld Bound* is hierarchical. First, the game is divided into *Construction Mode* and *Exploration Mode*, which the player switches between repeatedly as they advance through the game. In Exploration Mode, the player searches the environment for parts and batteries, both of which are used to build useful items in Construction Mode. In Construction Mode, the player builds 3D items using the parts and batteries they have collected by rotating and aligning parts with each other, one by one. Within each mode, there are multiple levels, each incorporating a different environment designed to incorporate a different mix of game features I hypothesized to embody my proposed four spatial game features (*object rotation*, *object alignment*, *landmark orientation*, and *navigation visualization*). I describe level-by-level differences in spatial features in more detail in the following subsections. Figure 3.4 shows examples of some of the levels included in each mode.

3.5.1 Spatial Game Features in Exploration Mode Levels

Canyon

In the first Exploration Mode level, Canyon, player movement is restricted to a narrow, winding canyon where the player cannot see very far ahead at any given time. Navigation is relatively simple as the player is restricted to going from one end of the canyon to the other or to various short offshoots that are all dead ends. The level is designed to require the use of extensive *landmark orientation* as nearby landmarks become more familiar after seeing them multiple times during level exploration. It is not designed to get players to use navigation visualization. I designed this first level to be the simplest both for the purpose

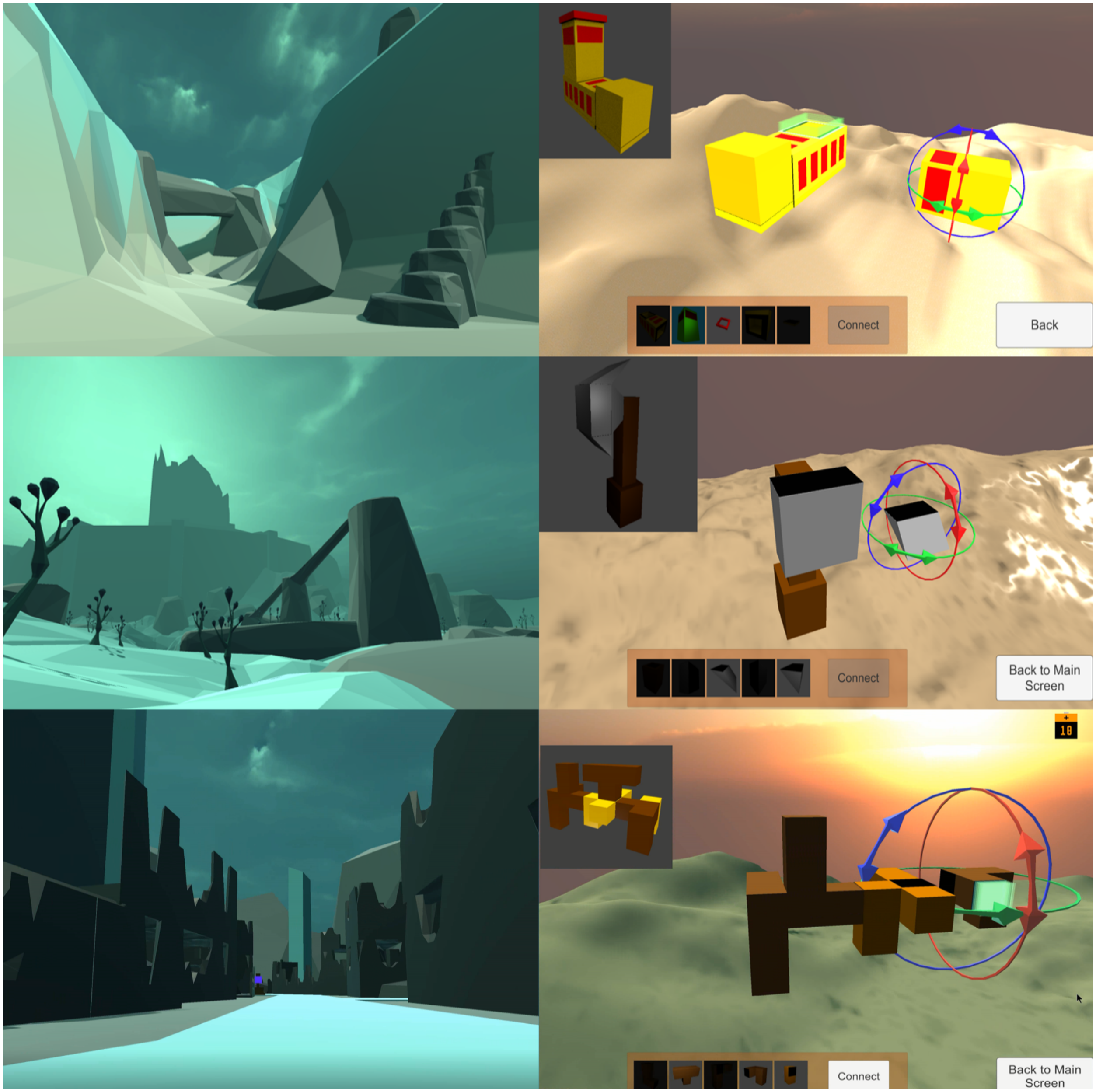


Figure 3.4: Game mechanics and features in *Homeworld Bound*, my spatial skill training game and testbed for different spatial game features. In Exploration Mode, players navigate through through the Canyon (top left), Highlands (middle left), and Ruined City (bottom left) collecting parts. This requires them to orient themselves in the environment relative to surrounding landmarks (*landmark orientation*) and determining a route and/or jumping location to get to hard-to-reach parts (*navigation visualization*). In Construction Mode, players build equipment (Rocket Boots, top right, Sledgehammer, middle right, and Ruined City Key, bottom right) using the parts they collected. This requires players to determine the correct rotation for a part in order for it to fit together correctly with the others (*object alignment*) and then execute the series of rotations necessary to obtain that correct alignment (*object rotation*).

of helping the player learn the game and to determine if the simple task of exploring an environment in first person without any sort of structure, as is often done in first person shooter games like *Medal of Honor* and *Unreal Tournament*, might be enough to tap into players' spatial skills.

Highlands

In contrast, in the Highlands level, movement is unrestricted across a wide open field, with a larger area for exploration. In addition, there are a series of large rocks spaced around the environment that must be climbed using a series of successive jumps, so vertical as well as horizontal navigation is required, utilizing the Rocket Boots item built in the previous level. Given that the parts and batteries at the top of large rocks are only visible from a significant distance away, players must plan out their multi-jump path to the top of the rock to get the item before they get too close to the rocks, requiring them to employ *navigation visualization*. *Landmark orientation* can also be used in this level by observing the different shapes of rocks, trees, and other structures, although the most visually prominent landmarks in this level serve as distal navigation cues: the outside of a walled city that dominates the horizon, and the tallest rock in the middle of the level, which can be reached with a series of ramps from other rocks not seen anywhere else in the level. Using landmarks that are visible from much further away than in the Canyon level allowed me to study how landmark distance (proximal or distal) may affect the extent to which landmark orientation taps into spatial skill given the findings of Castell et al. that proximal landmarks may reduce the need to employ mental rotation, an *intrinsic-dynamic* spatial skill hypothesized to be relevant to navigation skill [155].

Ruined City

In the final implemented Exploration Mode level, Ruined City, searching for item parts is accomplished in a different way. The six parts for the Ruined City Key, this level's item to build, are hidden in a series of partially ruined buildings. This makes it more difficult to find the parts since the city is full of very similar looking building ruins with multiple stories, and the buildings' walls obscure parts hidden within from view (see Figure 3.4). The purpose of making the part collection task more difficult in this level was to motivate the player to use a search strategy that more explicitly requires the use of *landmark orientation* and its associated *extrinsic-static* spatial skills. This strategy is supported by the presence of pictorial clues the player can easily find and collect on the main paths through the level.

Each clue shows a screenshot of the location of one Ruined City Key part from a camera angle that simultaneously prevents the part from being obscured by building walls and also includes enough background details for the player to be able to deduce with some effort in which building and room a part is likely to be. Background details that help orient the player include a colored lamp, the lamp's position relative to the main path through the level, and the shapes of the ruined building walls within which the part is to be found (see Figure 3.5). The colored lamps are spaced evenly throughout the level and designed to help the player find the rough area in which a part is located, while the rest of the background details help the player figure out in which specific building, floor, and room the part is located.

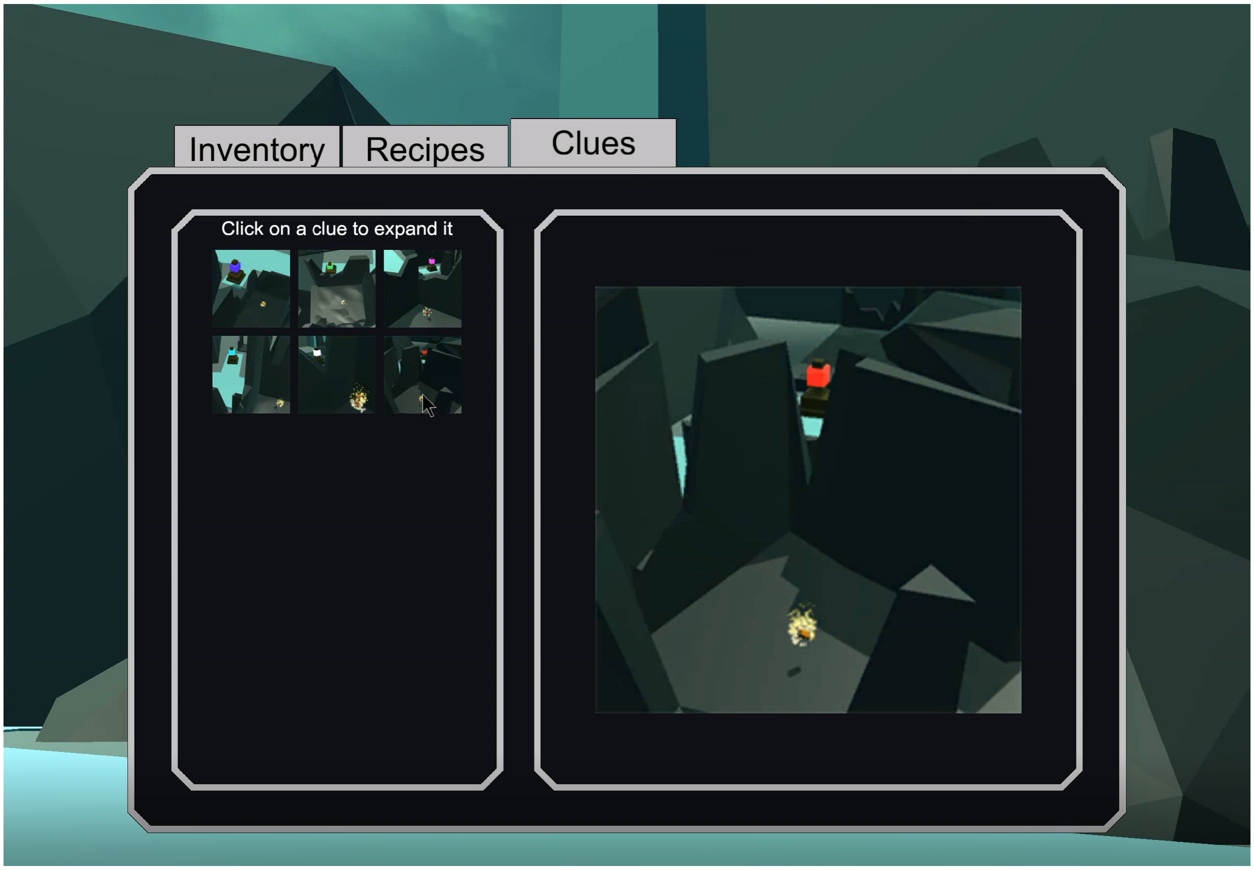


Figure 3.5: Screenshot of the clue interface in the Ruined City level of *Homeworld Bound*. As the player collects pictoral clues, they populate the picture grid on the left side of the screen. The player can click on on one of them to expand and display it on the right side of the screen. Players must use *landmark orientation* to figure out where they would have to be standing in the level to get the view of the item part shown in each clue.

Like the Highlands level, the Ruined City level is more vertical and requires the player to make multi-step jumps from certain locations to get to the higher floors on buildings. Thus, it may require the player to do some *navigation visualization* to plan how they will get to where they are trying to go. The more dense and cluttered structure of the level may make this task more difficult for the player and require more *navigation visualization* than in the Highlands level. As mentioned above, the use of pictorial clues also requires *landmark orientation* more explicitly than the unstructured exploration of the Highlands and Canyon levels and thus may do a better job of embodying this spatial game feature than the other two Exploration Mode levels. These pictorial clues were inspired in part by the game mechanics in *Portal 2*.

3.5.2 Spatial Game Features in Construction Mode Levels

Construction Mode Overview

The player's primary task in Construction Mode levels is to build an item from its constituent parts piece by piece in 3D. The step by step process is more complex than Exploration Mode levels (see Figure 3.4 for reference to specific UI features). At the start of the level, the player sees a 2D image in the upper left corner of the screen showing what the item should look like when it is finished. This image is static and thus cannot be rotated to see the item from different perspectives, so in order to see the correspondence between the placement of parts in the 2D image and the 3D parts they are manipulating, players must mentally rotate either the 3D parts in the environment or the 2D image, engaging their *intrinsic-dynamic* spatial skills. The first part of the item is already placed in the environment in the center of the screen and cannot be manipulated or rotated in any way since it serves as the construction's base.

The player's task is the following. First, they choose a part to place in the environment using the icons on the bottom of the screen. Next, they decide to which area on the base construction they will attempt to fuse their chosen part; each fusable area on each part is black. Once the player has selected one fusable area on the base construction and one on the selected part, they must determine how to rotate the selected part (*object rotation*) so that if it were pushed against the base construction at the selected black faces, it would fit together with the construction perfectly (*object alignment*). To make it easier for the player to understand how the selected part must be rotated to align correctly with the construction, the selected part moves so that it is next to the selected fusable area on the construction. Rotation along the x, y, and z axes is accomplished by clicking on the red, green, and blue rotation arrows which surround the selected part, and each rotation rotates the part 90 degrees in the direction indicated by the arrows.

Once the player decides that they have matched up the correct two black faces where the two parts should attach together and rotated the selected part to align it correctly, they can click the Fuse button to attempt a fuse. If the face matching and the part's current rotation are correct, the part is fused. If not, an error message pops up telling the player which of the two errors they made (wrong face or wrong rotation). A wrong face error indicates that the attachment attempt failed due to the two selected faces being incompatible; a wrong rotation error indicates that the attachment attempt failed because the part was not rotated to fit together correctly with the construction. The order of rotations and face selection is flexible; players can rotate first and then select, or select and then rotate, and selections of faces and parts can be changed at any time. The camera can be rotated around the construction and zoomed using the mouse. Once fused, a part becomes part of the construction and can have other parts fused onto it.

This process is repeated for each part attached to the construction until all parts are successfully fused or the player runs out of rotations (the number of remaining rotations is shown in the battery icon in the upper right corner of the screen). The player is free to choose a different part to attach at any time if the current one is proving difficult to attach. In addition, players can make as many different combinations of face selections as they want before attempting attachment. While players can also experiment with rotating a part in different ways, there is a soft limit on the number of rotations that can be performed in each Construction Mode level.

I decided to limit the number of rotations allowed in Construction Mode levels to encourage players to use *mental rotation* (an *intrinsic-dynamic* spatial skill corresponding to my *object rotation* and *object alignment* spatial game features) instead of trying to fuse every possible rotation of a part in succession until one is successful. Each rotation in Construction Mode uses up a small amount of the player's "battery power," which is determined by the number of batteries the player has collected in Exploration Mode. When their battery power runs out, players can no longer rotate objects and must switch back to Exploration Mode to collect more batteries before they can return to where they left off in Construction Mode. Batteries are automatically respawned elsewhere in a level when they are collected, ensuring that the player does not run out of them. This mechanic allows the player take a break from one in-game task and do something else for a while (the game design principle of *parallelism*), which has the added advantage of reducing player frustration and helping players come back to the main task cognitively refreshed [165]. The player is automatically returned to Exploration Mode to search for new parts once they finish a Construction Mode level.

Tutorial 1 and Tutorial 2

Once the player has collected all of the parts needed for the Rocket Boots item, they first must play through two easier tutorial levels, Tutorial 1 and Tutorial 2 (see Figure 3.6), in order to master the basics of Construction Mode gameplay they will need to complete the Rocket Boots level. I decided to add in these tutorial levels after repeat playtesting with children revealed that the whole process of rotating, aligning, selecting faces, and fusing took a while for players to understand. Tutorial 1 focuses on getting players to match up black fusing areas with the same shape by making all parts the same color and making each part and fusing area very different shapes. Tutorial 2, in contrast, focuses on getting the player to pay attention to the image of the finished item in the upper left corner by making fusing areas very similar shapes or identical; the player can no longer rely on matching by fusing area shape alone but must look at the picture to see where each part goes. The picture makes this very clear since all parts are colored differently. Thus, Tutorial 1 emphasizes shape matching, whereas Tutorial 2 emphasizes color and the relationship between the 3D construction and the 2D image of the finished item. Later Construction Mode levels combine both these elements in a variety of different ways.

Rocket Boots

The Rocket Boots level combines the features of Tutorial 1 (similarly colored parts in either red or yellow) and Tutorial 2 (similarly shaped rectangular parts), making this level harder than the preceding two. In order to succeed, the player must pay attention to the 2D image of the finished item and the subtle differences in the shapes of fusing areas simultaneously. Two parts have multiple axes of symmetry, allowing for multiple correct alignments when rotating and making the fusing task for these parts a little easier.

Sledgehammer

The Sledgehammer level preserves the high degree of symmetry of parts, the similarity of fusing areas, and the two color scheme of the Rocket Boots level, but introduces more complex shapes: triangular prisms and a trapezoidal prism. All fusing areas, however, remain rectangular or square. The more complex shapes allow the player to gain experience with mentally rotating and aligning more general types of 3D shapes.

Figure 3.6: Screenshots of the first (left) and the second (right) tutorial levels in the final version of *Homeworld Bound*. The first tutorial emphasizes shape matching, while the second emphasizes color and matching with the picture of the finished object in the upper left corner of the screen.

Ruined City Key

With every fusing area an identical square, and almost all parts the same color, the Ruined City Key level is designed so that players must rely heavily on the 2D static image of the finished construction to determine where to attach parts and how they should be rotated. The parts themselves are shaped like 3D *Tetris* blocks, meaning their shape is considerably more complex than parts in the previous levels. While previous levels could be completed without reference to the 2D image with a bit of guessing, this level was designed to make this guessing strategy much less fun and persuade players to analyze the correspondences between the 2D image and their 3D construction. The two color scheme is continued from previous levels.

3.6 PILOT TESTING

Homeworld Bound is described in the previous sections of this chapter in the state it was just before I used it in my first published study, which is discussed in the next chapter. However, multiple iterations of playtesting and development were necessary to get the game to a state where gameplay aligned with my four proposed spatial game features and represented a coherent, engaging whole. The following subsections describe the process of *Homeworld Bound*'s development up to this point.

Figure 3.7: Screenshots of the tutorial level (left) and Rocket Boots level (right) in the earliest functional prototype of *Homeworld Bound*.

3.6.1 Initial Prototype and Playtest

The first implementation of my concept for *Homeworld Bound* was a low fidelity but functional prototype of the Rocket Boots level (see Figure 3.7). Play testing was conducted by me and a collaborating graduate student in the Education department. We conducted the playtest at the Children's Museum of Indianapolis, inviting children in third through eighth grade (ages 8-14) to try out the two implemented levels in the functional prototype, the tutorial and the Rocket Boots level, during family visits to the museum. We observed children during gameplay and asked them to think aloud about their experience playing the game. We recorded the comments children made during gameplay as well as the amount of time they chose to play, how engaged they appeared, and what emotions they seemed to be expressing. Lastly, we conducted interviews with each child to identify which features of the game were perceived as the most positive or negative and how we might improve the game in its next iteration.

In total, five children played *Homeworld Bound*, with four completing both levels and the fifth choosing to end the game early. We observed high levels of engagement in all children; their eyes and attention were almost always on the game, and they leaned forward in their seats as they played. Children's facial expressions during gameplay conveyed primarily focus and confusion with occasional bouts of mild frustration. Four out of five players tried to figure out the game on their own, with only one turning to us for help. All five children reported finding the game fun and were unable to identify an aspect of the game they found boring, but found the task of trying to rotate and connect the parts correctly frustrating.

The playtest of my initial functional prototype revealed two ideas for possible revisions. First, children had a hard time understanding in what direction a part was rotating when they performed a rotation operation. This was likely due to a lack of rotation animation for each part; parts simply rotated instantaneously when one of the rotation buttons on the interface was clicked. I decided to add a more realistic animation to the rotation to show the part smoothly transitioning from one rotation to another to address this issue. I also worked on improving the background visuals for the next iteration as one player mentioned this as an area of improvement. Issues that remained unresolved into the next iteration included the lack of narrative (mentioned by one player) and most players' confusion about the step by step process of finding where to attach a part, rotating it to align correctly, and attempting to fuse.

3.6.2 Second Prototype and Playtest

In addition to the changes made to Construction Mode mentioned above as a result of my first playtest of *Homeworld Bound*, I implemented the first functional prototype of Exploration Mode with the help of a couple undergraduate students before the next playtest in Fall 2015. The goal of this playtest was to prepare the game for a controlled training study the next month. Given this short timeline for implementation and the small number of students working on the game, I decided to implement Exploration Mode in a different game engine to speed the development process. For this first prototype, I and the undergraduate students used the *Warcraft III World Editor* as our development engine. Originally designed for creating custom levels of the real time strategy game *Warcraft III*, the editor makes it easy to create 3D environments with a collection of premade texture and 3D model assets and simple editing tools. It also comes equipped with a simple scripting language that can be used to add more complex logic to the game, and a prebuilt character movement system and interface. Figure 3.8 shows screenshots of Construction Mode and Exploration Mode as they looked for the second iteration of my functional prototype.

Playtesting for this version was conducted with a larger group of 16 children at a local public elementary school, Barkstall Elementary. After installing *Warcraft III* on each of the computers in their computer lab with the help of Barkstall's IT personnel and adding the necessary game files for both Exploration and Construction Mode to the filesystem on each computer, I invited children to play both modes of the game in the lab. Each child completed two 45 minute sessions of gameplay spaced across two days, during which they alternated between Exploration Mode in *Warcraft III* and Construction Mode in Unity through the use of passwords they were given upon completing each level. Children played the game in groups of 8 simultaneously as I walked around and observed them, troubleshooting any points of confusion or bugs they encountered but otherwise just letting them play.

Figure 3.8: Screenshots of the Ruined City level in Exploration Mode implemented with the *Warcraft III World Editor* (left) and the Rocket Boots level implemented in Unity (right) in the second functional prototype of *Homeworld Bound*. Game modes were connected together via passwords given to the player after completing each level.

Overall, children seemed highly engaged with the game, although the process of switching between two separate games and entering passwords in both of them proved cumbersome and confusing for them. The playtest did not uncover any serious issues with player engagement or usability, except for a few glitches, so the next iteration of the game did not substantially change its content or gameplay apart from fixing bugs. However, children continued to find the multi-step process of part attachment in Construction Mode confusing, and some needed hints or additional explanations to understand it at first.

3.6.3 Controlled Training Study with Third Prototype

Method

In Fall 2015, I ran a controlled training study using the third iteration of *Homeworld Bound* resulting from the previous playtest at Barkstall Elementary School. I recruited 21 students in grades 3-5 (ages 8-11) at University Primary School, a private elementary school affiliated with the University of Illinois Urbana-Champaign, to participate. The study was conducted during school hours in 6 sessions over the course of 2 weeks. The purpose of this small-scale pilot training study was to see if playing *Homeworld Bound* could improve students' spatial skills relative to a nonspatial control game.

The active control group played *Little Alchemy*, a simple yet addictive webgame where players drag and drop different "elements" onto each other to combine them and create new "elements." The player starts with four basic elements and can create 360 unique elements in total. I selected this game for the active control group due to its lack of spatial gameplay elements and its potential to keep students engaged throughout the 6 sessions of the study.

All 21 students spent the first study session taking two different standardized tests of spatial skill. The first test was the Revised Purdue Spatial Visualization Test: Visualization of Rotations (PSVT:R) [166]. I chose to use the PSVT:R since it is a well-validated test of mental rotation (an *intrinsic-dynamic* spatial subskill [10] that maps well to the *object rotation* and *object alignment* tasks in Construction Mode). Furthermore, this specific test of mental rotation is widely used in studies analyzing the link between spatial skills and STEM proficiency [167, 168, 169, 170, 171]. Since the Revised PSVT:R was designed for populations age 13 and up, on the advice of the test's author, I reduced the number of questions from 30 to the first 20 (test questions are ordered from easiest to hardest) and increased the time limit from 20 to 25 minutes.

The second test administered was a version of Guay's Visualization of Views test adapted by Hegarty et al. [172, 55], a measure of *perspective-taking* (imagining how an object or scene would look when viewed from a different perspective [49]. I included this perspective-taking test to assess the *extrinsic-static* dimension of spatial ability [10] related to the task of *landmark orientation* in Exploration Mode. Since this test, like the Revised PSVT:R, was also designed for older populations, I increased the time limit from 8 to 15 minutes. When calculating Visualization of Views test scores, I used the adjustment for guessing recommended by the test's authors (subtracting the number of incorrect answers divided by the number of answer choices (6) from the total score). Given the likelihood of cognitive fatigue children would be likely to experience if a third test was added and the lack of well-validated tests of *extrinsic-dynamic* spatial skill, I could not add a third test to directly measure the impact of the game's *navigation visualization* tasks. However, research indicates that spatial skill training often has transfer effects from one spatial subskill to another [10], so measuring each dimension of spatial skill separately may not be necessary.

Students were assigned to play either *Homeworld Bound* or *Little Alchemy* using block random assignment with age and gender as the blocking variables and played their assigned game for the next four 45 minute sessions of the study. The final study session was a post-test using the same PSVT:R and Visualization of Views tests as before, followed by a short demographic survey.

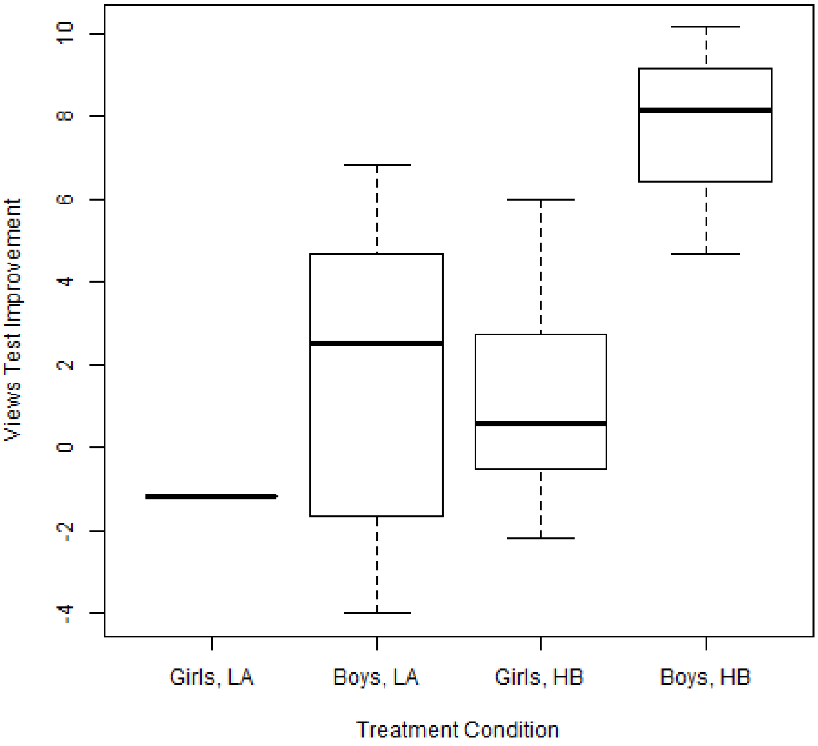


Figure 3.9: Children's improvement from pretest to post-test on the Visualization of Views Test, broken down by treatment condition and gender. LA = *Little Alchemy*, HB = *Homeworld Bound*. Note that there was only one girl in the *Little Alchemy* group due to the removal of an outlier. While these patterns do not translate to statistically significant interaction effects, they warrant further investigation as to why boys' performance is driving the statistically significant test improvement in the *Homeworld Bound* group.

Results

Table 3.2 shows a summary of boys' and girls' scores on each pretest and post-test of spatial skills. On the Views pretest, both girls' and boys' average scores ($\mu = 3.72$ and $\mu = 3.82$, respectively) were above chance performance (0 out of 24 since scoring already adjusts for guessing). On the PSVT:R pretest, however, only the boys' average was well above the level of chance ($\mu = 9.1$), while the girls' average was close to chance performance ($\mu = 5.9$, chance performance would be 5).

I analyzed the effects of training with either *Homeworld Bound* or *Little Alchemy* on spatial skill pretest to post-test improvement. I was also interested in investigating gender differences in pre- to post-test learning gains given that the age range of this study's participants was similar to the age at which the gender gap in spatial skills begins to emerge in

Visualization of Views

	Pretest μ	*Pretest* σ	*Posttest* μ	*Posttest* σ
Girls, LA	4.17	N/A*	3.00	N/A*
Boys, LA	4.92	7.48	6.55	7.13
Girls, HB	2.44	3.88	3.63	3.94
Boys, HB	1.22	3.08	8.89	1.98

Shortened Revised PSVT:R

	Pretest μ	*Pretest* σ	*Posttest* μ	*Posttest* σ
Girls, LA	10.50	4.95	10.00	8.49
Boys, LA	9.43	5.02	9.43	5.80
Girls, HB	4.75	2.92	5.13	2.80
Boys, HB	8.33	4.04	9.33	5.51

Table 3.2: Girls' and boys' scores on the Visualization of Views and the shortened (20 question) version of the Revised PSVT:R pretests and post-tests, broken down by experimental condition. LA = *Little Alchemy* group, HB = *Homeworld Bound* group.

earnest [21, 76]. In order to counteract these gender differences before they become a barrier to STEM access later in life, it is important that a spatial skill training intervention works for girls at least as well as it works for boys. Therefore, I conducted two ANCOVAs, one for each of the two spatial skill tests used (PSVT:R and Visualization of Views), each with post-test score as the dependent variable, treatment condition (*Homeworld Bound* or *Little Alchemy*) as the independent variable, and pretest score and gender as covariates. Due to the small sample size, I was unable to analyze the interaction effect of gender with treatment condition (to investigate whether different treatments are equally effective for each gender). All effect sizes reported use ω^2 values since the sample size is small [173].

Inspection of diagnostic plots for the PSVT:R ANCOVA showed that the ANCOVA assumptions of normality, homoscedasticity, and linearity were met, and there were no outliers. However, the interaction effects analysis revealed a marginally significant interaction effect between pretest score and treatment condition ($F(1, 16) = 3.42$, $p = 0.083$), indicating a possible violation of the homogeneity of slopes assumption. However, using the Johnson-Neyman procedure [174] to determine a region of significance within the range of the pretest scores yielded no significant regions, and thus a violation of the homogeneity of slopes assumption could not be confirmed. The ANCOVA showed no effect of treatment condition ($F(1, 16) = 0.20, p = 0.66, \omega^2 = -0.018$) or gender ($F(1, 16) = 0.17, p = 0.69, \omega^2 = -0.019$). Only the test-retest effect was significant ($F(1, 16) = 26.74$, $p < 0.0001$), with a large effect size of $\omega^2 = 0.58$.

For the Visualization of Views ANCOVA, inspection of diagnostic plots and interaction effects analysis revealed that ANCOVA assumptions of normality, homoscedasticity, linearity, and homoegeneity of regression slopes were met. However, one observation was identified as an influential outlier (Cook's distance = 1.01) due to a dramatic improvement between pretest and post-test (Δ score = 9.4, in a test with a maximum score of 24) and was thus removed from the dataset. After removing the outlier, the ANCOVA revealed significant effects of treatment condition ($F(1,15) = 5.14$, $p = 0.039$), retesting ($F(1,15) = 28.10$, $p < 0.0001$), and gender ($F(1,15) = 8.63$, $p = 0.010$) such that those who played *Homeworld Bound* scored higher than those who played *Little Alchemy*, boys scored higher than girls on the post-test relative to the pretest, and there was a large improvement overall from pretest to post-test ($\omega^2 = 0.47$). Effect sizes for training condition ($\omega^2 = 0.07$) and gender ($\omega^2 = 0.13$) were medium according to the ranges specified by Field [175]. As shown in Figure 3.9, visualizing the effects of treatment condition and gender together revealed that the Visualization of Views test score improvement observed in the *Homeworld Bound* condition was a result of improvement in boys' scores only.

3.6.4 Improvements for the Final Version of Homeworld Bound

Playtests of the first two functional prototypes of *Homeworld Bound* revealed that children were overall quite engaged when playing the game. However, they found the step by step process for completing levels in Construction Mode confusing and often needed help from study proctors to understand how it worked. In addition, switching between Exploration Mode implemented in the *Warcaft III World Editor* and Construction mode implemented in Unity via passwords was clumsy, error prone, and took time away from playing the game.

Before executing my next study using *Homeworld Bound*, I made the following three improvements to the game to address these limitations. First, I replaced the single block stacking tutorial level in Figure 3.7 with two more targeted and complex tutorial levels, Tutorial 1 and Tutorial 2, each designed to ease students gradually into the process of building in Construction Mode. I briefly considered adding a step by step text tutorial that guided the player through the process of attaching one part from beginning to end, but discarded this idea after another pilot test at the local Orpheum Children's Museum revealed that children did not read the tutorial text and were therefore more confused than before the tutorial was implemented. I settled on a minimalist tutorial that simply informed players of the controls and let them explore and figure it out on their own. Given that proctors would be present for future studies with children, I figured they could help fill in any gaps of understanding during gameplay.

Second, to reduce children's confusion in Construction Mode further, I sought a way to make the game's controls feel more intuitive and natural. To make rotation feel more intuitive, one of my undergraduate collaborators came up with the idea of and implemented a set of 3D curved arrows aligned around the axes of a part that more intuitively conveyed how to rotate a part a certain way than the previous prototype's confusing and ambiguous representation of rotation actions via interface buttons labeled "Rotate X", "Rotate Y", and "Rotate Z." To make camera movement feel more natural, he replaced the interface buttons for rotating the camera in discrete units along two axes with a click and drag mechanic that allowed free 360 degree rotation around the center of the construction area. He also added a zoom feature to the camera that could be accomplished by simply scrolling up or down on the mouse. To make it less confusing for kids to imagine how the selected part must be rotated to align correctly with the construction, my undergraduate collaborator wrote code to make the selected part move so that it is next to the selected fusable area on the construction once a fusable area had been selected on both the construction and the selected part.

Third, with more time before the next study was launched and with the help of a new team of undergraduate students, I implemented Exploration Mode in the Unity engine so that both modes could be played together in one cohesive game as intended, with mode switching happening automatically at the end of each level or with the click of a button. This implementation also served to address a major limitation of the *Warcraft III* version of Exploration Mode. Due to the limitations of the *Warcraft III* engine, it was not feasible to implement first person perspective, and so the game utilized a top-down, bird's eye view, with the player controlling the character from above. In contrast, the commercial games shown most consistently to train spatial skills, such as the shooters *Medal of Honor* and *Unreal Tournament* and the puzzle game *Portal 2* are first person games. Thus, I made the Unity implementation of Exploration Mode in first person perspective to be more similar to the navigational tasks required of players in these other games.

The end result of these three modifications was the final version of *Homeworld Bound* as shown in Figure 3.4 and as described in Sections 3.2, 3.3, 3.4, and 3.5.

3.7 DISCUSSION

This chapter discussed the design and initial evaluation process for *Homeworld Bound*, a game designed to train the spatial skills of children in late elementary school as well as to evaluate the efficacy of specific features at tapping into players' existing spatial skill. The initial iterative design process used to refine the game was crucial to achieving these twin goals, both in the period leading up to the small scale training study and in the period afterward, where several major modifications allowed me to realize the implementation of every spatial feature proposed in my theoretical mapping between spatial subskills and game features.

While the results from the training study indicate that *Homeworld Bound* may have some potential as a spatial skill training intervention, these results must be interpreted with caution given the low sample size used in this study of about 10 children per experimental group. In addition, all participants attended a private elementary school, and this may have exacerbated the gender effects observed since students attending private schools tend to be from high socioeconomic status (SES) backgrounds, and gender differences in spatial skill are greater among high SES populations [108]. Furthermore, improvement was only observed on the perspective-taking test and not on the mental rotation test, although this may have been due to floor effects for girls' scores on the PSVT:R test. The observation that the training effect for the perspective-taking test in the *Homeworld Bound* condition was due to boys' improvement alone and not any noticeable improvement on the part of girls is troubling. While these results might have simply been due to noise given the very small sample size, it was also possible that *Homeworld Bound* may not have been as effective a training tool for girls as it was for boys.

Understanding the reasons for an ineffective cognitive training intervention is difficult without first analyzing the extent to which various levels of the game did or did not tap into the cognitive skills to be trained. In the next chapter, I describe a follow-up correlation analysis I conducted to understand the relationship between spatial skill, in-game performance and behavior, and gender across the different levels of *Homeworld Bound* and diagnose potential problems with the game design.

Chapter 4: Untangling the Relationship Between Spatial Skills, Game Features, and Gender

In this chapter, I describe a correlation study conducted to assess the extent to which performance and behavior in the game I designed to train children's spatial skills, *Homeworld Bound*, are associated with players' current level of spatial skill and gender. Since *Homeworld Bound* was designed to separate different spatial game features into different levels, examining the correlations between players' spatial skill and their performance in the game allows me to understand which levels are doing a better or worse job of tapping into spatial skills. If performance on certain levels is highly correlated with spatial skill, this may be an indication that these levels are taxing the appropriate cognitive resources. Conversely, if performance on certain levels is not correlated with spatial skill, this indicates that those particular levels may need to be redesigned to require the use of spatial skill in completing the level.

In addition, analyzing the relationship between spatial skill, gender, and player behavior allows me to investigate the extent to which the game may be tapping into the spatial skills of and engaging one gender over another. This is a particularly important investigation in light of the gender gap in spatial skills emerging in late elementary school that favors boys and mirrors the STEM gender gap. It is crucial that interventions designed to train spatial skills work at least as well for populations that tend to have lower spatial skill (such as girls) as they do for higher spatial skill populations (such as boys) in order to allow lower spatial skill populations the chance to catch up on spatial skill development before it becomes a barrier to access to STEM coursework, majors, and careers later in life.

This chapter's principal contributions to the research literature are the following: 1) the presentation of the first empirical study analyzing the potential of different game features to tap into children's spatial skills, 2) how gender affects the extent to which different game features tap into children's spatial skills, and 3) practical recommendations for implementing these features in games to assess or train players' spatial skills based on the complex relationship between gender, spatial skill, and in-game behavior my results reveal. This chapter contains a substantial portion of my previously published work [40], coupled with some additional analyses.

4.1 RESEARCH QUESTIONS

RQ1: On which levels of *Homeworld Bound* is player performance associated with spatial skill?

RQ2: Is the association between player performance and spatial skill stronger for more difficult levels of *Homeworld Bound*?

RQ3: How does the level-by-level relationship between player performance and spatial skill differ by gender?

4.2 METHOD

I recruited 20 children ages 7-12 via a publicly accessible weekly newsletter distributed by the University of Illinois at Urbana-Champaign to participate (accompanied by their parents) in a 3 hour study session involving cognitive testing and *Homeworld Bound* gameplay. I chose to recruit children ages 7-12 because previous research has established that the effect size of gender differences in spatial skill is significant and roughly uniform across this entire age range [21, 76].

After parents signed a consent form and children verbally assented to participate, the children took a shortened version of the Revised Purdue Spatial Relations Test (PSVT:R) [166] as a pretest of spatial skills. Since the PSVT:R is designed for people age 12 and up, I shortened it from 30 questions to just the easiest 20 questions (the first 20 since test questions are ordered by difficulty) as recommended to me by the researcher who developed the test. I used a single pretest of spatial skills rather than administering enough tests to cover all 4 quadrants of the spatial skill framework introduced in Chapter 3 because this study would take place in a single session and I did not want the children participating in the study to become mentally fatigued before starting their multiple hour gameplay session. Next, children filled out a short demographic survey on their age, gender, and previous video game experience. Game experience metrics included a quantitative ranking of video game play frequency (1=no experience, 6=daily play) as well as a free response question asking children to list the games they play most often.

After completing the survey, each child played the game for 2 hours or until they finished the game, whichever came first. While parents accompanied their children for the duration of the study, the study proctors minimized parent-child interactions during gameplay by asking parents to physically distance themselves from their children, although parents were allowed to come around briefly to see what their children were doing. Children were allowed to stop playing whenever they wanted. After finishing the gameplay session, children completed a post-game survey in which they rated how fun, boring, easy, and frustrating they found the game and why on a 5 point Likert scale. They were also asked what they would change if they were making the game. Parents were compensated $10 an hour per child and given a $20 bonus in addition to the base rate if they and their child stayed for the entire 3 hours.

4.2.1 In-Game Player Behavior Metrics

I collected a large amount of player behavior data from participants. Data collected focused primarily on time taken to finish each level, number of errors made, and behaviors associated with player impulsiveness. I reasoned that more impulsive players would spend less time standing around thinking about where to go next in Exploration Mode and would spend less time trying to rotate parts in Construction Mode to the correct alignment before attempting to attach them. I therefore measured impulsiveness by percentage of time each player spent standing still versus moving in Exploration Mode and how many rotations each player performed in Construction Mode before trying to attach two parts. Analyzing time taken to finish levels and number of errors gives me a sense for how difficult the game was, while measuring player exploration and errors made allows me to see to what extent spatial skills are associated with two approaches to problem solving that education literature has shown to be important for learning: the *growth mindset* and *exploring the problem space*.

The growth mindset is an attitude towards learning that views mistakes and failures as the best possible learning opportunities. It embraces the 'fail fast, fail often' adage championed by many Silicon Valley entrepreneurs and has been associated with increased player engagement and motivation in video games [176]. Exploring the problem space involves trying out many different strategies, approaches, or directions to allow for cross-pollination between them. This idea has gained attention in the design community for its propensity to increase creativity and improve solution quality [177].

Thus, each of the main player behaviors I measure allow me to get a sense for how players' spatial skills are related to the game's difficulty (time spent in each level, number of errors), and different types of player strategies (impulsiveness, growth mindset, exploring the problem space) when experiencing specific game features at different levels of granularity (between the two game modes versus between individual levels within each mode). The specific player behavior metrics I collected for each game mode are summarized in Table 4.1. An explanation of some of the more complicated metrics follows.

Wrong Face and Wrong Rotation Errors

There are two types of errors a player can make in Construction Mode when attempting to attach two parts together. If the player has selected the correct two faces to attach and has rotated the part to be attached so that it is aligned correctly, then the attachment attempt succeeds. However, if the player selects two faces that are not supposed to be attached in the finished object (A and B in Figure 4.1) and tries to attach them, this is a *wrong face*

Exploration Mode	*Construction Mode*
Time spent playing	Time spent playing
Number of times stopped	Sessions
% time spent stopped	Rotations
Avg time spent stopped	Face and rotation errors
Number of batteries	RPAA

Table 4.1: The in-game player behavior metrics collected in each mode of the game. RPAA = Rotations per attachment attempt.

error. If the player manages to select the correct two faces to attach but the rotation of the part to be attached is incorrect and the player tries to attach them (C and D in Figure 4.1), this is a *wrong rotation error*. Thus, errors are only counted when the player clicks the "Connect" button to attach two parts, and each attachment attempt will result in at most one error. Players' part rotation and face selection actions before an attachment attempt is made (clicking the "Connect" button) are considered exploratory actions and not errors.

Number of Sessions

Since rotation operations in Construction Mode were "powered" by batteries that could be collected in Exploration Mode, players who ran out of battery power after performing too many rotations would have to switch from Construction Mode to Exploration Mode, collect more batteries, and then switch back to Construction Mode to continue building where they left off. I recorded the number of times players had to make this switch as an additional metric of how much difficulty they had with each Construction Mode level and to what degree their Construction Mode play experience was broken up into smaller units of time versus chunked into larger sessions.

Rotations and Rotations per Attachment Attempt

Players must rotate parts in 3D in Construction Mode in order to line up the two faces of each part correctly for attachment. Each rotation action the player performs with the interface controls corresponds to a 90 degree rotation of the current active object along one of the X, Y, or Z axes. I calculated a player's rotations per attachment attempt in order to see to what extent players were exploring the problem space visually (doing a lot of rotations before making an attachment attempt) or had a more impulsive, growth mindset-oriented strategy (doing fewer rotations before making an attachment attempt).

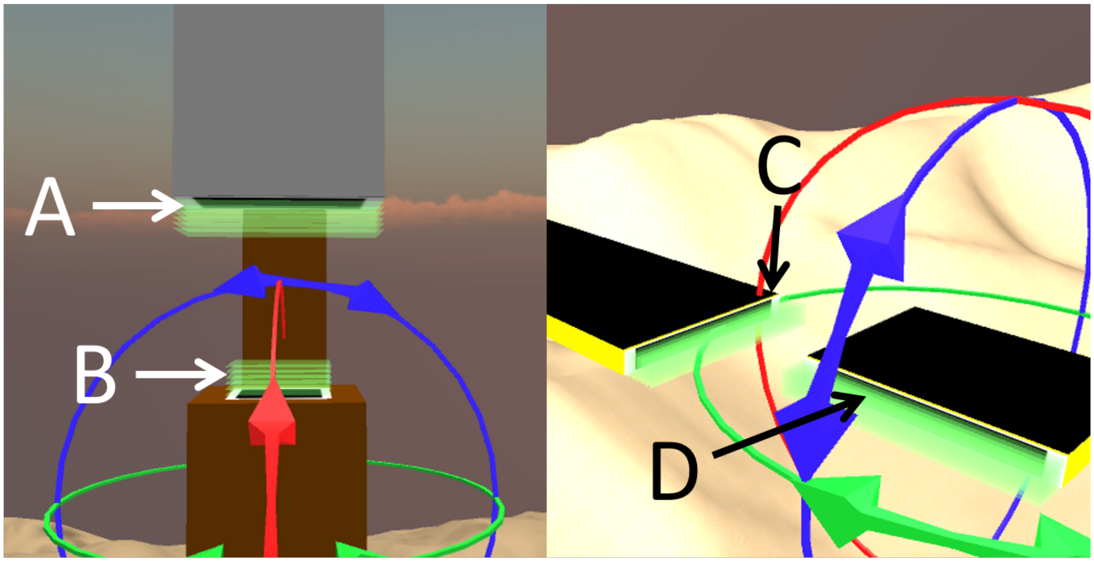


Figure 4.1: Faces A and B on the left are not the same size or shape, so trying to attach them would be a *wrong face error*. Faces C and D on the right have the same size and shape but are not yet aligned to face each other, so attaching them would be a *wrong rotation error*.

4.3 RESULTS

Of the 20 children who participated in the study, 10 were female, and ages ranged from 7 to 12 (median=10, mean=9.95). All but two had previously played video games, and all but one of those played games at least weekly. In addition, 11 (55%) were not able to complete the entire game. Of those who did not complete the entire game, 4 of them chose to quit early due to frustration, 2 quit early due to a prior commitment, and the remaining 5 played for the entire 2 hours but ran out of time to finish the game. Due to the relatively small number of players who made it as far as the Ruined City in Exploration and Construction Mode (n=9), I excluded data from these levels from my statistical analyses, leaving me with data from the first two Exploration Mode levels and the first four Construction Mode levels (Tutorial1, Tutorial2, and Rocket Boots in the Canyon and Sledgehammer in the Highlands).

The focus of my analysis was the correlation between in-game behaviors and children's scores on the spatial skill pretest. There were 20 questions on the pretest, and scores (number of questions answered correctly) were heavily skewed towards the low end ($\mu = 7.15$, median=6, $\sigma = 4.77$). The lowest score was 1 and the highest was 17.

To compare the effects of high-level features across the entire game and low-level features specific to certain parts of the game, I performed a hierarchical 2 stage correlation analysis,

Behavior	*Gender*	*Age*	*Pretest*	*Pretest (Boys)*	*Pretest (Girls)*
Time					
Exploration	n.s.	-0.71*	-0.55*	n.s.	n.s.
Canyon	n.s.	-0.65*	n.s.	n.s.	n.s.
Highlands	-0.48*	-0.62*	-0.57*	n.s.	n.s.
% Time Stood Still					
Total, Canyon, Highlands	n.s.	n.s.	n.s.	n.s.	n.s.
Avg Time Stopped					
Total, Canyon, Highlands	n.s.	n.s.	n.s.	n.s.	n.s.
Num. Times Stopped					
Total, Canyon, Highlands	n.s.	n.s.	n.s.	n.s.	n.s.
Num. Batteries					
Total, Canyon, Highlands	n.s.	n.s.	n.s.	n.s.	n.s.

Table 4.2: Behavior, demographic, and spatial skill pretest correlations using Spearman's ρ for Exploration Mode levels. $*p < 0.05$, n.s. = non-significant correlation. Gender was coded as 1=Female, 2=Male, so a positive correlation indicates a behavior associated more with males.

starting with mode-level player behaviors and then breaking them down further level-by-level in the second stage. My primary measures of interest for player behavior were time taken to finish all levels, errors made, and impulsiveness (as measured by percentage of time spent standing still versus moving and rotations per attachment attempt). The concrete metrics I used for each of these behaviors, which were the same for each stage of my hierarchical analysis, are summarized in Table 4.1. I will first present general results for all participants, and then investigate the effects of gender in later sections. My complete correlation analysis between spatial skill pretest scores, player behaviors, and demographics is summarized in Tables 4.2 and 4.3. I used Spearman's ρ for all analyses since pretest score and most behavioral measures had highly skewed, non-normal distributions. In addition, I used a p-value of 0.05 for all tests of statistical significance for correlations without a Bonferroni correction for multiple comparisons. I chose not to include an adjustment for multiple comparisons due to the already low power of my sample of $n = 20$ participants.

Behavior	*Gender*	*Age*	*Pretest*	*Pretest (Boys)*	*Pretest (Girls)*
Time					
Total, Tutorial1	n.s.	n.s.	n.s.	n.s.	n.s.
Tutorial2	n.s.	-0.50*	n.s.	n.s.	n.s.
Rocket Boots	n.s.	-0.68*	n.s.	n.s.	n.s.
Sledgehammer	n.s.	n.s.	n.s.	-0.64*	n.s.
Rotations					
Total, Tutorial1, Tutorial2	n.s.	n.s.	n.s.	n.s.	n.s.
Rocket Boots	n.s.	-0.58*	n.s.	-0.67*	n.s.
Sledgehammer	n.s.	n.s.	-0.56*	-0.94*	n.s.
Combined Errors					
Total	n.s.	n.s.	n.s.	-0.85*	n.s.
Rocket Boots	n.s.	-0.52*	n.s.	n.s.	n.s.
Tutorial1, Tutorial2, Sledgehammer	n.s.	n.s.	n.s.	n.s.	n.s.
Wrong Face					
Total	n.s.	n.s.	n.s.	-0.92*	n.s.
Tutorial1	n.s.	n.s.	n.s.	-0.67*	n.s.
Tutorial2	+0.49*	n.s.	n.s.	n.s.	n.s.
Rocket Boots, Sledgehammer	n.s.	n.s.	n.s.	n.s.	n.s.
Wrong Rotation					
Rocket Boots	n.s.	-0.66*	-0.45*	-0.65*	n.s.
Tutorial1, Tutorial2, Rocket Boots, Sledgehammer	n.s.	n.s.	n.s.	n.s.	n.s.
RPAA					
Total, Tutorial1, Tutorial2, Boots, Sledgehammer	n.s.	n.s.	n.s.	n.s.	n.s.
Num. Sessions					
Total	n.s.	n.s.	n.s.	n.s.	n.s.
Tutorial1, Tutorial2, Rocket Boots, Sledgehammer	n.s.	n.s.	n.s.	n.s.	n.s.

Table 4.3: Behavior, demographic, and spatial skill pretest correlations using Spearman's ρ for Construction Mode levels. $^*p < 0.05$, n.s. = non-significant correlation. Gender was coded as 1=Female, 2=Male, so a positive correlation indicates a behavior associated more with males. RPAA = Rotations per Attachment Attempt.

4.3.1 Exploration Mode Taps Into Spatial Skills

First, I analyzed the relationship between high-level player behaviors across all levels in Exploration and Construction Mode and pretest scores. Total time spent in Exploration Mode had a significant negative correlation with pretest score ($\rho = -0.55$, $p = 0.016$). None of the other high-level player behaviors were significantly correlated with pretest scores. Thus, children with higher pretest scores tended to finish Exploration Mode levels more quickly. Since time spent in Construction Mode as a whole was not significantly associated with pretest score, this result suggests that Exploration Mode as a whole requires more spatial skill than Construction Mode as a whole.

4.3.2 Specific Levels Tap Into Spatial Skills

To get a more detailed picture of the extent to which each game mode tapped children's spatial skills, I analyzed pretest scores and player behavior level-by-level in both Construction Mode and Exploration Mode.

In Exploration Mode, pretest score had a significant negative correlation with time spent in the Highlands ($\rho = -0.57$, $p = 0.01$) but no significant correlation with time spent in the Canyon level. Thus, only the Highlands level appears to be tapping into children's spatial skills.

In Construction Mode, number of wrong rotations performed in the Rocket Boots level had a significant negative correlation with pretest score ($\rho = -0.45$, $p = 0.047$). For the Sledgehammer level, pretest score and total number of rotations had a significant negative correlation ($\rho = -0.56$, $p = 0.015$). Thus, the Rocket Boots and Sledgehammer both tapped into players' spatial skills, but in different measures of performance. There was no significant relationship between performance in the Tutorial 1 and Tutorial 2 levels and spatial skill.

4.3.3 Gender and Age Differences

Boys ($\mu = 9.2$, median=8, $\sigma = 5.05$) scored higher on the pretest of spatial skills than girls ($\mu = 5.1$, median=3.5, $\sigma = 3.63$), but this difference was not significant ($t = -2.08$, $p = 0.053$). Since the pretest consisted of 20 multiple choice questions with 4 answer choices per question and 1 correct answer per question, it is interesting to note that boys' average performance was above chance, but girls' was not. While the difference in performance between boys and girls did not reach statistical significance, the direction of the difference is consistent with previous research establishing gender differences, not only in spatial skill but also on the particular psychometric test I used [78, 178]. It is likely that with a larger sample, a statistically significant difference in performance by gender would be detected.

Research has shown that boys tend to have higher spatial skill than girls and that this difference begins to emerge when children are around the same age as the children in my study [21]. Therefore, I investigated the extent to which there were gender or age differences in player behaviors that might influence how effectively the game taps into spatial skill for different demographics. Tables 4.2 and 4.3 present a complete summary of all player behavior, gender, and age correlation analyses I ran at both the mode-level and level-by-level. I used Spearman's *rho* for all analyses since I was analyzing a categorical variable (gender) and the behavioral measures tended to have highly skewed, non-normal distributions.

Age: Younger Players Find Game More Difficult

Age was associated with several mode-level behaviors. At the mode level, younger players spent more time in Exploration Mode and spent more time standing still in Exploration Mode. Age was also associated with level-by-level behaviors. Time spent in every level of Exploration Mode and the Tutorial 2 and Rocket Boots levels of Construction Mode had a significant negative correlation with age. Younger players also used more rotations and made more wrong rotation errors in the Rocket Boots level. These results demonstrate that Exploration Mode levels are harder for younger players, as is the Rocket Boots level and to a lesser extent the Tutorial 2 level in Construction Mode.

Gender: Girls take More Time, Boys Make More Errors

There were no significant correlations between gender and any mode-level player behaviors. However, there were a few gender differences in behavior in individual game levels. Girls took longer to complete the Highlands level, while boys made more wrong face errors in the Tutorial2 level. While these gender effects occur only for a small subset of behaviors in a small subset of the game's levels, they may indicate that girls and boys are using different strategies when they play certain levels.

For instance, since girls took longer to complete the Highlands level, they may have spent more time deciding where to go next than boys, who may prefer to act more impulsively, choosing a direction and walking there without worrying about whether it is the optimal direction to choose. Similarly, boys make more wrong face errors in the second tutorial level of Construction Mode, possibly because girls spend more time thinking about whether they matched up parts correctly before even trying to attach. This more impulsive, exploratory behavior on the part of boys and more premeditated, careful behavior on the part of girls may be related to spatial skill since time spent in the Highlands level was also related to spatial skill pretest scores.

4.3.4 Gender Effects on Relationship Between Spatial Skill and Performance

Since gender affected players' in-game behavior and pretest score, it may also affect the predictive power of spatial skill pretest scores on certain player behaviors. I therefore performed another correlation analysis between player behaviors in different parts of the game and pretest scores, but broken down by gender and age.

Boys' Pretests, Not Girls', Predictive of Behavior

For boys at the mode level, total number of errors and total number of wrong face errors had significant negative correlations with pretest scores ($\rho = -0.85$, $p = 0.0018$ and $\rho = -0.92$, $p = 0.0002$, respectively). Here, we see additional correlations missing from the combined gender sample: total errors and total wrong face errors. Thus, it appears that matching the correct two faces for part attachment tapped into boys' spatial skill in Construction Mode. However, unlike in the combined gender correlation analysis, the correlation between time spent in Exploration Mode and spatial skill failed to reach statistical significance for boys. Surprisingly, there were no significant correlations between mode-level behaviors and pretest scores for girls. Given that time spent in Exploration Mode only had a statistically significant relationship with spatial skill in the combined gender sample, it is likely that the lack of statistical significance for this relationship in either the boys' or girls' sample is a consequence of the loss of statistical power in these small samples ($n = 10$ each).

For boys, level-by-level analysis revealed that pretest scores were correlated with behavior in three different Construction Mode levels (Tutorial 1, Rocket Boots, and Sledgehammer) and no Exploration levels. Some of these correlations had also been observed in the combined gender sample. For the Rocket Boots level, as with the combined gender sample, there was a significant negative correlation between number of wrong rotations and pretest score, but it was stronger for boys ($\rho = -0.65$, $p = 0.04$). Likewise, the correlation between number of rotations used in the Sledgehammer level and pretest score was significant and stronger for boys ($\rho = -0.94$, $p < 0.0001$). The correlation between time spent in the Highlands Exploration Mode level and pretest score observed in the combined gender sample failed to reach statistical significance in the boys sample, likely due to reduced statistical power as a result of cutting the sample in half by gender.

Other relationships between pretest score and level-by-level performance that were not present in the combined gender sample emerged in the boys sample. Boys' pretest scores had a significant negative correlation with the number of wrong face errors in the Tutorial 1 level ($\rho = -0.67$, $p = 0.036$), and boys' pretest scores had a significant negative correlation with number of rotations in the Rocket Boots level ($\rho = -0.67$, $p = 0.034$). In addition, boys' pretest scores had a significant negative correlation with time spent in the Sledgehammer level ($\rho = -0.64$, $p = 0.046$).

In contrast, girls' pretest scores were not significantly associated with any level-by-level behaviors.

4.4 DISCUSSION

In this section, I discuss the results of my analysis of the relationship between spatial skill and player performance in different levels of *Homeworld Bound* (**RQ1**), the extent to which this relationship is dependent on a level's difficulty (**RQ2**), and the extent to which this relationship is dependent on a player's gender **RQ3**. I then conclude by discussing the design implications of these findings for spatial skill training video games and potential avenues for future work.

4.4.1 RQ1 and RQ2: What Homeworld Bound levels tap into players' spatial skills?

Exploration Mode

I found that completion time was related to spatial skill the Highlands level, but not in the Canyon level. This may be due to the fact that the Highlands level is larger than the Canyon and has a more open structure, requiring players to rely more on landmarks than nonstrategic wandering to navigate and find parts. Thus, players may be required to employ *landmark orientation* more in the Highlands level due to the increased number of possibilities for where parts may be and the need to remember where the player has already looked by reference to landmarks.

Another possible explanation is that players with higher spatial skill have an easier time jumping with the Rocket Boots, a skill required much more often in the Highlands than in the Canyon since many batteries in the Highlands level are accessible only via jumps. Many of the children frequently reported having difficulty with the jumps, and since jumping mechanics play a significant role in some games that have been demonstrated to improve spatial skills, such as *Portal 2* and *Super Mario* [5, 28], the simple acts of judging distance

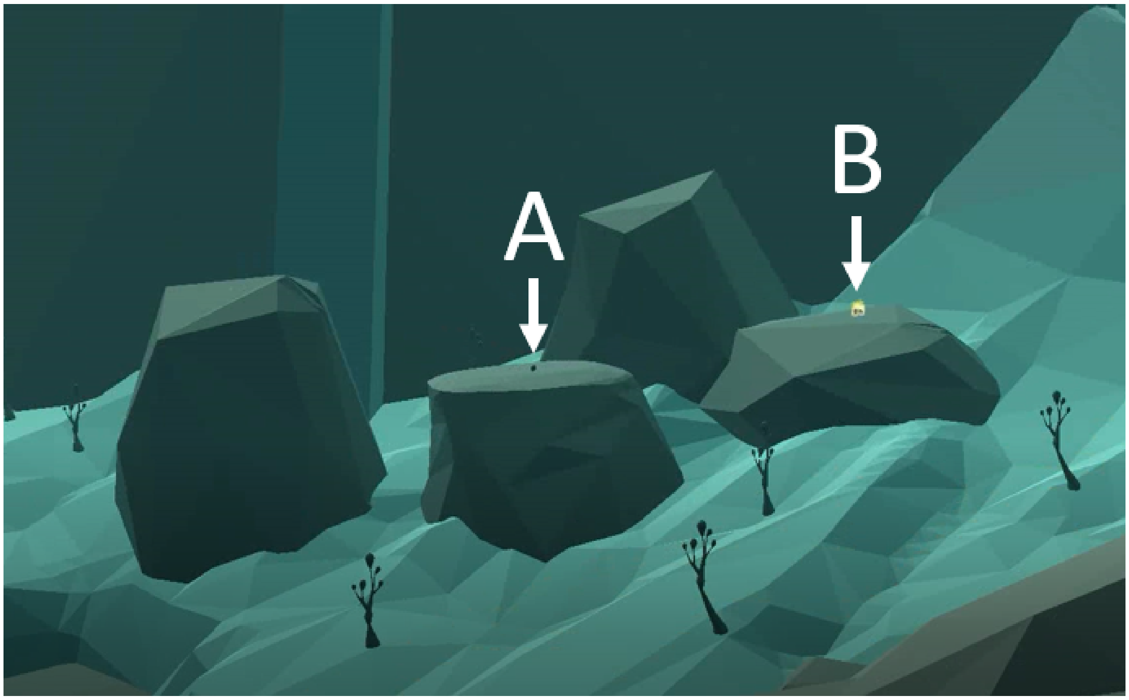


Figure 4.2: An example of a jumping sequence requiring *navigation visualization* in the Highlands level of *Homeworld Bound*. If the player wants to reach the battery (A), they must first jump onto the shorter rock on the right (B) since rock A is too tall to be reachable in one jump from the ground. The glowing object on rock B is an item part.

and controlling trajectory involved in such jumps may be taxing children's spatial skills as well. Thus, the jumping mechanic may simply make the Highlands level more difficult and therefore tax players' spatial skills more. In addition, some batteries were located on top of rocks that were too tall to be reached in one jump, so players had to figure out what shorter rocks nearby to jump on first in order to be high enough to get to the top of the tallest rock in one jump (see Figure 4.2 for an example). In short, players likely needed to spend time planning and picturing their route from the ground to the top of a rock via a series of jumps (*navigation visualization*). Thus, the Highlands level may be successful at tapping into players' spatial skills where the Canyon level is not due to the Highlands level's implementation of more vertical navigation and tasks requiring *navigation visualization*.

Construction Mode

I found that while there was no relationship between performance in the Tutorial 1 and Tutorial 2 levels and spatial skill, performance on both the Rocket Boots and Sledgehammer levels was associated with spatial skill in different ways. Spatial skill was associated with

Level	*Performance Metric*	*Mean*	*Median*
Tutorial 1	Completion Time	327	330
Tutorial 2	Completion Time	171	138
Rocket Boots	Completion Time	398	320
Sledgehammer	Completion Time	251	186
Tutorial 1	Total Rotations	37	37
Tutorial 2	Total Rotations	23	17
Rocket Boots	Total Rotations	39	31
Sledgehammer	Total Rotations	32	24
Tutorial 1	Wrong Face Errors	1	1
Tutorial 2	Wrong Face Errors	0	0
Rocket Boots	Wrong Face Errors	9	4
Sledgehammer	Wrong Face Errors	3	1
Tutorial 1	Wrong Rotation Errors	2	1
Tutorial 2	Wrong Rotation Errors	1	0
Rocket Boots	Wrong Rotation Errors	9	5
Sledgehammer	Wrong Rotation Errors	8	3

Table 4.4: Mean and median performance in each of the Construction Mode levels in *Homeworld Bound*. Completion times are in seconds. All numbers are rounded to the nearest integer.

number of wrong rotations in the Rocket Boots level, while in the Sledgehammer level, spatial skill was associated with the total number of rotations used. One possible reason is that the Rocket Boots level was more difficult than the Sledgehammer level, especially since players had to complete the Rocket Boots level before the Sledgehammer level and thus would have had more practice with Construction Mode levels before attempting the Sledgehammer level.

Player performance data supports the notion that the Rocket Boots level was the hardest Construction Mode level. As Table 4.4 shows, there was high variance in individual player performance on each Construction Mode level, but regardless of whether the median or mean is used, players made many more wrong face and wrong rotation errors on the Rocket Boots level than on other levels. The amount of time and rotations needed to complete the Rocket Boots level also appears to be much higher than in the Tutorial 2 or Sledgehammer levels, although these performance metrics were not significantly related to players' spatial skill.

Interestingly, time spent and rotations used were similar between the Tutorial 1 and Rocket Boots levels. The difficulty of the Tutorial 1 level in terms of time and rotations but not errors, as well as the lack of correlation between performance and spatial skill in this level, suggests that its difficulty may have mainly been due to players getting used to the controls

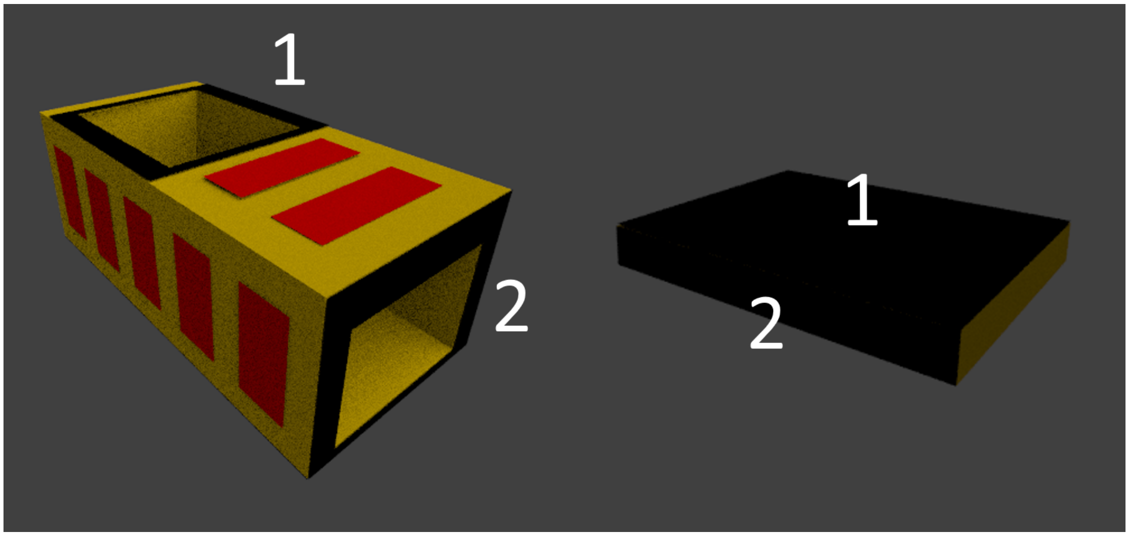


Figure 4.3: The two most difficult parts to attach in the Rocket Boots Construction Mode level, the main body (left) and the sole of the toebox (right). Players had to use the location of two different attachment faces (1 and 2 on both parts) to correctly align and attach these parts.

and learning how to play, since it was the first Construction Mode level encountered. Thus, Tutorial 1's spatial elements may have been simple enough to not tax players' spatial skills at a detectable level. Tutorial 2 may have failed to tax players' spatial skills for the same reason since it seems to have been the easiest Construction Mode level of all.

Similarly, it is interesting to note there was no relation between number of wrong face errors and spatial skill in the Rocket Boots level, despite the large number of wrong face errors players made in this level. One possible reason for this may be that there are a large number of very similarly shaped fusing areas on parts in this level (6 in total, each shaped like a roughly square picture frame). If players did not pay attention to the reference image of the finished item in the upper left corner, they might resort to brute force, trying to fuse each part on every single one of these similar areas until the game switched from giving them wrong face errors to wrong rotation errors.

Despite being designed to be even harder than the Rocket Boots level, the Sledgehammer level seemed to be easier than the Rocket Boots level, although harder than the tutorial levels for the most part based on Table 4.4. My observation that children had a lot of difficulty attaching two parts in particular in the Rocket Boots level may explain why. As Figure 4.3 shows, both parts had two different attachment regions, and the player needed to note the location of both of them in order to determine how each part should be rotated to correctly align it with the part they needed to attach it to. The Sledgehammer level, while utilizing more complex part shapes and more parts than the two tutorial levels, did not have parts requiring this more complex, multiple step *object alignment* task, which may explain why players' spatial skill was related to the number of wrong rotation errors they made in the Rocket Boots level but not in the Sledgehammer level.

However, the total number of rotations used was related to spatial skill in the Sledgehammer level but not the Rocket Boots level. Thus, predicting the result of a given rotation operation (*object rotation*) may have been more spatially demanding in the Sledgehammer level. This is perhaps due to the more complex part shapes used in the Sledgehammer level; the Rocket Boots level had primarily solid or hollowed out cuboids, whereas the Sledgehammer introduced triangular and trapezoidal prisms for the first time. Thus, the Rocket Boots level seems to tap into players' spatial skills via more difficult *object alignment* tasks, whereas the Sledgehammer level seems to tap into players' spatial skills via more difficult *object rotation* tasks.

Summary

In summary, my results indicate that all of the spatial game features I hypothesized would tap into players' spatial skills in Chapter 3, as implemented in *Homeworld Bound*, did indeed tap into players' spatial skills in certain Construction Mode and Exploration Mode levels (**RQ1**). I found that fewer wrong rotation errors in the Rocket Boots level as well as fewer rotations in the Sledgehammer level (corresponding to the game features *object alignment* and *object rotation*) were correlated with higher spatial skill. I also found that less time spent in the Highlands level was correlated with higher spatial skill, suggesting that the combination of *landmark orientation* and *navigation visualization* tasks required for the Highlands level tapped into players' spatial skill. The greater complexity of navigation tasks and the implementation of *navigation visualization* tasks in the Highlands level may explain why this level was successful at tapping into players' spatial skills where the Canyon level was not.

Level difficulty alone was not enough to tap into players' spatial skills, as evidenced by the lack of relationship between spatial skill and performance in the Tutorial 1 level, which was about the same level of difficulty for players as the Rocket Boots level in terms of completion time and total rotations used. This suggests that the difficulty of certain spatial operations contained in a level, such as those in my four feature model (*object alignment*, *object rotation*, *landmark orientation*, and *navigation visualization*), is more important in determining the extent to which it taps into players' spatial skill than general difficulty alone (**RQ2**). Examining the difficulty of each spatial operation on a level-by-level basis can therefore be used as a diagnostic tool for identifying how certain levels might be modified to improve their ability to tax players' spatial skills. The more each level in a spatial skill training game is able to tap into players' spatial skill during gameplay, the more likely it is to actually produce measurable training effects over time.

4.4.2 RQ3: Gender Differences in Correlations and Behavior

The finding that girls' spatial skill pretest scores did not predict any level-by-level or mode-level behaviors, while boys' pretest scores predicted five level-by-level and two mode-level behaviors, is very surprising (**RQ3**). It seems that the spatial features as implemented in my game may not be tapping into girls' spatial skills, but it is not clear why.

First, I wondered if girls were simply not as interested in the game as boys and therefore used less strategic behaviors when playing. To see if this was likely, I analyzed Spearman correlations between gender and children's self-reported 5-point Likert scale measures of fun, easiness, boredom, and frustration from the post-game survey. There was no significant correlation between gender and any of the four self-report measures. I also examined children's explanations for why they rated fun, easiness, boredom, and frustration the way they did and what they would change about the game, but there did not appear to be a distinctive gender difference in the responses. Therefore, girls did not seem to have a more negative experience than boys.

If the girls in my study were less familiar with construction and first person exploration games like *Homeworld Bound*, this may have caused the difference in girls' and boys' in-game strategies. I analyzed boys' and girls' self-reported lists of games they played to see if there was evidence for this hypothesis. Out of 20 children, 15 provided information about the games they played. Both genders reported playing a diverse set of games, including racing, first person shooter, construction, sports, and puzzle games. The only noticeable difference was that 7 boys (a majority) reported playing construction games like *Minecraft* and *Roblox*, whereas only 3 girls (a minority) did. Both *Minecraft* and *Roblox*, like *Homeworld Bound*, allow the player to explore and collect materials from a virtual world and use those materials to build objects. Since fewer girls had experience with this kind of game, perhaps they spent more time familiarizing themselves with the controls and play style, which may have affected their in-game behavior.

Another possibility is that girls may be using different strategies while playing to reduce cognitive load. For example, girls' longer completion times for the Highlands level may be due to the use of less cognitively taxing nonspatial strategies to circumvent this difficulty (such as random wandering until all parts are found). However, without additional data beyond these quantitative measures, it is difficult to know for certain why girls took longer to complete the Highlands level.

A second and perhaps more likely explanation for the observed gender differences is that the particular psychometric test I used for the pretest, the Revised Purdue Spatial Visualization Test: Rotations (PSVT:R), may not have been sensitive enough to capture differences in spatial skill among children with low spatial skill - those who performed no better than chance, and who were disproportionately girls. I chose this test because to the best of my knowledge, no spatial skill tests currently exist for children in the age range I targeted with this study. Every spatial skill test I could find was designed either for very young children (ages 0-3) or for adults. I felt that the PSVT:R was the best possible option given that it was designed for subjects ages 12 and up. I attempted to reduce the difficulty of the test by eliminating the 10 hardest questions, but it is likely that the test needs to be made even easier to accurately assess the spatial skill of children in late elementary school.

4.4.3 Design Implications

These findings have several important implications for game designers interested in assessing their players' spatial skill. First, spatial skills do seem to be correlated with different dimensions of player performance in Exploration Mode and Construction Mode as a whole, as well as in individual levels within each mode. The Highlands level in Exploration Mode and the Rocket Boots and Sledgehammer levels in Construction Mode seem particularly effective at tapping into players' spatial skills.

Given that the most likely reasons for the stronger correlation between performance and spatial skill in the Highlands level are its larger, more open, more vertical, and more complex navigation requirements overall, game designers may therefore consider incorporating large, open spaces that require players to recognize landmarks from many different angles in order to most effectively tap into players' spatial skills via *landmark orientation*, or possibly include more tasks involving planning of multiple step navigation, such as that required to make a series of vertical jumps to reach a battery, to get players to exercises their *navigation visualization* skills more. To determine which explanation - larger level, more open level structure, more vertical navigation (jumping), or some combination thereof - is more likely, it is necessary to further isolate these different subfeatures of the Highlands level in separate levels so the effect of each feature on the relationship between spatial skill and performance can be isolated. This is an excellent direction for future work but lies outside the scope of this dissertation.

Since the Rocket Boots level's strongest relationship between performance and spatial skill is in players' spatial skills via *object alignment*, whereas the Sledgehammer level seems to tap into players' spatial skills via *object rotation*, giving each level a unique feature found in the other may help make both more spatially demanding. For instance, the use of more complex parts in the Rocket Boots level may require players to perform more spatially taxing *object rotation* tasks, whereas including parts whose successful alignment depends on attending to multiple areas on the construction at once may increase the amount of spatial cognition necessary for the level's *object alignment* tasks. In Chapter 6, I discuss how I implemented both of these modifications to make these levels more cognitively taxing for players. However, more complex part shapes and more complex part alignment tasks are features that could be incorporated into any construction-based spatial skill training game, not just *Homeworld Bound*.

In Exploration Mode, only the more complex, more vertical first person navigation in the Highlands level tapped into players' spatial skills, while the very simple first person navigation in the Canyon level did not. Like the Highlands level, many other games that have shown spatial skill training effects require the player to navigate an environment in first person by either walking or jumping: *Medal of Honor, Unreal Tournament,* and *Portal 2*, for example. However, *Portal 2* includes many additional game mechanics, such as aiming and shooting a "portal gun" at walls to create portals for spatial teleportation and disabling enemy turret guns that shoot at the player.

The fact that the Highlands level includes none of these additional navigational features yet still manages to tap into players' spatial skill suggests that the act of first person navigation alone (provided the task is sufficiently difficult) may be enough to make first person games like *Medal of Honor* and *Unreal Tournament* effective at training spatial skills. This makes intuitive sense since first person exploration is how we experience the world around us as humans. However, it still may be the case that the more complex navigational requirements in *Portal 2* (portal-based navigation, see Chapter3, Section 3.3 for a more detailed explanation) tax players' spatial skills more than more basic navigation, just as the more open, larger level design of the Highlands level and its correspondingly more difficult navigation tasks may have contributed to the significant relationship between spatial skill and performance in this level in a way that the simpler navigation in the Canyon level could not. This notion is supported by a study by Castell et al., who found that introducing proximal (closer) landmarks into a navigation game reduced players' need to use mental rotation when playing the game [155].

While *Tetris* and Construction Mode both require the player to rotate and fit blocky shapes together, there are also a number of substantial differences between the two games. *Tetris* is a 2D game with possible rotations along only one axis, while Construction Mode allows the player to make rotations along each of the X, Y, and Z axes in 3D. In addition, there is an inherent time-sensitivity to actions in *Tetris* (act too slowly and your blocks will be placed very poorly, resulting in a game over before too long), with the result that expert Tetris players tend to rotate *Tetris* blocks very quickly in order to assess all of their possible options before they run out of time to place the block [179]. There is no such time sensitivity in Construction Mode. In fact, Construction Mode encourages careful, deliberate thought when rotating objects due to the constraint that each rotation uses up battery power, which the player can only replenish by leaving Construction Mode to collect more batteries.

The only feature common to both *Tetris* and Construction Mode is the requirement of rotating and fitting objects together. The ability of Construction Mode, and especially the Rocket Boots level, to tap into players' spatial skills despite its few similarities to *Tetris* suggests that the simple act of rotating objects and deciding how to fit them together taps spatial skills in different types of spatial environments and in games with different priorities for speed and accuracy. Therefore, *object rotation* and *object alignment* appear to be good features to include in a game for assessing and training players' spatial skill and are generalizable to many different types of games, although construction-based and puzzle games are probably the most natural fit for these two features.

Given my finding that girls spent more time in the Highlands level than boys and boys made more wrong face errors than girls in the Tutorial 2 level, it may be that boys and girls tend to utilize different strategies to complete each of these levels. However, with only quantitative timing data, it is difficult to test this explanation. Collecting additional data about players in future studies, such as path traces in Exploration Mode levels, post-game interviews, or questionnaires assessing players' strategy use in the game, could be one way of analyzing the reasons for gender differences in navigational strategies and behavior in each level. In addition, it is important to note that while more exploratory and risk-taking behavior (such as boys' greater wrong face error rate) is beneficial up to a certain point in a learning experience, too little forethought about one's actions could result in random, nonstrategic actions that help the player brute force their way to success but do not tap into the (spatial) skills that the intervention is designed to train. Similarly, waiting too long to make a move for fear of making a mistake can hinder the learning process. Thus, a balance is needed between the more implusive and more cautious behavior I observed in boys and girls, respectively. One way to encourage players to achieve this balance might be to implement a machine learning algorithm that detects when the player is acting too cautiously or too impulsively and either reduces the cost of errors (for more cautious players) or increases the cost of errors (for more impulsive players).

Lastly, gender differences in player behavior and the degree to which pretest scores predicted in-game behaviors indicate that game designers need to consider gender differences in both spatial skill and player behavior if they want to ensure that a spatial skill training intervention works at least as well for girls as for boys. The data collected in my study do not shed much light on the mystery of the gender differences in correlations I observed; girls did not seem to be less motivated to play the game and were only a little less familiar with similar games than boys. There is a need for future work exploring in more detail not just what, but why these gender differences in the relationship between spatial skill and performance exist and if they are consistent across different age groups or found only in the particular age range or demographic of children I studied.

4.4.4 Limitations and Future Work

The biggest limitation of this study is its low sample size of $n = 20$ children. With a sample of this size, my analysis lacked sufficient power to detect medium or small effects (with 20 participants and a 95% confidence level, I had an 83% chance of detecting an effect of size $r = 0.6$ or higher). Since increasing the confidence level via an adjustment for multiple comparisons would have further reduced power, I elected not to adjust for multiple comparisons. However, given that I performed a total of 210 statistical comparisons in my statistical analysis, it is very likely that a good portion could be spurious findings and therefore the findings in this chapter should be treated as preliminary. In addition, my sample was likely biased towards children of University of Illinois staff given my recruitment method. Additional correlation studies with larger, more diverse samples are needed to confirm these findings and determine the extent to which they generalize to other populations. Furthermore, while my study was able to identify certain features in *Homeworld Bound* that tap in to children's spatial skills, it cannot verify that these features actually train spatial skills. A controlled study with pre- and post-tests of spatial skill is needed to establish a causal relationship between specific game features and spatial skill development.

In Chapter 7, I address each of these limitations in a controlled study analyzing the training effects of a newer version of *Homeworld Bound* and its level-by-level ability to tap into players' spatial skills with a large sample of low spatial skill college students.

This work had three other limitations that could be addressed in future work outside of the scope of this dissertation. First, using completion time, time spent standing still, and number of batteries collected as the sole measures of performance in Exploration Mode may have caused my analysis to miss connections between more sophisticated measures of in-game performance and spatial skill. One such measure might be the efficiency of paths taken through the level to collect items; the shorter the path and the fewer times the player returns to an area they had already checked before, the more likely it is that the player may be using *landmark orientation* skills to engage in strategic navigation as opposed to the brute force method of random wandering that is unlikely to tap into players' spatial skills. Future work could investigate the implementation of automatic path trace metrics to facilitate their analysis as measures of strategic performance.

Another limitation of my study was that the incentive structure used for participation (hourly compensation and a bonus for staying the full 3 hours, both monetary) is that it may have caused children who got bored or frustrated and wanted to stop early to keep playing the game, but without motivation to progress, potentially affecting their performance in later levels of *Homeworld Bound*. Future work could utilize an incentive scheme that, in addition to providing a base amount for participation, gives participants extra rewards for progress made in the game rather than simply the amount of time spent on it to discourage unmotivated, low effort gameplay.

A final limitation of this work is that the ability to detect relationships between children's spatial skill and performance in *Homeworld Bound* may have been limited by floor effects on the spatial skill pretest, especially for girls. Future work studying children's spatial skills should investigate the use of a more sensitive pretest of spatial skill than the PSVT:R (such as the MRT-Animals or MRT-Letters mental rotation tests used by Jansen et al. and Neuburger et al. [180, 23]) to determine whether the gender difference in the predictive power of player behavior on pretest scores is due to boys and girls using different strategies, floor effects for girls' PSVT:R test scores, or some other cause.

Chapter 5: A Player-Centric Approach to Designing Spatial Skill Training Games

Designing spatial skill training games that contain the right set of features to tap into players' spatial skills and develop them through gameplay is essential for them to be effective as interventions, but it is not enough. Prior research has found that game-based spatial skill training interventions and more traditional spatial workbook exercises are about equally effective [1]. Thus, the main advantage to going to the trouble of developing a functional game-based intervention lies in the motivational power of video games. Good video games naturally incorporate experiences that promote players' feelings of *competence* and *autonomy*, which have been shown to promote motivation, enjoyment and feelings of well-being in players [25, 66]. Without this motivational benefit, there is little reason to invest the time and effort into creating a game.

Furthermore, the kind of games one person likes may be very different from what another likes, which in turn may affect their motivation to keep playing, or to start playing in the first place. While in-school or laboratory interventions, in which the learner receives course credit or monetary compensation for participation, may motivate the learner to play via these extrinsic incentives, those who do not enjoy the game being played are not likely to continue to play it long term of their own volition outside of the classroom or laboratory. In addition, their motivation during an in-school or laboratory intervention may be lower than those who enjoy the game more, causing them to potentially not try as hard, not progress as much, and not learn as much. Therefore, it is essential to understand the preferences and motivations of the target audience during the game's design process to ensure that the target audience can reap the full motivational benefits of the game-based intervention.

I argue that students with the lowest levels of spatial skill should be the target audience for spatial skill training games given that they stand the most to benefit from them. Students with low spatial skill tend to struggle in introductory courses and are more likely to drop out of STEM majors - unless they can bring their spatial skills up to a certain "threshold" of ability that gets them through early STEM coursework [1, 10]. This is especially a problem for female students given the consistent gender gap in spatial skills [21, 77, 22], which may contribute to the gender gap in many STEM fields [8]. Bringing low spatial skill students' spatial skills up to a certain threshold could be one way of reducing this gender gap in STEM, and in general allowing more students who otherwise might drop out to continue in STEM majors and then on to STEM careers in the future.

Unfortunately, most of the games that have been successful at training players' spatial skills in the laboratory may be most appealing to the subset of the population that already has higher levels of spatial skills: male action video game players. Most video games shown empirically to train spatial skills are action games, and those who play action video games more often generally have higher spatial skills [32, 143, 31]. In addition, men and boys tend to enjoy action video games more than women and girls [99, 32, 34, 181], and previously studied games that train spatial skills tend to be overwhelmingly commercial games, which for decades have been designed with men and boys as the target audience [182, 183, 184].

In essence, those with low spatial skill, especially women and girls, are not being served by current approaches to finding game-based spatial skill training interventions. Designing training games with low spatial skill students in mind is therefore essential for addressing this problem and helping more underrepresented students pursue STEM careers. The current work takes a player-centered approach [74, 75] to designing games for this target population by asking directly for their input about what they like in a gaming experience. Combining this information with demographic characteristics allows me to present a *player persona* [185] of sorts to help game designers understand the gaming preferences and demographics of low spatial skill populations.

The study presented in this chapter extends previous work on the relationship between demographics, video game play, and spatial skill in several ways. First, I combine predictors of spatial skill from several different studies: video gameplay habits, genre preferences, gender, age, and SES, in order to build a more complex model of spatial skill predictors than any of them alone and to provide a more specific picture of the low spatial skill population. Second, I include participants from three distinct populations: online college-age adults, students from a non-selective, lower SES status high school, and students from an academically selective, higher SES status high school. This diversity of sampling allows my findings to be more generalizable than studies utilizing only a single population. Third, I analyze a set of predictive variables that has not yet been studied: motivations and emotional gratifications (emotional experiences appreciated during media use [186]) in gaming. These are important aspects of player experience [24, 187, 188, 189, 190] that can help me build a more in-depth model of low spatial skill populations and understand not just what, but why certain genres or patterns of play might appeal to them.

This chapter makes two main contributions to the research literature on spatial skills training: 1) a deeper and more complex understanding of the relationship between pre-existing spatial skill and video game play habits, preferences, and underlying gaming motivations and 2) a set of design recommendations for spatial skill training games that align with the preferences of low spatial skill populations of high school students and college-age adults from diverse backgrounds. This chapter contains a substantial portion of my previously published work [41].

5.1 RESEARCH QUESTIONS

The research study described in this chapter was guided by the following four questions:

RQ1: What video gameplay habits and preferences predict spatial skill independent of gender, age, and population?

RQ2: What motivations for playing video games predict spatial skill independent of gender, age, and population?

RQ3: What emotional gratifications in video games predict spatial skill independent of gender, age, and population?

RQ4: What are the specific gaming habits and preferences of those with the lowest levels of spatial skill?

5.2 METHODS

I conducted an online and in-school study to assess the relationship between spatial skill and gaming preferences. The study consisted of a timed test of spatial skill followed by a series of questionnaires asking about participants' gaming habits and preferences. Thus, the data I collected about habits and preferences was based entirely on self-report measures, which can often differ significantly from their actual behavior [191, 192]. However, even if participants' expressed desire of what is important to them in a game is inaccurate, it is still valuable for the purpose of designing a game to appeal to them because it can still tell me what people may look for first when choosing a new game to play.

5.2.1 Recruitment

I recruited three different populations in the age range 12-22. The first population was a non-selective high school serving primarily low SES students; about 63% of students are eligible to receive free or reduced lunch, or other low income family services. The second population was an academically selective high school serving primarily higher income families (only about 9% of students are eligible for free or reduced lunch). The third population consisted of college-age adults (ages 18-22) recruited from a large public university, a community college, and various online sources. These three populations were selected to obtain a sample in my target age range that was as diverse as possible. I chose to conduct my study with this age range to strike a balance between a younger population with more time to benefit from spatial skill training and my desire to build upon previous findings in the spatial skill literature, which focuses almost exclusively on college-age adults.

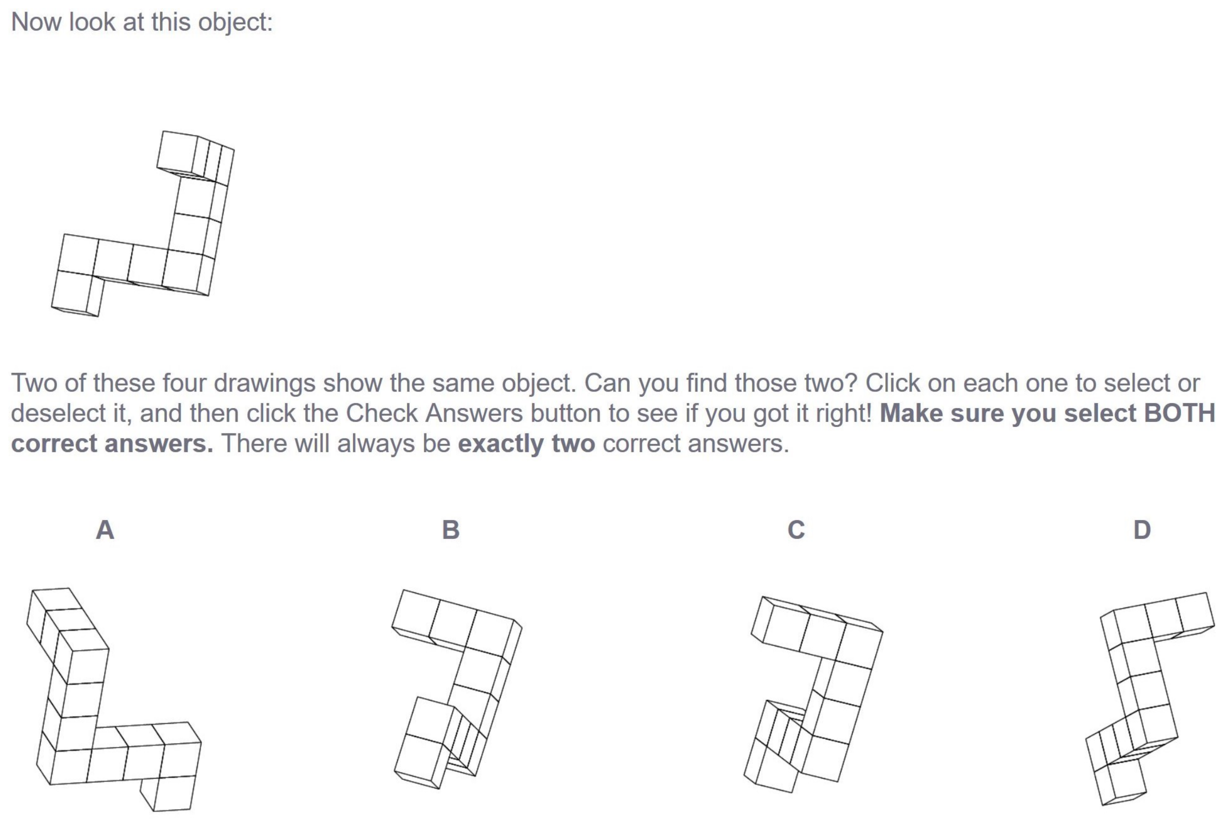
For the college-age population, flyers were posted around the University of Illinois at Urbana-Champaign campus and Parkland College (a community college in Champaign) as well as at libraries and coffee shops in the Urbana-Champaign area. Online advertisements were posted on Facebook, Reddit's r/SampleSize subreddit, and in campus email newsletter, and participants could take the survey online at any time and anywhere that they had internet access. At the two high schools, the survey was incorporated into the school day as a class activity that students could participate in with parental permission and consent forms.

There was no monetary compensation for completing the survey. Instead, I offered a different kind of reward to participants designed to motivate them to take both the test of spatial skill and the questions about their gaming preferences seriously: an opportunity to find out what their primary motivations for gaming were and how well they performed on the spatial test compared to average U.S. adult performance. This form of reward has been used successfully on LabInTheWild.org to attract a large, diverse array of people to participate in online psychology experiments [193].

5.2.2 Survey Procedure

Upon beginning the survey, participants were asked to read and electronically sign either an assent form (for high schoolers) or a consent form (for online participants). The form explained the purpose of the survey and that participants would receive a summary of their performance on the test and their motivations for gaming at the end. Once participants gave their assent or consent, they began a short timed test of spatial skill: an online version of the redrawn Vandenberg and Kuse Mental Rotations Test (MRT-A) [194].

The MRT is one of the most commonly used assessments of spatial skills [102, 32, 3, 5] and has the advantage of being short, making it feasible for an online study. It consists of two blocks of 12 multiple choice questions. Three minutes are given to complete each block, with a break of two minutes in between (my online version also allowed participants to continue to the next block before the two minutes were up if they wanted). For each question, participants must select from the available answers which two represent the exact same object as a given exemplar figure (see Figure 5.1). The MRT includes a set of written instructions and four practice problems with correct answers provided to ensure participants understand the task before they start, which I reproduced in the online version. In between survey administration at the high schools and online deployment, I implemented logging of time spent on the test as a way of checking whether participants took the test seriously.



Now look at this object:

Two of these four drawings show the same object. Can you find those two? Click on each one to select or deselect it, and then click the Check Answers button to see if you got it right! **Make sure you select BOTH correct answers.** There will always be **exactly two** correct answers.

Figure 5.1: A practice question from the online MRT test. An exemplar figure is shown, and test takers must choose which 2 of the 4 drawings below correspond to an identical figure. The correct answers in this case are A and C.

After finishing the test, participants were asked about their gaming habits and preferences. The first survey question asked if the participant had ever played video games. If they had, they were asked a series of follow-up questions related to how recently and how often they played video games (for how many years, hours per week, length of play session), then asked to name their 3 favorite genres and their top 3 favorite games (digital or non-digital). If the participant indicated they had never played video games, they were asked the same questions about games in general, and the questions about recent play, years of play, and favorite video game genres were omitted.

Next, all participants completed the Digital Games Motivation Scale (DGMS), an internationally validated questionnaire used to assess different motivations for playing games [195, 196], and an Emotional Gratifications questionnaire to assess participants' most valued emotional experiences in games (see Table 5.1). Developed by Bartsch [186], it was originally designed for movie and television experiences, but it has been adapted previously by other researchers for video games [197] - I use an adapted version similar to theirs. I included these questionnaires to provide insight into the "why" behind participants' gaming habits and genre preferences. I added an attention check question to both questionnaires, which asked the participant to select a specific answer choice. Those who failed to answer with the requested choice for either questionnaire would be marked as failing the attention check.

Finally, participants completed an optional demographic survey, which asked them to state their gender, age, and country of residence. I also asked participants if they had completed this survey already and provided a text box for them to mention any technical difficulties they had encountered on the spatial skill test or surveys. Once this section was complete, participants saw a summary page describing their performance on the spatial skill test relative to the United States adult average, as well as bar graphs showing their strongest motivations for playing games and their most valued emotional experiences in games, which corresponded to participants' scores on each construct in the DGMS and Emotional Gratifications questionnaires, respectively.

5.2.3 Data Preparation

In total, I gathered data from 506 participants (235 from the selective high school, 63 from the non-selective high school, and 208 from the online survey for adults).

For the public selective and non-selective high school samples, I removed the data of participants who failed attention checks on the DGMS or Emotional Gratifications questionnaire or did not complete the entire survey (selective: n=14, non-selective: n=22) as well as those who reported having technical problems during the spatial skill test that might have affected their performance (selective: n=1, non-selective: 0). This left me with 220 and 41 participants for the selective and non-selective high schools, respectively.

For the online sample, I removed the data of participants who were not between the ages of 18 and 22 (n=73), followed by those who failed to input a valid age (n=23) since the focus of my analysis was on college-age participants only. I also removed the data of those who indicated that this was not their first time taking my survey to avoid duplicate responses (n=9), those who failed attention checks on the questionnaires or did not complete the entire survey (n=8), and those who reported technical problems (n=7). Two additional participants took less than 30 seconds to complete each section of the test (less than 3 seconds per question) and scored 5 out of 24 possible points, below the level of chance. This indicated that these participants did not take the test seriously, and thus I omitted their data as well, leaving me with a final count of 89 online participants.

5.2.4 Data Summary

My cleaned sample consisted of 350 participants. In the selective high school sample, 48% (106) identified as female, 50% (111) as male, and 2% (3) as a different gender, of whom one specified their gender as Genderfluid. In the non-selective high school sample, 41% (17) students identified as female, 56% as male, and the remaining student as Transgender Male. The age range of selective high school students was 12-17, while for the non-selective high school it was 14-22 (6 chose not to answer, only one student reported an age higher than 18). All but two participants in the online sample reported their country of residence as the United States (98%). The remaining two were from Singapore and South Korea, respectively. In addition, 2% of the selective high school students (5) and 7% of the online sample (6) reported never having played video games. All of the non-selective high school students reported playing video games. Cronbach's α for each construct on the DGMS and Emotional Gratifications questionnaire ranged from 0.73-0.91, indicating good reliability.

5.2.5 Grouping Games & Game Genres

The 15 video game genre list used in my study is adapted from some of my prior work outside of this dissertation [198] and includes the following genres: Role-Playing Game (RPG), Action, First Person Shooter (FPS), Strategy, Adventure, Simulation, Music, Fighting, Family, Racing, Fitness, Sports, Platformer, Puzzle, and Other. The Action genre included in my 15 genre list is intended to be a catch-all category for games that people generally consider to be action games but that do not fall into any of the other action-related categories (e.g. arcade games).

However, prior work analyzing the relationship between spatial skills and video game genre preferences generally categorizes games and game genres into a more manageable number of categories. Usually, there are two: "Action" or "Non-Action" [112, 114, 27, 141, 142, 31, 143]. However, there is substantial disagreement about what defines an action game [137]. Therefore, I categorized my participants' favorite genres as "Action" or "Non-Action" first using a more broad set of criteria and then using a more restrictive set.

Looking at the academic research literature on spatial skills and action gameplay, I found that some studies equated first person shooters with action games [3, 142, 143], while others named some exemplar action games [114, 115, 116, 141] with genres corresponding to the following genres in my original 15 item list: Action, Platformer, First Person Shooter, Sports, Simulation, Fighting, and Racing. In the industry sphere, Ernest Adams' *Fundamentals of Game Design* reference book mentions a few subgenres of the Action game genre, two of which correspond to genres in my original list: Fighting and Platformer [144]. A list from a recent LifeWire article includes the subgenres Shooter and Platformer [145]. TvTropes' action subgenres of Platformer, Fighting, and First-Person Shooter [146] are found on my 15 genre list, as are BoardGameGeek's subgenres of Fighting and Platformer [199].

Taking the disjunction of all of these definitions as my broad definition of action games, I ended up with the following list of "Action" genres: Action, Platformer, First-Person Shooter, Sports, Simulation, Fighting, and Racing. However, only a single game in Castel et al's list of "Action" games was tagged as Simulation: *NHL 2002*. Since this game was also tagged as Sports, I decided to remove Simulation from my list, leaving me with Action, Platformer, First-Person Shooter, Sports, Fighting, and Racing. If a participant's list of favorite genres included any genre from my action genre list, I set the variable Action Favorite Genre (Broad) to 1 (true), and if not, 0 (false). For my strict definition of Action genre (Action Favorite Genre (Restrictive)), I took the conjunction of all definitions from the literature, leaving me with the First-Person Shooter (and the Action genre by definition).

Another approach to grouping video game genres is latent class analysis, as in Quaiser-Pohl et al. [32]. Using this approach, I clustered video game-playing participants' favorite genres from my 15 item list with the R package `poLCA` using 2-7 class solutions. Each solution was run 100 times with a maximum number of 5000 iterations. Due to the large number of parameters (participants could select up to 3 favorite genres), I used the Akaike Information Criteria (AIC) (and interpretability) to evaluate each solution. The four class solution provided the best balance between low AIC and interpretability. I interpreted the four classes as "Action Gamers" (favoring Action, FPS, Fighting, and Other), "Cognitive Gamers" (favoring Strategy and Puzzle), "Role-Playing Gamers" (favoring RPG and Simulation), and "Sports/Social Gamers" (favoring Racing, Fitness, Sports, and Family). I assigned each participant their predicted genre class as the variable Favorite Genre Class.

To categorize participants' favorite games as Action or Non-Action, I used genre tags from the review-aggregation site Metacritic. If a game's tags included the word "Action" (e.g., "Action", "Action Adventure", "Action RPG"), I counted the game as Action; if not or if the listed game was not a video game, I counted it as Non-Action. If any of a participant's favorite games was an action game according to this definition, I assigned the participant a value of 1 (true) for the Action Favorite Game variable, and 0 (false) if otherwise. If the game was a video game but not listed on Metacritic (e.g. "Brawl Stars"), or its description was too vague to uniquely identify it (e.g., "Mario", "Telltale Games"), I marked it as "neither" and looked at the rest of the games the participant listed. If no other games were categorized as Action (meaning that this game would be the deciding factor in whether Action Favorite Game was 1 or 0), I omitted the participant's data from the dataset (n=14).

Measure	Scale
Gender	F/M*
Age	Number
Population	NSHS/SHS/Online
Played Videogames Recently	Y/N
How Long Played Videogames	1-5 (< 6 mo.-10+ yrs)
Weekly Hours	Number
Session Duration	1-5(< 15 min.-4+ hrs)
Action Fav. Game	Y/N*
Action Fav. Genre (Broad)	Y/N*
Action Fav. Genre (Restrictive)	Y/N*
Fav. Genre Class	Y/N*
Habit	1-5 (Disagree-Agree)
Moral Self-Reaction	1-5 (Disagree-Agree)
Agency	1-5 (Disagree-Agree)
Narrative	1-5 (Not-Very Important)
Escapism	1-5 (Not-Very Important)
Pastime	1-5 (Not-Very Important)
Performance	1-5 (Not-Very Important)
Social	1-5 (Not-Very Important)
Contemplative Experiences	1-5 (Disagree-Agree)
Fun	1-5 (Disagree-Agree)
Thrill	1-5 (Disagree-Agree)
Character Engagement	1-5 (Disagree-Agree)
Vicarious Release of Emotions	1-5 (Disagree-Agree)
Empathic Sadness	1-5 (Disagree-Agree)
Social Sharing of Emotions	1-5 (Disagree-Agree)

Table 5.1: Hierarchical Regression Measures. *See Grouping Games & Game Genres section. NSHS = non-selective high school, SHS = selective high school.

5.3 RESULTS

My analysis of the data proceeded in two stages. First, to understand what gaming preferences predicted spatial skill, I conducted hierarchical regressions analyzing the relationship between demographics, gaming habits and preferences, and spatial skills. Next, I used the results of the regression analysis to identify the subset of my sample with the lowest spatial skills and characterize their gaming habits and preferences in order to develop a set of recommendations for designers of spatial skill training games.

5.3.1 Hierarchical Regression Analysis

My regression analysis consisted of 3 hierarchical regressions with my entire sample of video game players ($n = 350$). According to the criteria used by Wilson Van Voorhis et al., this sample size provides sufficient statistical power for the number of variables I am analyzing [200]. I chose a hierarchical analysis because I was interested primarily in gaming preferences as predictors of spatial skill after taking into consideration demographic variables' predictive power. For each regression, participants' score on the spatial skill test was the dependent variable. I entered the following demographic variables in the first block: gender (only male and female were used due to the small number of participants (5) identifying as a different gender), age, and population (selective high school, non-selective high school, online). I entered the following gaming experience variables in the second block (See Table 5.1 for details):

1. **Habits Regression**: Played Videogames Recently, How Long Played Videogames, Weekly Hours, Session Duration, Action Favorite Game, Action Favorite Genre (Broad), Action Favorite Genre (Restrictive), and Favorite Genre Class.
2. **Motivations Regression**: Habit, Moral Self-Reaction, Agency, Narrative, Escapism, Pastime, Performance, and Social constructs from the DGMS [196].
3. **Emotional Gratifications Regression**: Contemplative Experiences, Fun, Thrill, Character Engagement, Vicarious Release of Emotions, Empathic Sadness, and Social Sharing of Emotions constructs from the Emotional Gratifications questionnaire [186].

All regressions were performed using the statistical software package R. For each regression described below, diagnostic plots indicated that the assumptions of linearity of the data, normality of residuals, homoscedasticity, and independence of observations were met, and no variables had variance inflation factors greater than 2. All reported β values are unstandardized regression coefficients.

Video Game Players Only

Initially, I analyzed the predictors of spatial skill for only those participants who had played video games before, since those who had never played video games would be unable to indicate their video game genre preferences or answer the Played Videogames Recently and How Long Played Videogames questions. My first regression model, Habits, looked at predictors of spatial skill related to participants' gameplay habits and genre preferences. Before running the model, I noticed that two participants had given extremely high answers for the Weekly Hours question: 90 and 100 hours, so I omitted these participants' data from my Habits model. With only the first block added, the model was significant ($F(4, 259) = 12.51$, $p < 0.0001$) and explained 15% of the variance in the data (adjusted $R^2 = 0.15$). Only male gender ($\beta = 3.43$, $t = 5.57$, $p < 0.0001$) was a significant predictor of spatial skill. In the second block of habits variables, Action Favorite Game (Broad) was the only significant predictor ($\beta = 1.81$, $t = 2.07$, $p = 0.040$), but the second block did not significantly improve the model ($F(9, 250) = 1.24$, $p = 0.27$).

There were a significant number of missing Weekly Hours responses (n=52), especially from the non-selective high school. Including only the participants who submitted answers to the weekly hours question would have excluded a large number of participants' data from the analysis. Therefore, I reran the Habits model again but dropped the Weekly Hours variable. In the new Habits model, the first block was significant ($F(4, 301) = 16.46$, $p < 0.0001$) and explained 17% of the data's variance. Male gender ($\beta = 3.56$, $t = 6.25$, $p < 0.0001$) and being a selective high school student ($\beta = 2.52$, $t = 2.03$, $p = 0.043$) were both predictors of spatial skill. Adding the second block of habits variables did not significantly improve the model ($F(8, 293) = 1.34$, $p = 0.22$).

My second model, Motivations for Playing, also had a significant first block ($F(4, 321) = 18.00$, $p < 0.0001$) explaining 17% of the variance, with male gender ($\beta = 3.52$, $t = 6.58$, $p < 0.0001$) and selective school population ($\beta = 2.45$, $t = 2.075$, $p = 0.039$) as the only predictors of spatial skill. Adding the second block of motivation constructs from the DGMS did not reveal any significant predictors, explained only 1% of additional variance, and did not improve the model significantly ($F(8, 313) = 1.41$, $p = 0.19$).

My third and final model, Emotional Gratifications for Playing, revealed the same pattern of results: significant first block ($F(4, 321) = 18.00$, $p < 0.0001$) explaining 17% of the variance, with male gender ($\beta = 3.53$, $t = 6.58$, $p < 0.0001$) and membership in the public selective school population ($\beta = 2.45$, $t = 2.08$, $p = 0.039$) as the only predictors of spatial skill. Adding the second block of emotion constructs explained only 2% of additional variance and did not improve the model significantly ($F(7, 314) = 1.98$, $p = 0.057$).

In summary, my models showed consistently that among participants who played video games, no gaming habits or preferences (**RQ1**), no motivations for gaming (**RQ2**), and no emotional gratifications (**RQ3**) improved the model's predictive power over and above what the first block of demographic factors provided. Only gender and population were predictive of spatial skill.

Adding in Non-Video Game Players

While participants who reported never playing video games constituted a relatively small portion of my sample (11 participants, 3%), excluding them might bias my data more in favor of action video gamers, since non-video game players by definition do not play action video games. For this reason, I decided to redo the above analyses to include non-video game players. In order to do this, I had to remove the variables Played Videogames Recently and How Long Played Videogames from the Habits model since they were questions about video gameplay habits and thus not applicable to non-video game players. In addition, I assigned a value of 0 (false) to the Action Favorite Game and Action Favorite Genre variables for each non-video game player in the sample.

Rerunning the Habits, Motivations, and Emotions regression models, I found that the Habits model stayed mostly the same. However, the first block (gender, age, and population) explained more of the variance (18%), and the second block of the Motivations for Playing model became a significant improvement over the first block ($F(8, 324) = 2.05$, $p = 0.040$), explaining an additional 2% of the variance. Within the second block, habit was the only DGMS construct associated with spatial skill ($\beta = 0.27$, $t = 2.63$, $p = 0.009$), and the association had little practical importance given its low beta value; increasing average habit score from 1, the minimum possible, to 5, the maximum possible, would only add about one point to the predicted spatial skills test score. The Emotional Gratifications model did not change significantly (gender and population were the only predictors, and the second block did not improve the model). All in all, adding non-video game players to my models did not change them in any significant way.

5.3.2 Low Spatial Skill Population Preferences

Taken together, my regression models suggest that the only predictors of spatial skill across my three study populations were gender and population, with male participants and students at the selective high school scoring higher on the spatial skills test than females and participants from the other two populations. Male gender and being a student at the

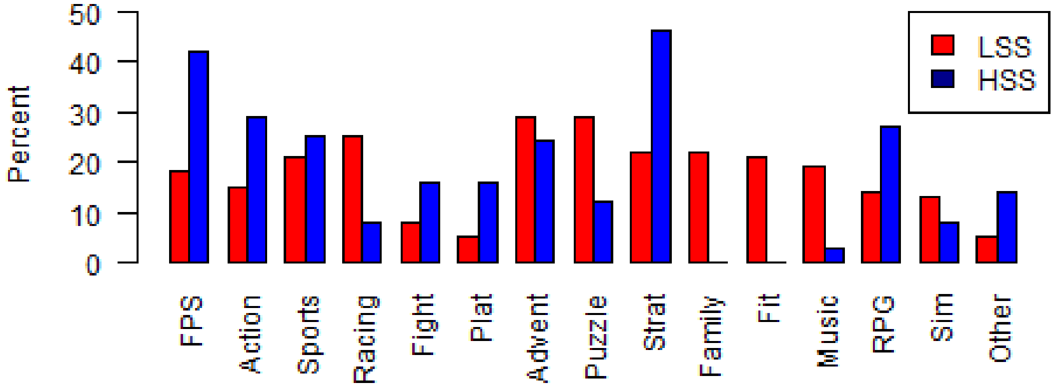


Figure 5.2: Video game genre preferences for the LSS group, as compared to the HSS group. FPS = First Person Shooter, RPG = Role-Playing Game. Percentages do not add up to 100 because participants could choose up to 3 favorite genres.

selective high school each add about 2.5-3.5 points to one's spatial skill test score, meaning that a male selective high school student is predicted to score 6 points higher (out of 24 possible points) than a female participant from the online population or the non-selective high school.

Thus, my data suggests that female participants from the online or non-selective high school populations constitute the lowest spatial skill group in my sample. To provide insight into what this demographic looks for in a game, I analyzed their gaming habits, genre preferences, motivations, and emotional gratifications. While I could have used a threshold value of spatial skill test score to segment my population into low and high spatial skill groups instead, I chose to segment based on the results of my regressions instead because there currently exists no agreed upon threshold to distinguish between "low" and "high" spatial skill people for the mental rotation test I used (the Vandenberg and Kuse MRT), so any threshold I decided on would result in a rather arbitrary segmentation of the data. Henceforth, for the sake of brevity, I will refer to the subset of my participants who are female and come from the online or non-selective high school as the low spatial skills (LSS) group (n=85).

The LSS group reported playing games most often in fairly short sessions; their most popular answer choice was "15-59 minutes" (43%), and about equal numbers of them chose "Less than 15 minutes" (22%) and "1-2 hours" (24%). Reported weekly hours of gameplay tended to be somewhat low in the LSS group. Those in the LSS group who answered the question about weekly gameplay hours (69%) reported a mean of 2.95 hours a week (median = 2, min = 0, max = 20).

Figure 5.2 summarizes the LSS group's favorite video game genres compared to the preferences of the participant group with the highest spatial skill (male selective high school students, HSS). Six LSS participants did not provide favorite video game genres because they reported not playing video games. The two most popular genres in the LSS group were Adventure and Puzzle, which were each chosen by 32% of the group. While 67% of the LSS group's video game players chose favorite genres that fell into my broadly-defined "Action" genre (Action, Platformer, First-Person Shooter, Sports, Fighting, and Racing), Figure 5.2 shows that most of this "Action" preference is due to a preference for the Racing and Sports genres. Genres in my more restrictive "Action" game grouping (including only the First-Person Shooter and Action categories) were much more popular with the HSS group (chosen by 42% and 29%, respectively) than with the LSS group (chosen by 19% and 16%, respectively).

I looked up the Metacritic genre tags for each participant's favorite video games in order to understand participants' genre preferences in more detail. LSS group members listed 101 favorite video games in total, which generated 79 unique tags. Each game had 2-6 tags. Although the Miscellaneous and General tags occurred very frequently, I chose to ignore them as they were not descriptive and always occurred in the presence of more descriptive genre tags. Eight LSS participants (9%) did not list any favorite games.

Analysis of LSS participants' favorite games revealed a pattern not evident in the favorite genre data: while only a little over half of the LSS group (59%) listed at least one video game, Action game preferences were strong among those who did. The four most popular video game genre tags were all highly related to both my more broad and more restrictive definitions of "Action" genre: Action, Action-Adventure, First-Person, and Shooter. Among those in the LSS group who listed at least one video game (n=45), 67% named a game tagged as Action, and the tags Action-Adventure, First-Person, and Shooter were each named by 29% of them. In total, 80% of those who listed at least one video game named at least one favorite game with a tag including the word "Action" (Action, Action-Adventure, and Action-RPG). In addition, only 16% of the LSS group listed both video games and non-video games as favorites, suggesting that there may be two distinct groups of game type preferences among the LSS group: digital and non-digital.

Overall, the LSS group indicated that they were moderately motivated by most of the DGMS constructs. They generally felt somewhat positive about spending time playing games, scoring a median of 3.67 on the Moral Self-Reaction construct. In addition, they were moderately motivated by the desire to perform and achieve (Performance, median = 3.33), a sense of agency (Agency, median = 3), the in-game narrative (Narrative, median = 3), playing just to pass the time (Pastime, median = 3), and the desire to escape from daily life (Escapism, median = 2.67). LSS participants were less motivated by getting to interact with other players (Social, median = 2.33), and playing out of habit (Habit, median = 2), each with a particularly low mode of 1. Thus, most LSS participants have a diverse array of motivations for playing games, but may not be particularly regular or social gamers.

The Emotional Gratifications questionnaire revealed that the LSS group valued mainly pleasurable, hedonistic emotional experiences in games; the only constructs on which they scored higher than 3 (Neither Agree Nor Disagree) were Fun (median=3.75) and Thrill (median=3.5). LSS group members scored a median of 2-3 on the remaining constructs of Character Engagement, Social Sharing of Emotions, Contemplative Experiences, Vicarious Release of Emotions, and Empathic Sadness. While the distribution of Character Engagement scores seemed to be somewhat bimodal with peaks at 1 and 4, suggesting two distinct camps of pro- and anti-character engagement, scores on the remaining constructs were consistently low, indicating that the LSS group did not value these more neutral to negative emotional experiences very much in gameplay. This was especially true of Empathic Sadness, which had a mode of 1.

Taken together, my findings regarding the gaming habits, preferences, and motivations of LSS participants (**RQ4**) suggest that this subgroup is split fairly evenly between a preference for digital games and a preference for non-digital games. Those whose favorite games are video games tend to favor certain subgenres of Action video games as well as the non-Action genres, but the LSS group in general seems to enjoy the Adventure and Puzzle genres the most. Overall, LSS participants prefer short play sessions, have many different motivations for gaming, and value the emotions of fun and thrill the most in gaming experiences, but do not generally play habitually or socially and do not value more negative emotional experiences in games.

5.4 DESIGN RECOMMENDATIONS

Focusing on the low spatial skill (LSS) group in my sample and analyzing their gaming habits, preferences, and underlying motivations for gaming allowed me to obtain a more detailed picture of the kinds of games they might be the most interested in. Here, I summarize this picture and provide several recommendations to designers of game-based spatial skill training interventions based on my findings.

Facilitate Short Gameplay Sessions

Since the LSS group reported playing in short sessions (about 15-59 minutes), reported a low number of hours played per week, and were motivated to play games in part by the desire to just pass the time, I recommend that game interventions focus on providing a gameplay experience that is easy to engage and disengage with to facilitate short sessions. One way to do this might be to make game levels completable in less than 15 minutes each, allowing players to feel a sense of accomplishment and progression despite short play sessions. Mobile games are particularly well-suited to short sessions [201].

Promote Simple Fun and Thrill

LSS participants' responses to the Emotional Gratifications questionnaire indicated that they valued hedonistic emotional experiences in games and put less value in emotional experiences that were more social and cognitive in nature or more negative. Game designers can accommodate these emotional gratifications by designing games to emphasize the more immediate pleasures of gameplay - simple fun, of course, but also thrill. Thrill can be elicited in gameplay by creating tense, suspenseful situations, such as the pressure to overcome a challenge within a certain time limit or complete a mission objective without being discovered and attacked by enemy forces.

Focus on Adventure and Puzzle Genres

The LSS group's gaming preferences seemed to be split along a digital divide: my analysis of their favorite games revealed that a large portion of them seemed to prefer action video games, while many others preferred non-digital games, as evidenced by their responses to the question about favorite games. How can game designers reconcile these two sets of preferences in practice? Here, my findings regarding the LSS group's video game genre preferences may provide insights. Overall, the most popular genres with the LSS group were Adventure and Puzzle, which may indicate some common ground between digital and non-digital gamers' preferences.

Game designers may therefore want to focus on these two genres when designing spatial skill training video games, especially since they lend themselves well to being combined with other genres - like the action games many LSS gamers enjoy. For instance, fast-paced first-person shooter gameplay could be combined with an overarching story, as is done in many Action-Adventure games, and would also support one of the LSS group's stronger gameplay motivations: Narrative. To accommodate non-digital players, Adventure and Puzzle games could be adapted to non-digital formats. Text-based adventures could be designed with spatial features (e.g., having to navigate through buildings or caves and gradually build a mental map of the area as features are described to the player), and many board games exist already that present puzzle-esque spatial challenges, such as the laying out of complex tunnel pathways in the board game *Saboteur* or the spatial planning required for moves in checkers and chess.

Another advantage of focusing on Adventure and Puzzle genres is that they may be easier to combine with spatially-relevant features. While very little is known empirically about which game features are spatially relevant, some preliminary steps in this direction have been taken by me and my colleagues Xiao et al., who found that performance on first person exploration and 3D object construction tasks within a computer game correlated with spatial skill [202, 156]. In addition, Chang et al. and Mazalek et al. found that a first person exploration VR game with tangibles improved players' spatial skills in the short term [152, 153]. Each of these in-game tasks map well to the Adventure and Puzzle genres and demonstrate how a synergy between spatially-relevant features and LSS population preferences might be achieved. However, these studies are preliminary work with underpowered samples, so instead of or in addition to the feature sets they recommend, game designers may want to try incorporating features found in games shown empirically to train spatial skills, such as *Medal of Honor* (Action, First-Person Shooter), *Portal 2* (Action, First-Person Shooter, Puzzle), or *Super Mario 64* (Action, Platformer).

5.5 DISCUSSION

In this study, I investigated four research questions, **RQ1, RQ2, RQ3** and **RQ4,** regarding the gaming habits, preferences, and motivations of low spatial skill teens and young adults in order to provide player-centered design guidance for game-based spatial skill training interventions that aim to increase students' efficacy in STEM majors and careers. My findings are consistent with prior work showing a male and high socioeconomic status advantage in spatial skill [108, 109, 21, 10, 111], but inconsistent with prior work showing a relationship between action gaming and spatial skill [112, 114, 27, 141, 143, 31].

A likely reason for this failure to replicate is that I analyzed only preference for action games, whereas these previous works all measured actual frequency of action gameplay, which may be more relevant to spatial skills. Although I did attempt to measure frequency of gameplay in terms of estimated weekly hours of gameplay, the significant number of participants who left the question on weekly gameplay hours blank prevented me from analyzing action gameplay frequency. Given that a mere preference for action games was not related to spatial skills in this study, it may be that the relationship between spatial skills and action gaming found in prior work is due primarily to the development of spatial skills via action gameplay rather than a preference on the part of those with high spatial skill for action video games.

Another potential reason for my failure to replicate the action gaming and spatial skill relationship may be the difference between the definitions of "Action game" I used and the definitions other researchers have used. My broad definition of Action games based on my literature review (encompassing the First-Person Shooter, Platformer, Sports, Fighting, Racing, and Action categories from my 15 item genre list) was probably broader than what most researchers define as Action, since I took the (inclusive) disjunction of all genres defined by researchers as "Action" to compose my own definition. Given the lack of consensus about what defines an Action game in the research literature, I recommend the development of a formal schema for game classification to facilitate more consistency between different researchers' work and allow different studies to be more easily compared.

My different results may also be due to the fact that nearly all prior work used extreme groups analysis [112, 27, 124, 114, 141, 143], and had very low sample sizes ($n \leq 20$ per comparison group for individual studies). Extreme groups analysis and low sample size can sometimes lead to overestimated effect sizes, a scenario less likely with my larger sample size and regression analysis [203, 204, 205]. I encourage future work in this area also utilizing larger sample sizes and regression analysis instead of separating data into arbitrary groups and increasing the risk of inaccurate results.

In addition, the over-representation of selective (high SES) high school students in my sample relative to the other populations likely caused my results to be less representative of the general population of 14-22 year olds. Since previous work has shown that the gender gap in spatial skills is larger for those with high SES [108], one potential concern is that over-representation of the high SES high school students in my sample may have led to an overestimation of the gender gap in this work. However, this is not likely since the gender gap in performance on the spatial skill pretest was very similar across the three populations I studied. Each population had a mean gender gap in test performance of 3.5-4.4 and the low SES high school students, not the high SES high school students, had the biggest gender gap in performance. Still, over-representation of the high SES selective high school students could have influenced the study's results in other ways. Future work should address this limitation by conducting research studies with a wider range of schools, particularly those in lower income communities who attend nonselective public schools. In order to do this, especially with public schools in lower income communities, it is important that researchers spend the time to establish a relationship with such schools and design research studies in collaboration with them with the twin goals of 1) not interfering with existing required curricula and 2) ensuring that the schools and students are getting something of equal value out of the study to avoid overburdening cash-strapped school districts that are already pressed for time to cover required content.

Another factor potentially affecting my results was that the method for administering the exam was different for the high school populations and the college-age online population. While high schoolers at the low SES high school were given the option to participate in the study instead of attending one of their math classes and participated in the study while I supervised them as a proctor, the online population could do the study whenever and wherever they wanted to without a proctor supervising them, and the students from the high SES high school completed the study as part of the required curriculum. Since there was no monetary reward for completing the survey, those who elected to participate in the online population may have been more intrinsically motivated to take the spatial skills test and survey seriously than the high schoolers, many of whom may have chosen to participate just to get out of class (at the low SES high school) or were required to participate regardless of interest (at the high SES high school). Evidence of this difference in motivation is seen in the fact that a much greater proportion of the low SES high school student population failed attention checks on the surveys than in the other two populations. Overall, differences in levels of motivation to take the study seriously may have adversely affected the accuracy of high school students' responses.

Furthermore, response rates for the low SES high school students were low overall. The study was advertised to several classes at this high school, but in all but one classroom, the majority of students did not return the parental consent form they were asked to get signed to participate in the study. Future work is needed to develop a more effective incentive scheme for motivating students to participate in research studies like mine that involve cognitive testing and long questionnaires and to take the tests and questionnaires seriously. Giving course credit or monetary rewards for completion may increase response rate but does not completely address the problem of motivation; an extrinsic incentive may undermine the motivation of intrinsically motivated students to participate [65] and may cause them to fill in meaningless answers on tests and surveys just to get the reward quickly [206].

Presenting more pre-study material explaining the relevance of the study contents to students' lives or personalizing the post-study summary of each student's performance on the test and answers on the survey more could help improve response rate and quality of responses while avoiding undermining intrinsic motivation. For example, instead of giving students a single number to represent their score on the spatial skills test, they could be given a “spatial cognition profile” based on the kinds of errors they made on the test, and a similar “gaming motivation profile” could be devised to personalize the report of participants’ primary gaming motivations. This, combined with taking more time to talk about the wide variety of tasks and jobs which require spatial skills and what specifically students will get out of the study in more detail before asking for participation could improve response rates and quality of responses in future studies.

Another limitation of my results is that my models only explained 15%-19% of the variance in participants’ spatial skills. Future work could introduce more potentially relevant variables, such as frequency and type of spatial non-video game activities [111, 207], which might help not only predict spatial skill more accurately but provide more design recommendations for spatial skill training interventions targeting populations that do not enjoy digital games, or any kind of game at all.

Given my inconclusive results, I advise game designers to not worry too much about whether or not to use action games for training interventions. It is far more important to incorporate the more specific, fine-grained gaming preferences of low spatial skill populations so that the game intervention is actually something they would want to play. By asking more detailed and fine-grained questions about participants’ gaming habits, preferences, and motivations, I was able to provide more sophisticated, concrete recommendations than would have been possible with more simplistic measures of gameplay habits and genre preferences that have typically been used in past work [124, 112, 114, 27, 141].

Since I found that LSS group members seemed divided between a preference for digital and non-digital games, one might reasonably ask if this group is indeed the best target audience for video game-based spatial skill training interventions. I believe it is; while there was certainly a digital/non-digital split in game preference, the majority (59%) of the LSS group named a video game as one of their favorite games, indicating that a large chunk of the low spatial skill population could be open to a video game-based intervention. However, I am not suggesting that those with a preference for non-digital games should be ignored; rather, I recommend game designers consider how to apply my design recommendations to both digital and non-digital spatial skill training games. Designing for digital and non-digital interventions allows game designers to target a wider section of the low spatial skill population - those who have the most to gain from spatial skill training interventions that can help them achieve the threshold of spatial skill necessary to succeed in STEM majors and future careers [1, 10].

I set out in this chapter to guide game designers in a player-centered approach to spatial skill training game design to improve the STEM efficacy of low spatial skill students, but the approach is applicable to the design of any educational or cognitive training game.

What is exciting about the possibility of using games as interventions is not simply that the intervention will train a skill, but that people will actually want to do the training, just for the intrinsic fun of it. These same skills, after all, can be trained in laboratory or classroom settings using traditional workbook exercises, but this requires extrinsic compensation, whether in the form of money or course credit; as soon as the extrinsic compensation ends, participants are likely to stop training by themselves. The intrinsic fun offered by game training - if the game is designed with the target population's preferences in mind - offers a way for those who stand the most to benefit from training to obtain these benefits relatively painlessly, utilizing their leisure time for *informal learning* [208] rather than replacing the precious few moments of leisure they have with something that feels like work.

Chapter 6: Homeworld Bound: Redux - Revising Homeworld Bound For Improved Learning and Player Experience

This chapter describes the set of revisions made to my spatial skill training game *Homeworld Bound* as a result of the findings of Chapters 4 and 5 with the goal of improving the connection between spatial skill and in-game tasks as well as improving the player experience for the game's target audience: low spatial skill students. With these improvements, the new version of the game, *Homeworld Bound: Redux*, is better positioned to tap into students' spatial skills more consistently and strongly across different levels containing different combinations of spatial features and tasks, and is more likely to be enjoyable and engaging for low spatial skill students. Another major change I made to the game was to change the target audience from elementary school students ages 8-11 to low spatial skill high school students and college-age adults since the latter was the population whose gaming preferences I identified in Chapter 5. Redesigning the game to make it more appropriate for this older age group (mainly to make it more challenging) allows me to implement my design recommendations from Chapter 5 in a game that fits the target audience of those recommendations.

Game-based interventions targeted at children have an advantage in that the intervention has more time to train students' spatial skills up to the threshold necessary to succeed in STEM coursework [1] and may thus prevent student avoidance of STEM electives early on in school. However, an older audience of high school students and college-age adults still has time to benefit from spatial skill training interventions, which could make the difference in terms of what major a student settles on during college, encouraging more students to persist in STEM majors who might otherwise drop out due to the spatial demands of introductory STEM courses. An intervention designed for college-age adults also can be implemented as a required part of introductory STEM courses by professors to improve students' success rate in the course, an approach used successfully by Sorby et al. [128, 129, 131, 209]. Implementation of a game-based spatial skill training intervention directly in a course also allows researchers to study the training effects and player experience at scale (introductory STEM courses in many public universities in the United States tend to be quite large) with the intended target audience of the game: students at risk of giving up on STEM coursework or majors due to their low spatial skill.

To help researchers and game designers understand what game features are most important for a game's effectiveness at training spatial skills and what features are most important for enhancing the player experience of low spatial skill students, I designed *Homeworld Bound: Redux*, the revised version of the original *Homeworld Bound* game described in Chapter 3.

The design revisions I made for *Homeworld Bound: Redux* are grounded in the theoretical framework I developed for mapping between spatial operations and game features (Chapter 3, the empirical findings of my first attempt at implementing game features derived from this mapping (Chapter 4), and my recommendations for designing spatial skill training games for low spatial skill populations [41].

This chapter's principal contributions to the research literature on spatial skill training games are 1) a demonstration of how to implement spatial features derived from my theoretical framework in Chapter 3 and the player experience features I recommended in Chapter 5 in a spatial skill training game for low spatial skill young adults, 2) an open-source, scalable spatial skill training game that can be modified, used for future research studies, and set up in classrooms by the larger community of cognitive training game designers, researchers, and teachers. This chapter contains an expanded, mmore detailed version of my previously published work [42].

6.1 GAME REVISIONS TO ENHANCE SPATIAL FEATURES AND PLAYER EXPERIENCE

Given the lack of relationship between different player performance and spatial skill in several levels I found in my study investigating the relationship between spatial skill and performance in the original *Homeworld Bound* (Chapter 3), there were several opportunities for improving the extent to which each level tapped into players' spatial skills via *object rotation*, *object alignment*, *landmark orientation*, and *navigation visualization.* In addition to enhancing the spatial features of *Homeworld Bound*, I modified the game to be more in line with the recommendations I made in Chapter 5 for designing spatial skill training games that are appealing to the population that stands the most to benefit from them: those with low spatial skill [41]. These player experience (PX) recommendations focused on 1) *facilitating short gameplay sessions*, 2) *promoting simple fun and thrill*, and 3) *focusing on the Adventure and Puzzle genres.*

6.1.1 More Complex Spatial Features

The first substantial revisions to the original *Homeworld Bound* involved making the game's spatial features more complex and thus more likely to require players to tap into their spatial skills during gameplay in both Construction and Exploration Mode.

Figure 6.1: Screenshots of Exploration Mode (left) and Construction Mode levels in *Homeworld Bound: Redux*. Shown here are levels with time limits: Rocket Boots part collection (upper left) and Sledgehammer part collection (bottom left) in Exploration Mode and Rocket Boots construction (upper right) and Sledgehammer construction (bottom right) in Construction Mode. Only levels pertaining to the collection of parts for and construction of more complex items like the Rocket Boots and Sledgehammer have a time limit.

Construction Mode

In Construction Mode, I focused on increasing the complexity of part shapes for the Rocket Boots and Sledgehammer levels (see Figure 6.1 for screenshots of the new versions). Due to time constraints, the Ruined City Key level was not implemented in *Homeworld Bound: Redux*. In the Rocket Boots level, I replaced some of the cuboid parts with trapezoidal prisms, increased the number of parts, and gave parts with similar shapes different main colors. I introduced trapezoidal prisms to the Rocket Boots level since they were already present in the original Sledgehammer level and I hypothesized that the difficulty of predicting the result of rotating this more complex part shape (*object rotation*) was the reason for the relationship between spatial skill and performance on the Sledgehammer level that I found in Chapter 4.

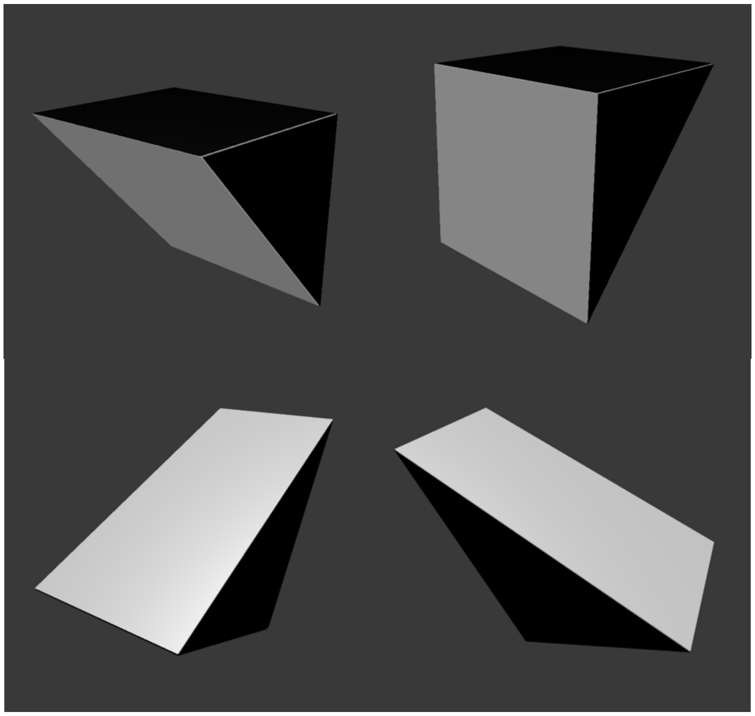


Figure 6.2: The two chiral pairs of parts (first pair on top, second pair on bottom) in the Sledgehammer Construction Mode level in *Homeworld Bound: Redux*. Each member of a pair is identical except for the location of the black fusing area on its side; on one part the fusing area is on its left side, and on the other, its right side. These fusing areas must be attached together during construction, and thus each pair is chiral; neither of the parts can be rotated to look identical to the other. And since each of these parts has two fusing areas (one on top, and one on the side), figuring out the rotation and alignment necessary to successfully fuse them to the construction requires a more complex *object alignment* procedure similar to that required in the original Rocket Boots level (see Chapter 4).

I increased the number of parts used and gave parts with similar shapes different colors in the Rocket Boots level to make it more difficult to complete this level without reference to the 2D image of the finished construction in the upper left corner given my hypothesis in Chapter 4 that players might be using a brute force strategy for figuring out which parts to attach in this level instead of using the intended spatial strategy of establishing correspondences between the 2D representations of parts in the image and their 3D counterparts. Establishing these correspondences would require players to mentally rotate either the parts shown in the 2D image or the 3D parts on their screen and thus tap into their *intrinsic-dynamic* spatial skills more deeply to complete the level.

In the Sledgehammer level, I increased the number and complexity of parts by splitting some of them in half. This produced chiral pairs: pairs of parts that looked alike except that they were mirror images of each other and could not be rotated to look identical to one another. I introduced these chiral pairs with the goal of tapping more deeply into players' *intrinsic-dynamic* spatial skills since distinguishing each part within a chiral pair from the other involves the *intrinsic-dynamic* spatial task of mentally rotating one of the parts to see that it cannot be superposed upon the other. This task is central to many of the most widely used tests of *intrinsic-dynamic* spatial skill assessment, such as those used by Shepard and Metzler [157] and Vandenburg and Kuse [210] in their seminal work. In addition, these chiral pairs each require a more complex *object alignment* procedure involving multiple fusable areas that I hypothesized was related to the link between spatial skill and performance in the original Rocket Boots level in Chapter 4 (see Figure 6.2 for an example of the alignment procedure for a chiral pair).

A final revision I made to Construction Mode levels to increase the difficulty of the *intrinsic-dynamic* mental rotation tasks they required was to remove the automatic movement of parts close to selected fuse areas once two fuse areas were selected. While this feature aided children's understanding of what it meant for the fusable areas of the construction and the selected part to be aligned and thus aided them in figuring out the correct rotation, I felt that this aid might make the *object rotation* and *object alignment* tasks required to fuse a part too easy for the adult target audience of *Homeworld Bound: Redux*.

Exploration Mode

In Exploration Mode, I made two main changes to increase the complexity of the navigation task in the Canyon level. I focused my revisions on the Canyon level since I found no relationship between completion time in the Canyon level and spatial skill in my correlation study with *Homeworld Bound* (Chapter 4). My goal with the revisions was to make the level's navigation more strategic and challenging to tap into players' spatial skills better.

First, I worked with one of my undergraduate collaborators to redesign the Canyon level so that there were multiple ways to reach each area of the level, some shorter than others. We implemented this redesign by adding tunnels between certain areas of the level that provided a shorter route than the original path between the two areas the player would have had to take. The goal of adding these alternative paths was to make it more difficult for players to find parts by engaging in the kind of nonstrategic wandering I observed in earlier playtests that reduced the need to employ navigational spatial skills. With the increased number of possibilities and forking paths in the Canyon level, players might find wandering more

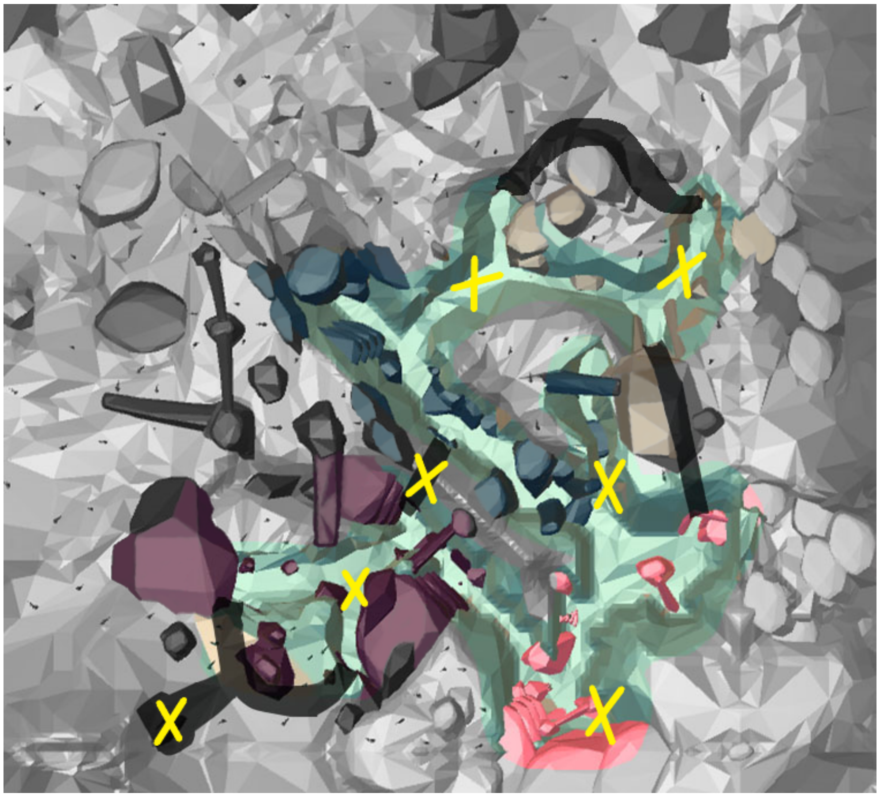


Figure 6.3: Bird's eye view map of the revised Canyon level. The distinct regions of rock formation colors can be seen here (purple, blue, pink, and gray) as well as the new tunnels added to create alternative paths through the level (overlaid in black). Yellow Xs indicate the location of Rocket Boots parts in the timed level.

tedious and be motivated to use *landmark orientation* to find parts quicker by recognizing and avoiding paths they had already traversed.

The second change I made to the Canyon level was to make landmarks more distinctive through the use of distinct color variations to further encourage players to use *landmark orientation* to navigate through the level. As Figure 6.3 shows, in *Homeworld Bound: Redux*, the Canyon level is divided into four sections by the color of the rock formations in each: blue, gray, purple, and pink. This variety of colors was chosen to make the color differences noticeable while still fitting in with the level's overall color scheme of turquoise.

6.1.2 Time Limits to Increase Difficulty and Suspense

To further discourage the nonstrategic wandering in Exploration Mode I observed during playtests with the original *Homeworld Bound* and motivate players to engage in strategies utilizing *landmark orientation* and *navigation visualization* instead, I added a time limit and a static map showing the location of each item part to certain Exploration Mode levels (See Figure 6.1). Allowing players to orient themselves by determining the correspondences between the bird's eye view of the landmarks on the map and the landmarks visible from their own egocentric perspective in the environment (such as tunnel entrances and colorful rock formations, see Figure 6.3) may encourage players to use a *landmark orientation* strategy for navigating and finding item parts rather than wandering around until they stumble upon something. In addition, if players do not plan out a route on the map ahead of time, it is very difficult to find all the item parts shown on the map within the time limit, so the combination of map and time limit may also prompt players to engage in *navigation visualization* at the start of the level. Including the map and time limit features in only certain Exploration Mode levels allows researchers to use *Homeworld Bound: Redux* to evaluate the added value of these new features for tapping into players' spatial skills in comparison to levels lacking these features.

I also added a time limit to the more complex Construction Mode levels (Rocket Boots and Sledgehammer) to increase the difficulty of the spatial operations players were required to perform in these levels: the *object rotation* and *object alignment* operations that tap into players' *intrinsic-dynamic* spatial skill. This same time pressure is present in levels of *Tetris*, a game successful at improving players' mental rotation (an *intrinsic-dynamic* spatial skill) and in standardized tests of spatial skill. Furthermore, the fact that prior work has shown that gender differences in scores on standardized tests of spatial skill are often more pronounced when time limits are more stringent suggests that not only accuracy but also speed are necessary to achieve higher levels of spatial skill. Thus, including a time limit in only certain Construction Mode levels allows researchers to evaluate the extent to which time pressure by itself might be an important feature to include in spatial training tasks to motivate players to perform accurate spatial operations more quickly and thus tax their spatial skills more fully.

To increase the hedonistic feeling of thrill during gameplay (**PX Recommendation #2**), I combined three new game elements in timed Construction Mode and Exploration Mode levels: the time limit mentioned above, a narrative explanation for the time limit, and suspenseful music. Before each timed Exploration Mode level started, I added a narrative explanation for the timer ("there is an evil power here that can sense the energy output of the fully charged Fuser. Use the Fuser to find the locations of hidden building materials, but work quickly so that you aren't detected"). In both Exploration and Construction Mode levels, I worked together with one of my undergraduate collaborators to change the in-game music to a more ominous and urgent version when there was less than 1 minute left on the timer (both the regular and ominous versions of the music for timed levels were original compositions by my collaborator). The narrative explanation for the time limit and sense of urgency also served the purpose of increasing the presence of narrative in the game, one of low spatial skill players' stronger gaming motivations and a key component of one of their favorite video game genres, Adventure (**PX Recommendation #3**).

6.1.3 A New Purpose for Batteries

Given the new adult target audience, Construction Mode levels in *Homeworld Bound: Redux* needed to be more difficult so players could not brute force them by accumulating a very large number of batteries and then trying every possible rotation at every possible fusing area, thus avoiding the use of spatial skills to complete the level. Previous research analyzing strategies used by *Tetris* players supports this notion; Kirsh and Maglio found that *Tetris* players could avoid using the *intrinsic-dynamic* spatial skill mental rotation by simply performing a quick succession of actual rotations on *Tetris* blocks as they fell to more easily see all the possibilities (referred to as *epistemic* actions by Kirsh and Maglio) [211].

To prevent brute force methods, the number of rotations allowed in each Construction Mode level in *Homeworld Bound: Redux* is not determined by how many batteries the player collects. Instead, rotations are limited to a fixed amount that is just slightly more than the minimum number of rotations required for the particular level (allowing for a couple mistakes or less efficient multi-step rotations). When players use up all their rotations in a Construction Mode level, the construction breaks apart and they must restart the level from the beginning instead of getting to collect more batteries and resuming the level where they left off. Thus, players must plan out and visualize the rotations they will perform even more carefully in *Homeworld Bound: Redux*, requiring them to think more about *object rotation* and *object alignment* to avoid wasting rotations and having to restart the level.

The decoupling of batteries with rotations in Construction Mode rendered batteries in their current form useless. However, instead of removing batteries from the game altogether, I changed the way that batteries were integrated into gameplay. Instead of being an infinite resource the player can collect to get more and more rotation operations in Construction Mode levels, batteries must be collected in pieces and built in Construction Mode themselves. Rather than providing "power" for rotation operations in Construction Mode, batteries now allowed players to charge up the Fuser to full power, which is required to build more complex items like the Rocket Boots and Sledgehammer. I added this narrative explanation for batteries due to the moderate importance my new target audience of low spatial skill young adults placed in in-game narratives and its important role in one of their preferred video game genres, Adventure (**PX Recommendation #3**).

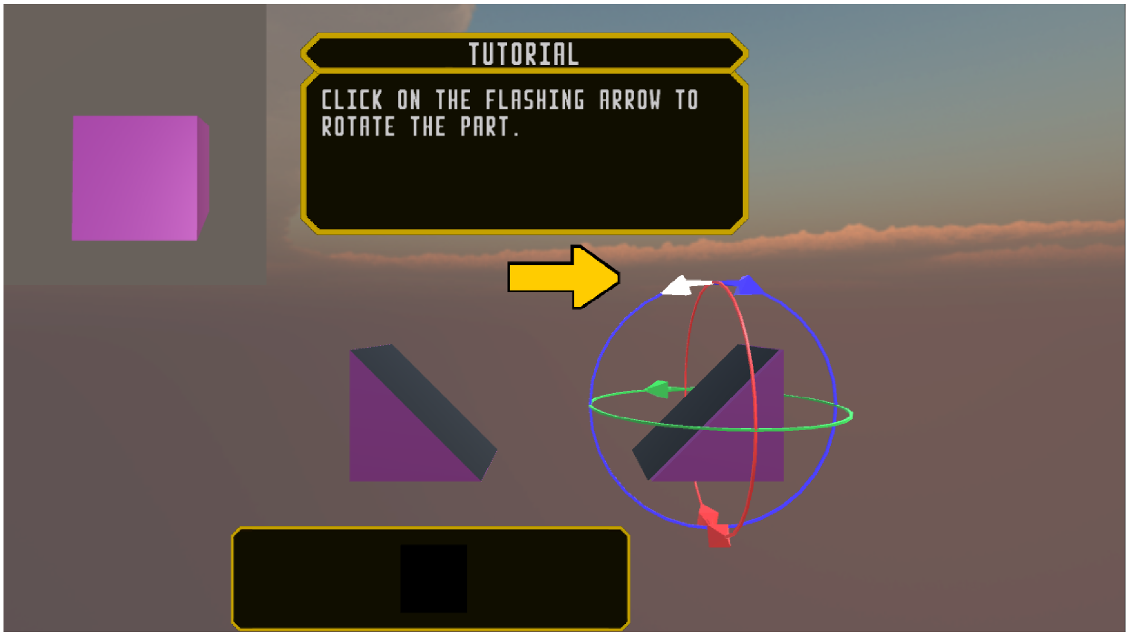
This design decision resulted in 8 new Construction Mode levels, one for each battery. Adding these additional 8 Construction Mode levels to *Homeworld Bound: Redux* served two main purposes. First, it helped balance the time spent in the game between Exploration and Construction mode. In the original game, *Homeworld Bound*, players spent about 80% of their time in Exploration Mode, meaning that there was much less time spent on the part of the game designed to train *intrinsic-dynamic* spatial skills using *object rotation* and *object alignment* tasks. With more Construction Mode levels, more time could be devoted to training *intrinsic-dynamic* spatial skills and more combinations of game features could be tested for their effectiveness at tapping into players' spatial skills.

Another benefit of adding more Construction Mode levels was that it broke up the otherwise lengthy Exploration Mode levels (Canyon and Highlands) into shorter sublevels that could each be completed in a few minutes. The Canyon and Highlands levels each got partitioned into 5 shorter levels: one leading up to each of the 4 new Construction Mode levels and the final one leading up to the original Rocket Boots level (Canyon) or the original Sledgehammer level (Highlands). Splitting up Exploration Mode levels in this way allowed for shorter Exploration Mode level completion times, which in turn made the game more suitable for shorter gameplay sessions (**PX Recommendation #1**). Longer levels might force players to play longer than they wanted to or were able to since stopping gameplay in the middle of a level would break the player's flow and require more time to remember what they were in the middle of upon returning to the game.

6.1.4 Improved Scaffolding for Construction Mode

Since I had repeatedly observed that players struggled to understand how gameplay in Construction Mode worked, both with the children in my previous pilot studies and adult testers, I revised the tutorial section of Construction Mode to improve the learning experience. The final tutorial design I settled on for the original version of the game was minimalist since I found that children ignored the text in a more structured competence-promoting tutorial, and I wanted to give them more autonomy in gameplay. However, one of my undergraduate colleagues helped me realize that in order to provide appropriate scaffolding, I needed to strike a balance between the competence and autonomy-promoting aspects of my previous attempts at the tutorial. The final version of the tutorial, developed by my both of us, achieved this balance by *scaffolding* the learning experience.

Scaffolding is a method rooted in the learning sciences literature and has been shown to be empirically effective at enhancing learning [67]. To scaffold a learning experience is to provide a high level of initial support, guidance, and feedback on the learner's actions,


TUTORIAL

CLICK ON THE FLASHING ARROW TO ROTATE THE PART.

Figure 6.4: Screenshot of the new tutorial level for Construction Mode in *Homeworld Bound: Redux*. The player is given a restricted set of controls and simple step-by-step instructions for the entire process of selecting a part, choosing where to fuse it to, rotating it, and fusing it. Each step of the tutorial must be followed in order to progress to the next step.

and then gradually remove this support as the learner progresses [68]. The scaffold helps the player progress initially, increasing their sense of competence, while its gradual removal increases their sense of autonomy. In turn, these feelings of autonomy and competence increase the player's intrinsic motivation to play [63, 61, 25].

The new, scaffolded tutorial for Construction Mode in *Homeworld Bound: Redux* consists of one simple, structured level that introduces each basic operation required in Construction Mode one by one, and introduces supplemental controls later at the beginning of the first battery level. At the beginning of the tutorial level, the player is walked through a simplified version of the process of part selection, fusing area selection, rotation, and attachment. At each step of the process, the player is told what to do and is not allowed to perform a wrong action. In addition, the camera is kept fixed at a zoom level and angle that allows the player to see everything necessary to complete one part attachment and no more. This restriction also serves the purpose of exposing the player more gradually to the controls and features of Construction Mode so they are not overwhelmed. Explanatory text is provided at each step to help the player understand how the process works (see Figure 6.4).

Once the tutorial level is complete, the player is introduced to the camera zoom and camera rotation controls at the beginning of the first battery level and is required to try each out once before the rest of the level's controls are enabled. Once the level starts, the player encounters wrong rotation and wrong face errors for the first time with explanations about what part of the fusing process is incorrect (either "INCORRECT ALIGNMENT" or "INCOMPATIBLE LOCATION", respectively).

I implemented this new scaffolded tutorial to improve the player experience of learning Construction Mode gameplay. In the original *Homeworld Bound*, the tutorial levels were often progression stoppers for players such that if proctors had not been present at studies using the game, players would not have been able to progress past the tutorial level and may have quit the study in frustration. The improved tutorial in *Homeworld Bound: Redux* is designed to allow players to play the game as intended - without recourse to external help. Eliminating the need for proctors to explain Construction Mode's tutorial means the game can be more easily used in large classrooms and outside of the classroom as well - where most commercial games are played anyway.

6.1.5 Enabling Online, Remote Play

The last major game revision I made was to make *Homworld Bound: Redux* playable online. The original *Homeworld Bound* was playable only via an executable file on a local Windows or Mac computer. For a study involving n computers, the file had to be copied onto each of the n different computers, and two different executable files were needed for Windows and Mac computers. Furthermore, all player data the game recorded was dumped into a directory on each local machine, meaning that to get all of it for data analysis, I had to copy each player data file off every one of the n computers used for the study. All this is to say that the requirement to run a local executable to play prevented studies using the game from scaling well. Running the game locally also made it extremely difficult to run remote studies and collect data from them, a threat to the ecological validity of future studies aiming to analyze the player experience of the game since laboratory or classroom studies with proctors present are not the natural environment in which a person typically plays games for leisure.

To address these limitations, I developed *Homeworld Bound: Redux* as an online game via Unity's WebGL build option. Once compiled to WebAssembly, a Unity game can be embedded and run in a webpage, and telemetry data from the game can be output to a central database for rapid and scalable data collection and analysis. When a study is conducted using *Homeworld Bound: Redux*, participants need only be directed to the webpage hosting the game and then instructed on how to quit the game so that their telemetry data can be sent to the central database at the end of each study session. Thus, conducting online, remote, and anonymous studies with *Homeworld Bound: Redux* is straightforward and enables study participants to play the game in a more naturalistic play context than a laboratory or a classroom.

6.2 DISCUSSION

This chapter described the revisions made to *Homeworld Bound*, my game for training children's spatial skills, to improve its ability to tap into players' spatial skill, improve the player experience for its new target audience of low spatial skill young adult players, make the game suitably challenging for low spatial skill young adults, and enable the game to be played online. Revisions to the game's spatial features were made by analyzing empirical data from the correlation study I describe in Chapter 4 in combination with my theoretical framework for mapping game features to spatial skills I describe in Chapter 3 to identify specific game features in certain levels that may have allowed players to circumvent the use of spatial skills during gameplay and then adding incentives for players to adopt spatial strategies instead. Revisions to improve the player experience of the game were made based on the recommendations of my player-centric analysis of what low spatial skill young adults from diverse backgrounds look for in a game as described in Chapter 5. Additional revisions to make the game more challenging for adult players and playable online were implemented over the course of several iterations of informal playtesting. These four major categories of revisions resulted in *Homeworld Bound: Redux*, a spatial skill training game for low spatial skill young adults that can be easily implemented and scaled in classrooms, laboratory studies, and any remote location where players have access to a computer and an internet connection.

The current WebGL build of *Homeworld Bound: Redux* as well as its open-source code is available for download at `https://github.com/hwauck/homeworld-bound` for researchers, game designers, and educators to use and modify for future studies evaluating the game's training effectiveness. It is my hope that sharing the code for *Homeworld Bound: Redux* will inspire other researchers to continue to build on the game and make further improvements to its spatial features and player experience features.

Of course, the question of whether *Homeworld Bound: Redux* actually can train players' spatial skills and actually appeals to its target audience remains. In the next chapter, I give a partial answer to these questions using a controlled training study deployed in the context of an introductory STEM course.

Chapter 7: Evaluating Homeworld Bound: Redux with Low Spatial Skill Students

The preceding chapter of this dissertation described the design and revision process for creating *Homeworld Bound: Redux*, the latest version of my game for training spatial skills. The game was revised from its original version, *Homeworld Bound*, to do a better job of tapping into players' spatial skills and improve the player experience for its new target audience of low spatial skill young adults. The next logical step is, of course, to evaluate the extent to which the game's revisions actually accomplish these goals, using the data-driven and player-centric approach I have introduced in the preceding chapters. In this chapter, I discuss the results of a study I conducted in an introductory programming course for non-STEM majors at the University of Illinois Urbana-Champaign to evaluate *Homeworld Bound: Redux* in terms of both training effectiveness and motivational appeal for low spatial skill college students.

My evaluation focuses on low spatial skill college students because they have the most to gain from spatial skill training interventions. Low spatial skill serves as a barrier to success in introductory STEM coursework and majors [Uttal and Cohen 2012], and may be exacerbating the gender gap in STEM since women consistently make up a disproportionate number of those with low spatial skill [21, 76, 8]. Thus, low spatial skill students represent a critical target audience for spatial skill training interventions aiming to help more students succeed in STEM coursework, persist in STEM majors, and pursue STEM careers.

The evaluation studies I conducted for *Homeworld Bound* in Chapters 3 and 4 suffered from the limitation of lower sample sizes, hindering their generalizability and reliability. This limitation is also shared by other studies evaluating the training effects of non-commercial spatial skill training games [152, 153]. In addition, game-based spatial skill training studies in general do not investigate training effects on low spatial skill populations specifically, limiting our ability to know if a training intervention will be effective for this critical target audience. In this chapter, I contribute the first study evaluating the spatial skill training effectiveness of a non-commercial game that is 1) conducted in a player-centric way, using its target audience of low spatial skill students as participants and 2) conducted with a sample size large enough to observe reliable training effects.

However, it is not enough to study the training effects of a game-based training intervention with a low spatial skill population; for a truly player-centric approach to game evaluation, the motivation of the target audience to play must be taken into account as well. Low spatial skill students may not enjoy the kinds of games that are most frequently studied in the spatial skill training literature [32, 34, 41], which could remove the fun from the game experience and thus remove one of the key advantages games have over other training interventions. In addition, students with low spatial skill may find a spatial skill training game too difficult if it is not designed with low spatial skill students in mind, potentially leading them to give up on the intervention when implemented as a course assignment. The study I describe in this chapter contributes the first investigation in the research literature of low spatial skill students' motivation to complete a game-based intervention relative to other spatial training interventions.

Demographic factors like prior gaming experience and gender may also affect motivation to engage with game-based spatial interventions since action video game players tend to have higher spatial skill than those who do not play action games [124, 112, 115] and men tend to play more action games than women [118, 119, 34, 120]. These demographic factors may lead to differential training effects as well [3, 126, 4, 30]. It is therefore critical to ensure that a game-based spatial skill training intervention is effective and motivating for demographic groups that may be less interested in playing existing commercial action games like *Unreal Tournament* and *Medal of Honor* that have been shown to be effective for a general population. Therefore, in this chapter I extend my player-centric analysis of training effects and enjoyment to not just low spatial skill students, but also to demographic groupings by gender and prior gaming experience to evaluate whether the intervention's benefits extend only to certain demographics. This chapter thus contributes a deeper understanding of how different demographics may benefit differently from game-based spatial skill training interventions to the research literature.

A particular advantage to using a non-commercial, modular game like *Homeworld Bound: Redux* as a spatial skill training intervention is that features hypothesized to be relevant to a game's training effectiveness can be manipulated and isolated across different levels to analyze their empirical ability to tap into players' spatial skills. This data-driven analysis, in turn, can be used to identify levels of the game that are doing a good job of tapping into players' spatial skills and diagnose levels that are not tapping sufficiently into players' spatial skills. Comparing the different features and difficulty across these levels can provide insights into how the less effective levels might be improved to tax players' spatial skills more and increase the game's training effectiveness. The study I describe in this chapter contributes the most detailed analysis to date of the effectiveness of multiple types of game features at tapping in to players' spatial skills, providing researchers and game designers with insights that can be used to increase a spatial skill training game's likelihood of actually achieving training effects and deepen our theoretical understanding of the mechanisms behind spatial skill use in digital games.

7.1 RESEARCH QUESTIONS

RQ1: How does the spatial skill training effectiveness of *Homeworld Bound: Redux* compare with the training effectiveness of online spatial exercises and a graphics programming activity for undergraduate non-STEM majors?

RQ2: How motivated are undergraduate non-STEM majors to play *Homeworld Bound: Redux* compared to completing traditional spatial workbook exercises or a graphics programming activity?

RQ3: What is the relationship between spatial skill improvement and demographic characteristics such as current spatial skill level, prior gaming experience, and gender?

RQ4: What is the relationship between enjoyment of different spatial skill training interventions and demographic characteristics such as current spatial skill level, prior gaming experience, and gender?

RQ5: What is the relationship between spatial skill and performance in different levels of *Homeworld Bound: Redux*?

7.2 METHODS

To answer my research questions, I conducted a training study in the context of an introductory computer science course for non-STEM majors. I decided to conduct the study with this class because its students were likely to have lower spatial skill than average given that spatial skill tends to be higher among engineering and other STEM majors [2]. Another advantage to conducting the study in this course was the large class size, which enabled me to be more confident in my estimates of effect size and do follow-up subgroup and correlation analyses. I launched the study at the very beginning of the course in August 2019. All study activities were incorporated into the class as assignments by the instructor and were graded for completion only.

Students took an online spatial skill pretest in the first week of class. The pretest was an online version of the Revised Visualization of Rotations Test (PSVT:R) (Figure 7.1) [166], one of the most commonly used tests of spatial skill in engineering education. The online version was created by some of my collaborators for a different project and automatically randomizes the order of questions and answer choices each time the test is taken to reduce test-retest effects. The online platform logs the time spent on the test as a way of checking whether students took the test seriously. If a student failed to answer all 30 questions on the test within the specified time limit of 20 minutes, they were redirected to a page telling them they were finished, and their answers for the remaining questions were marked as blank. Students were also asked to report any technical difficulties they encountered while taking the test to the instructor or researchers.

One week after taking the test, students participated in a training intervention in their lab section. Since there were 15 lab sections, I assigned 5 sections to each of the following three training conditions: *Homeworld Bound: Redux*, Spatial Exercises, and Python Graphics. For more details about each training condition, see the following subsection. When choosing which sections to assign to which condition, I controlled for time of day and day of week (ensuring sections at different times of day and days of the week were as evenly distributed as possible over the three conditions), and after this did my best to ensure that conditions were evenly distributed over different teaching assistants (some teaching assistants taught multiple sections). I did not randomly assign individual students to conditions since seeing other students in the same section doing a different intervention might affect students performance and motivation on theirs.

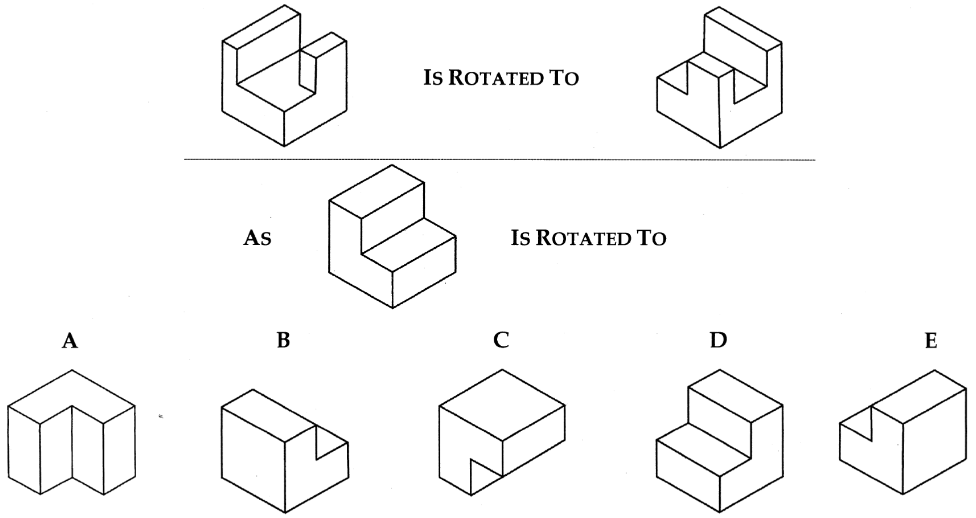


Figure 7.1: A sample question from the Revised PSVT:R test of mental rotation. Participants are shown an exemplar figure before and after being rotated a certain way and must decide which of the answer choices results from performing the same rotation on the second exemplar figure. In this case, the correct answer is D.

At the beginning of the lab period, students were informed that since spatial skills are strong predictors of success in computer science courses, they would be completing a spatial activity. Students then worked on their assigned intervention for 70 minutes of the 80 minute lab period, or until they finished the activity. Once the 70 minutes were up or students had finished the activity, they were directed to complete the enjoyment subscale

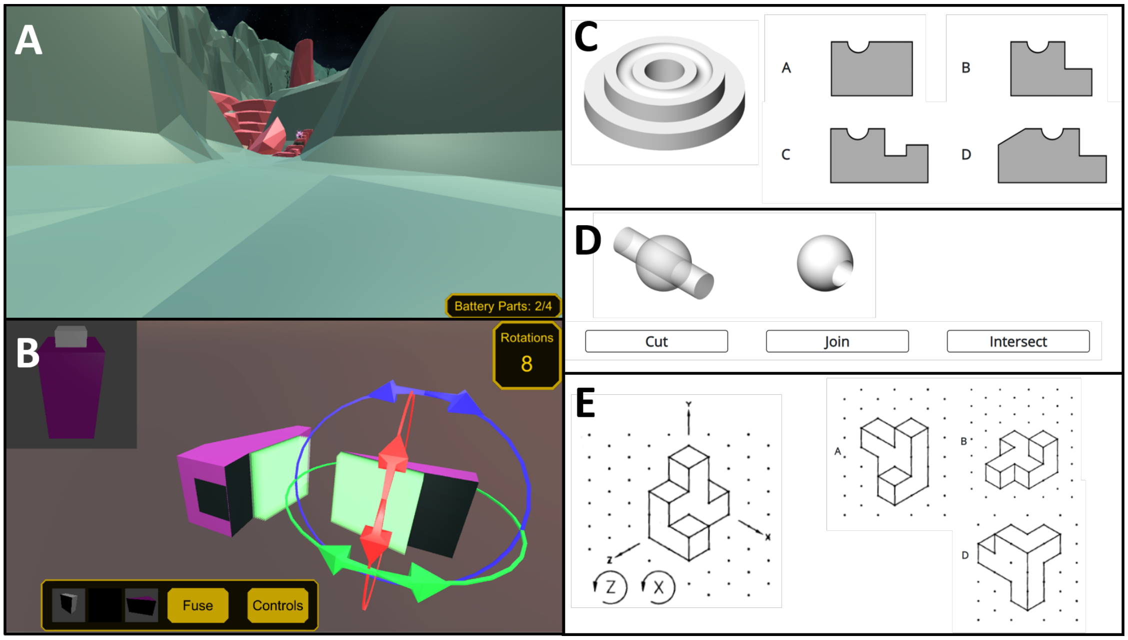


Figure 7.2: Sample screens from the two spatial training interventions. *Left:* screenshots of *Homeworld Bound: Redux's* Exploration Mode **(A)** and Construction Mode **(B)**. *Right:* Examples of three question types used in the Spatial Exercises intervention: Choosing the 2D object that produced the shown 3D solid of revolution **(C)**, determining which operation has been performed on the two intersecting objects shown on the left to produce the object on the right **(D)**, and selecting which of the answer views on the right is produced by performing the two indicated 90 degree axis rotations on the figure on the left **(E)**.

of the Intrinsic Motivation Inventory (IMI), a well-validated measure of intrinsic motivation for a task [212, 213, 214] consisting of 7 7-point Likert scale questions. I asked students to rate enjoyment in order to compare not only training effects, but also students' motivation to do the activity, across conditions. If students finished their intervention before the class ended, teaching assistants checked the assignment for completion and reasonable effort and then allowed students to leave once they had completed the IMI.

A week after completing the intervention, students were asked to take an online post-test and demographic questionnaire. The post-test was the same as that used for the pretest. After completing the post-test, students read an informed consent form and were given the option to allow the use of their data for this study. Lastly, students completed the questionnaire, which included questions about gender, growth mindset, what kinds of games (digital or otherwise) they played most often, and how much time they spent weekly playing video games. Asking these questions allowed me to perform more detailed subgroup analyses of training effects.

7.2.1 Training Interventions

Homeworld Bound: Redux

Homeworld Bound: Redux is a first person, online computer game designed to train spatial skills. The design of the game is based on a combination of two sets of game features, as described in Chapter 6. Game features relevant to spatial skills were proposed using a mapping between different spatial subskills in the spatial skill typology proposed by Chatterjee [45] and adopted by Uttal et al. in their meta-analysis of the spatial skill training literature [10] and features drawn from other video games that have been empirically demonstrated to train players' spatial skills [4, 112, 3, 5]. These features were then incorporated as the game's core gameplay mechanics: *object rotation* (rotating 2D or 3D in-game objects relative to the player), *object alignment* (aligning matching surfaces on objects to fit them together), *landmark orientation* (orienting oneself in first person by using nearby visual landmarks), and *navigation visualization* (visualizing one's path to get from one point to another in an environment before beginning the movement) [40]. In Chapter 5, I proposed another set of game features appealing to players with low levels of spatial skill [41] and later incorporated them into *Homeworld Bound: Redux* as described in Chapter 6 to improve the game's appeal to this subset of the population that stands the most to gain from spatial skill training interventions [42].

The game's premise is that the player is stranded after a crash on an alien planet and must explore and salvage scraps from the surrounding environment with which to rebuild their ship. Gameplay alternates between two modes (see Figure 7.2): Exploration Mode, where the player explores a 3D environment in first person, collecting scrap parts from the environment, and Construction Mode, where the player constructs useful items out of the scrap parts they have collected piece by piece by choosing parts to attach, choosing the specific area on each part where the parts will attach, rotating them to align correctly, and then pushing and "fusing" them together. Constructed items can then be used in Exploration Mode to access new areas of the game and progress. Exploration Mode was designed to incorporate the *landmark orientation* and *navigation visualization* spatial game features, while Construction Mode was designed to incorporate the *object rotation* and *object alignment* features.

Before constructing an item, the player must first collect parts for and build four batteries, which are designed to be simpler warm-up levels preparing the player to build the item that will allow them to access a new area. Battery construction levels tend to have fewer parts than item construction levels, and limit only the number of rotations the player can perform on parts, while item construction levels limit both rotations and time. If the player runs out of rotations (or time, when applicable) in a Construction Mode level, the player must restart the level. Similarly, Exploration Mode levels where the player is collecting battery parts have no time limit, while Exploration Mode levels where the player is collecting item parts have a time limit and a static map that must be used to navigate to all the item parts within the time limit.

I chose to study the training effects of *Homeworld Bound: Redux* because while multiple papers have been published about its design and development process [40, 42], it has not yet been evaluated for training effectiveness or for its purported benefits as a training tool more motivating than traditional spatial training exercises for low spatial skill students. In addition, since the game contains detailed reporting of in-game player behavior and performance data for each level, I can analyze to what extent the features of each game level are actually tapping into players' pre-existing spatial skill to look for clues as to how different levels may be contributing more or less to the game's training effectiveness.

Spatial Exercises

As a point of comparison with the game-based intervention *Homeworld Bound: Redux*, I wanted to include a more traditional spatial training condition without game elements in order to determine if *Homeworld Bound: Redux* does indeed offer advantages in terms of motivation over a more traditional method of training that has already been shown to train college students' spatial skills. For this condition, I chose to use a subset of Sorby's spatial skill training exercises for first year engineering students. These workbook exercises, in their original form, included a wide variety of different activities, from sketching isometric views of objects to mental rotation and visualizing cross sections and revolved solids about axes and are designed to be used in the context of a college-level course dedicated to spatial skill training [128, 129]. They are the best example I know of in the research literature of empirically-proven spatial skill exercises designed specifically to help students succeed in a STEM discipline (engineering).

Previous work has adapted these workbook exercises for a digital format, enabling them to scale better in large classrooms [215]. With the help of collaborators who had implemented the digital version of the workbook exercises, I selected a subset of the digital version of the exercises to use in the current study that were appropriately challenging for non-STEM majors. Since Sorby's exercises were designed for engineering majors, who typically have higher spatial skill than students in non-STEM majors [2], my collaborators and I selected six sets of the easiest exercises that targeted spatial subskills most similar to the ones targeted by

Homeworld Bound: Redux in order to make a fairer comparison between the two interventions in my study. The exercises we chose focused on visualizing cross sections of objects, revolving solids about axes, and mental rotation around multiple axes (see Figure 7.2 for an example of each kind of exercise). We chose six sets of exercises since my collaborators estimated that this amount would take students at least an hour to finish based on their prior experience using the digital versions of Sorby's exercises in engineering classrooms.

Python Graphics

The third training intervention in my study was designed to be more of a regular classroom activity for the course I administered the study within - a simple programming lab using Python's Turtle graphics library, which enables lines and shapes to be drawn on a canvas with simple Python commands. This lab was chosen for this study among other labs in the course because it appeared to be a spatial activity. However, I considered it unlikely to train the kind of spatial skills I was studying given that the activity was confined to drawing simple 2D shapes such as squares and triangles (students' main tasks were to draw a line, draw a square, and draw an equilateral triangle next to the square), which I felt would not develop students' generalizable mental rotation skills. The lab was designed to be completable within the 70 minute lab session for all students.

7.3 RESULTS

My results consisted of a primary analysis of training effects and enjoyment of each intervention (**RQ1** and **RQ2**), followed by secondary analyses of individual characteristics that might affect training effectiveness and enjoyment (**RQ3** and **RQ4**), and a correlation analysis aimed at determining which levels of *Homeworld Bound: Redux* may be contributing the most or least to its training effectiveness (**RQ5**).

7.3.1 Data Cleaning

In total, 545 students completed all parts of the study (pretest, in-class training intervention, IMI enjoyment subscale, post-test, and demographic survey), and 490 (90%) consented to have their data used for this study. Within this group, I removed the data of students who experienced technical problems taking the pretest or post-test (31). In order to determine which students did not take the test seriously, I looked at the relationship between time spent on the test and test score. Inspecting plots of time spent versus score for the

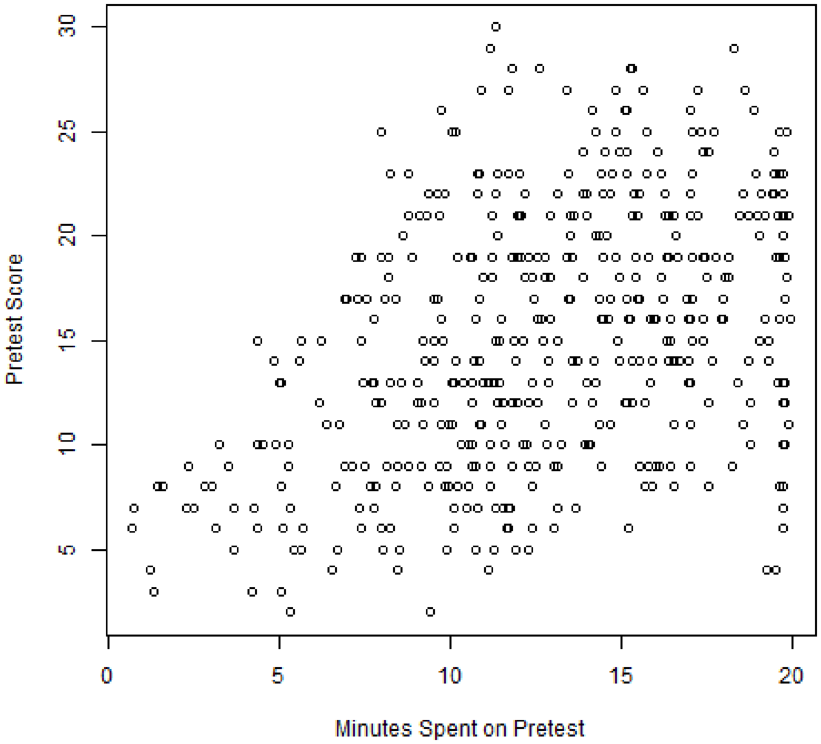


Figure 7.3: Plot of the relationship between score and time spent on the pretest used to establish a cutoff for removing the data of students who did not take the test seriously. Students were given 20 minutes to take the test, so the positive correlation between time spent on the test and test score for those who finished the test in well under 20 minutes (the left side of the graph) indicates that these participants were likely not taking the test seriously.

pretest and post-test (see Figures 7.3 and 7.4), I observed what appeared to be a positive correlation between time spent and test score for those who spent the least amount of time on the test up until a certain threshold of time spent, such that no one who spent less time than the threshold scored above around 15 points. For the pretest, this threshold seemed to be at about 8 minutes, whereas for the post-test it was about 5 minutes. Therefore, I decided to omit the data of all students who spent less than 8 minutes on the pretest or less than 5 minutes on the post-test since they used far less time than they were given for the test and received a low score. This criteria eliminated 116 additional students, leaving me with a total of 343 students in the dataset.

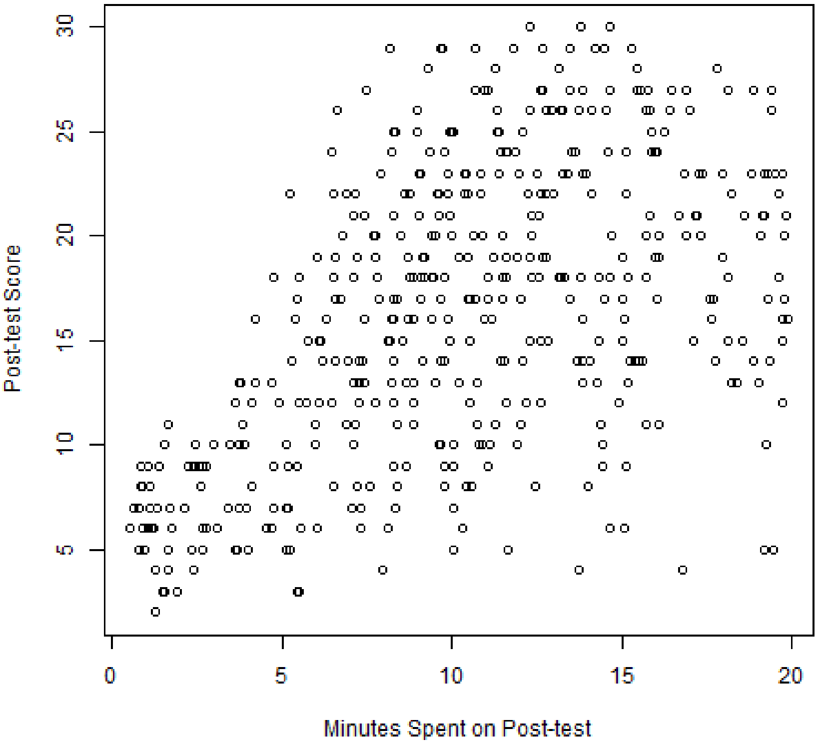


Figure 7.4: Plot of the relationship between score and time spent on the post-test used to establish a cutoff for removing the data of students who did not take the test seriously. Students were given 20 minutes to take the test, so the positive correlation between score and time spent on the test for those who finished the test in well under 20 minutes (the left side of the graph) indicates that these participants were likely not taking the test seriously.

7.3.2 Participant Characteristics

Of the students in the cleaned dataset, 59% identified as male, 40% as female, and the remaining 1% identified as nonbinary, genderfluid, gender non-conforming, or not specifying. Students' self-reported hours of video game play per week were mostly evenly distributed between the four lowest answer choices: 0 hours, Less than 1 hour, 1-2 hours, and 3-7 hours per week, with 51% of the sample answering either 0 hours or less than 1 hour. The distribution of scores on the spatial skill pretest is shown in Figure 7.5. Scores overall ($\mu = 16.44$, median = 17) were lower than the cutoff used in previous engineering education research for identifying low spatial skill students (18/30 or 60%) [131, 134].

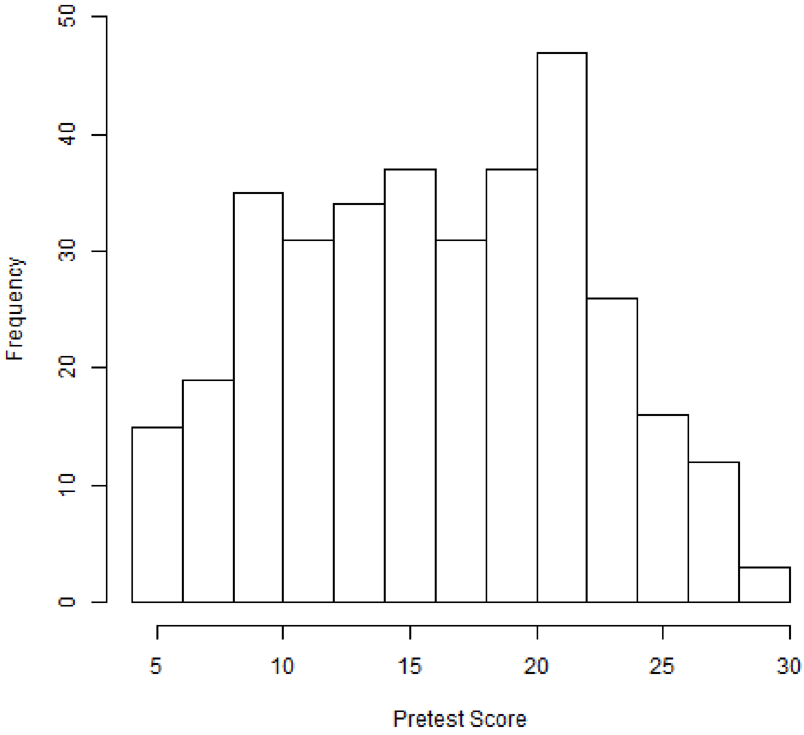


Figure 7.5: Histogram showing the distribution of Revised PSVT:R pretest scores. Scores had a roughly normal but left-skewed distribution, and mean and median performance fell below the threshold used in engineering education research for identifying low spatial skill students (18 out of a maximum of 30 points).

7.3.3 Training Effects and Enjoyment (RQ1 and RQ2)

To determine how training effectiveness compared across conditions for low spatial skill students (**RQ1**), I analyzed the data of students who scored 18 points out of 30 (60%) or lower on the spatial skills pretest ($n = 202$). Of these students, 72 were in the Spatial Exercises condition, and 65 were in each of the two remaining conditions (*Homeworld Bound: Redux* and Python Graphics). I used this cutoff since it is the same one Sorby et al. and Onyancha et al. used to determine if engineering students' spatial skills fell below a desired level of competence [131, 134]. As a check, I performed the same analyses using a 21/30 cutoff that is used by engineering professors at my university to identify first year students in need of extra spatial skill training, and found that the results did not materially differ between the two thresholds.

With this sample of low spatial skill students, I ran an ANCOVA with post-test score as the dependent variable, treatment condition (*Homeworld Bound: Redux*, Spatial Exercises, or Python Graphics) as the independent variable, and pretest score as a covariate. Based on inspection of diagnostic plots and interaction effects analysis, ANCOVA assumptions of normality, homoscedasticity, linearity, and homoegeneity of regression slopes were met. There was no significant effect of treatment condition on post-test score, controlling for pretest score ($F(2, 198) = 0.33$, $p = 0.72$), all Cohen's $d < 0.10$.

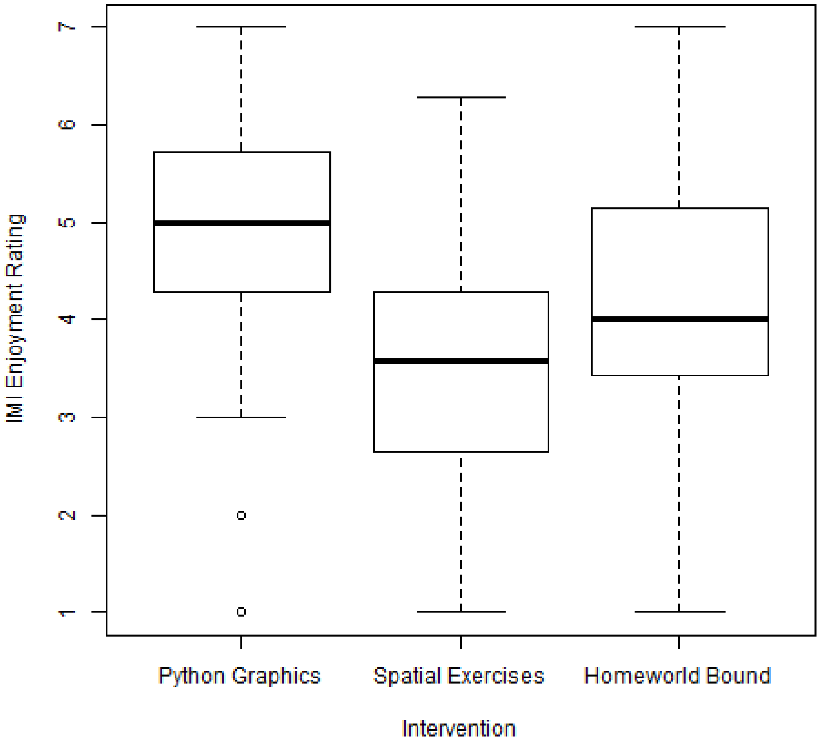


Figure 7.6: Intrinsic Motivation Inventory (IMI) ratings of enjoyment for each training intervention.

To determine which spatial training condition was most motivating for low spatial skill students (**RQ2**), I performed a one-way ANOVA using the same dataset as described above, with average IMI score as the dependent variable and treatment condition as the independent variable. Cronbach's α for the IMI was 0.93, indicating good reliability, and inspection of diagnostic plots confirmed that assumptions of normality and homoscedasticity were met. The ANOVA revealed a significant effect of treatment on IMI score ($F(2, 196) = 21.54$, $p < 0.0001$). A posthoc Tukey test showed that all pairwise differences were significant; the Python Graphics condition ($\mu = 4.90$) was rated higher than both the Spatial Exercises ($\mu = 3.44$, $d = 1.23$, $p < 0.0001$) and *Homeworld Bound: Redux* ($\mu = 4.20$, $d = 0.52$, $p = 0.0076$), and *Homeworld Bound: Redux* was rated higher than the Spatial Exercises

($d = 0.56$, $p = 0.0027$) (Figure 7.6). Thus, students reported enjoying the Python Graphics condition the most, but enjoyed *Homeworld Bound: Redux* more than the Spatial Exercises.

7.3.4 Predictors of Improvement and Enjoyment (RQ3 and RQ4)

To determine what individual characteristics and demographic factors may have influenced training effectiveness in each condition (**RQ3**), I conducted three linear regressions (one for each treatment condition) with pretest to post-test spatial skill improvement as the dependent measure and pretest score, weekly hours of video game play, and gender as predictors. Since the group of participants who specified a gender other than male or female was too small to include in a linear regression analysis ($n = 3$), I removed them from the dataset for the purposes of this analysis, but otherwise used my entire dataset ($n = 340$). Based on inspection of diagnostic plots, linear regression assumptions of linearity, normality, and homoscedasticity were met, and variance inflation factors were less than 2 for all linear regressions, indicating no multicollinearity problems. For each linear regression analysis, I report the standardized regression coefficient β in addition to the unstandardized regression coefficient B to aid in effect size comparisons between coefficients.

In the *Homeworld Bound: Redux* condition, pretest score ($B = -0.19$, $\beta = -0.26$, $t = -2.64$, $p = 0.0096$) was the only significant predictor of spatial skill improvement such that those who did worse on the spatial skills pretest improved more from pretest to post-test. However, the overall model was not significant and had a low R^2 value (adjusted $R^2 = 0.037$, $F(3, 103) = 2.36$, $p = 0.076$). For the Spatial Exercises condition, pretest score was the only significant predictor ($B = -0.37$, $\beta = -0.43$, $t = -5.03$, $p < 0.0001$), but unlike in the *Homeworld Bound: Redux* condition, the overall model was significant (adjusted $R^2 = 0.18$, $F(3, 119) = 9.89$, $p < 0.0001$). Similarly, for the Python Graphics condition, pretest score was the only significant predictor ($B = -0.34$, $\beta = -0.43$, $t = -4.93$, $p < 0.0001$), and the overall model was significant (adjusted $R^2 = 0.16$, $F(3, 106) = 8.15$, $p < 0.0001$).

To determine what individual characteristics and demographic factors may have influenced enjoyment of training interventions (**RQ4**), I conducted one linear regression for each treatment condition in the same manner as described above, with IMI enjoyment score as the dependent measure and pretest score, weekly hours of video game play, and gender as predictors. In the *Homeworld Bound: Redux* condition, there were no significant predictors of enjoyment. The model overall was significant, but had a low R^2 value (adjusted $R^2 = 0.056$, $F(3, 100) = 3.05$, $p = 0.032$). For the Spatial Exercises condition, only pretest score was a significant predictor of enjoyment ($B = 0.040$, $\beta = 0.20$, $t = 2.27$, $p = 0.025$) such that those with higher pretest scores enjoyed the intervention more, and the model was significant (adjusted $R^2 = 0.14$, $F(3, 119) = 7.61$, $p = 0.00011$). In the Python Graphics condition, there were no significant predictors of enjoyment. The model overall was significant but had a low R^2 value (adjusted $R^2 = 0.047$, $F(3, 106) = 2.78$, $p = 0.045$).

7.3.5 *Homeworld Bound: Redux* Level Performance and Spatial Skill (RQ5)

To understand the relative effectiveness of different parts of *Homeworld Bound: Redux* at tapping into players' spatial skills and discover potential ways of improving the game's training effectiveness (**RQ5**), I conducted a linear regression analysis investigating the relationship between students' in-game performance and their current level of spatial skill in the *Homeworld Bound: Redux* condition, as measured by pretest score. There were a total of 107 students in the *Homeworld Bound: Redux* condition. However, a series of unreproducible browser crashes during the study caused a loss of player performance data for a subset of students in this condition. These crashes affected the data of 31 students, so after eliminating these students, I was left with $n = 76$ students for the linear regression analysis. For all linear regression analyses in this section, inspection of diagnostic plots, linear regression assumptions of linearity, normality, and homoscedasticity were met for each regression, and variance inflation factors were less than 2, indicating no multicollinearity. For each linear regression analysis, I report the standardized regression coefficient β in addition to the unstandardized regression coefficient B to aid in effect size comparisons between coefficients.

Overall, the number of levels students completed in *Homeworld Bound: Redux* varied widely, ranging from a minimum of 5 to a maximum of 20. On average, students completed 10.45 levels (median: 9.5), and the distribution was heavily skewed towards the low end of the range (only a handful of students completed a large number of levels). My first analysis evaluated whether performance in *Homeworld Bound: Redux* as a whole, as measured by number of levels completed, was related to current spatial skill level. I performed a linear regression analysis for this purpose with pretest score as the dependent measure, number of levels completed as the independent variable, and self-reported weekly hours of video game play as a covariate to control for the effect of prior video game experience. The model was significant overall (adjusted $R^2 = 0.21$, $F(3, 73) = 10.73$, $p < 0.0001$). Levels completed was a significant predictor of pretest score after controlling for weekly video game hours ($B = 0.73$, $\beta = 0.48$, $t = 4.33$, $p < 0.0001$), while weekly video game hours was not after controlling for levels completed ($B = -0.03$, $\beta = -0.0068$, $t = -0.061$, $p = 0.95$).

To get a more fine-grained understanding of how performance in *Homeworld Bound: Redux* was related to current spatial skill level, I conducted a series of regressions looking at the relationship between students' spatial skill pretest score (dependent measure) and performance in individual *Homeworld Bound: Redux* levels (predictor) while controlling for prior gaming experience (covariate). Since the number of students who had completed more than 8 levels was significantly smaller than the generally recommended sample size for linear regression [200], I analyzed performance for only the first 8 levels (the first 4 levels of Exploration Mode and the first 4 levels of Construction Mode).

		Coefficient values				**Model Values**			
Level	*Measure*	B	β	t	p	R^2 *(adj)*	*df*	F	p
Exp B1	Time	0.0038	0.078	0.67	0.51	0.0084	73	1.32	0.27
Exp B2	Time	−0.0099	−0.10	−0.85	0.40	0.012	73	1.45	0.24
Exp B3	Time	−0.0066	−0.11	−0.97	0.34	0.011	72	1.39	0.25
Exp B4	Time	−0.0019	−0.085	−0.65	0.52	−0.0068	58	0.80	0.46
Exp All	Time	−0.00095	−0.053	−0.39	0.70	−0.011	58	0.66	0.52
Con B1	Attempts	−0.58	−0.19	−1.62	0.11	0.037	73	2.44	0.094
Con B1	Time*	−0.0069	−0.26	−2.27	0.026	0.068	73	3.73	0.029
Con B1	Errors*	−0.29	−0.25	−2.20	0.031	0.062	72	3.43	0.037
Con B2	Attempts**	−0.83	−0.43	−4.07	0.00012	0.19	73	9.60	0.0002
Con B2	Time**	−0.0080	−0.46	−4.46	0.0001	0.22	73	11.33	0.0001
Con B2	Errors*	−0.090	−0.27	−2.38	0.020	0.074	73	4.00	0.022
Con B3	Attempts**	−1.16	−0.41	−3.58	0.00068	0.16	62	7.09	0.0017
Con B3	Time**	−0.0071	−0.50	−4.49	0.0001	0.23	62	11.33	0.0001
Con B3	Errors*	−0.10	−0.32	−2.68	0.0094	0.091	62	4.22	0.019
Con B4	Attempts	−0.13	−0.047	−0.31	0.76	−0.039	45	0.13	0.88
Con B4	Time	−0.0014	−0.087	−0.56	0.58	−0.034	45	0.24	0.79
Con B4	Errors	0.022	0.064	0.42	0.64	0.0074	45	−0.037	0.85

Table 7.1: Summary of regression analysis results for first 8 *Homeworld Bound: Redux* levels (Exp = Exploration, Con = Construction). B is the unstandardized regression coefficient, β is the standardized regression coefficient. $*p < \alpha$ without Bonferroni correction ($\alpha = 0.05$), $**p < \alpha$ with Bonferroni correction ($\alpha = 0.05/17 = 0.0029$) for β coefficients. p-values less than 0.0001 are reported as 0.0001. Time is measured in seconds. Exp All refers to the combined "All Batteries" level. For the sake of brevity, B and β values for weekly video game hours are not shown.

Since the first 4 levels of Exploration Mode (Exploration B1-Exploration B4) are battery part collection levels and therefore contain no timed levels, the only measure of performance in these levels is completion time. Furthermore, since parts can be collected in any order in Exploration Mode and an Exploration Mode level ends each time a certain number of battery parts have been collected, I decided to also conduct a regression treating all 4 battery part collection levels as one big level (Exploration All). Since none of the first 4 Construction Mode levels (Construction B1-Construction B4) are timed levels either, the measures of performance for these levels are number of attempts, completion time, and number of errors made. Every time the player must restart a Construction Mode level due to running out of rotation operations, a new attempt is counted. An error is counted whenever the player attempts to attach two parts together in Construction mode and either a) has selected two parts that cannot be attached to each other, b) has selected two parts that can be attached to each other, but has selected the wrong areas on each part to attach, or c) has selected the correct areas on each part to attach, but one of the parts is not rotated correctly to align with the other so they can be pushed together. I performed one linear regression per performance measure per level, for a total of 17 linear regressions.

The results of the regression analysis are summarized in Table 7.1. Overall, better performance in most Construction Mode levels is associated with higher spatial skill pretest scores even after controlling for prior video game experience, especially for B2 and B3. In contrast, performance was not associated with spatial skill pretest score in any Exploration Mode levels. Controlling for performance, weekly hours of video games played was not associated with pretest score in any linear regression. Performance was comparable across levels except for B1 having noticeably fewer errors (see Table 7.2).

7.4 DISCUSSION AND FUTURE WORK

7.4.1 RQ1 & RQ2: Training & Motivational Effects for Low Spatial Skill Students

Overall, I found no evidence of a training effect for any intervention (**RQ1**). It is not too surprising that I failed to find training effects for the spatial interventions given that most game-based spatial skill training studies use a training period of at least several hours [112, 3, 126, 150, 147, 5], in contrast to the 70 minute training period I used, a product of my decision to sacrifice training time for a larger sample of low spatial skill students. In addition, the Spatial Exercises and Python Graphics interventions did not take the full class period as intended for most students, so average training duration was even shorter in these conditions (around 30 minutes). Despite this, I still believed I might observe training effects in the *Homeworld Bound: Redux* or *Spatial Exercises* conditions since other controlled studies have found spatial skill training effects using training times of an hour or less [152, 30, 153].

Level	*Measure*	*Median*	*Mean*
Exploration B1	Time	108.36	138.83
Exploration B2	Time	72.80	91.23
Exploration B3	Time	137.15	147.99
Exploration B4	Time	196.72	308.97
Construction B1	Attempts	2	2.71
Construction B2	Attempts	3	3.67
Construction B3	Attempts	3	3.6
Construction B4	Attempts	3	3.04
Construction B1	Time	292.23	325.36
Construction B2	Time	378.34	448.89
Construction B3	Time	709.69	827.83
Construction B4	Time	482.02	579.46
Construction B1	Errors	3.5	5.53
Construction B2	Errors	5.5	11.43
Construction B3	Errors	13.0	18.12
Construction B4	Errors	6.0	10.67

Table 7.2: Median and mean performance on first 8 *Homeworld Bound: Redux* levels. Time is measured in seconds. Since there were an even number of data points, medians were calculated by taking the mean of the middle two data points.

One key difference between these other short interval training studies and my training study was that each of them conducted pretests, interventions, and post-tests within a very short time period. Mazalek et al. and Chang et al. conducted the pretest, intervention, and post-test all in the same day, and Cherney et al. conducted them within 48 hours of each other, while I spaced my pretest, intervention, and post-test of spatial skills out to ensure that there was at least a week long gap between the pretest and the intervention and between the intervention and the post-test. Adding these gaps allowed me to minimize possible test-retest effects and test the longer term transferability of each training intervention. Therefore, it is likely that spatial skill training lasting an hour or less results in only very short term training effects, whereas longer training periods are necessary to observe lasting training effects. Future work should evaluate the training effects of playing *Homeworld Bound: Redux* with a much longer training period, spaced across multiple weeks.

Low spatial skill students enjoyed *Homeworld Bound: Redux* more than the other intervention intended to train spatial skills, Spatial Exercises (**RQ2**), with a medium effect size ($d = 0.56$, 0.76 points on a 7 point scale). This finding provides evidence that the game-based intervention *Homeworld Bound: Redux* provides motivational benefits over more traditional workbook exercises. However, it is interesting to note that the active control activity, the Python Graphics lab, was enjoyed more than either of the training interventions. This may be due to low spatial skill students preferring less spatially demanding activities. Another possibility is that the Python Graphics lab was rated higher because it may have been easier to complete than the other two; many students finished this activity in 20-30 minutes, while students in the Spatial Exercises condition typically took 30-35 minutes to finish, and students playing *Homeworld Bound: Redux* had enough content to play for the entire 70 minute lab period. Future work should investigate the extent to which student enjoyment of an intervention may change over the course of a longer term intervention.

7.4.2 RQ3 and RQ4: Influence of Demographics on Training Effects and Enjoyment

Analyzing potential demographic differences in pretest to post-test improvement revealed that pretest score was the only significant predictor in all three conditions; those with lower pretest scores improved more on the post-test, while there was no significant effect of weekly hours of video game play or gender on test score improvement. The standardized regression coefficients for pretest score in each condition ranged from -0.26 to -0.43, indicating that an increase of one standard deviation (5.83 points) on the pretest would predict a decrease of anywhere between 1.08 points and 1.77 points in pretest to post-test improvement. This somewhat modest influence is likely due to a ceiling effect since there is more room for improvement on lower pretest scores than on high ones and the effect was observed in all three conditions.

When looking at the same demographic predictors for enjoyment, only pretest score was a significant predictor of intervention enjoyment, and only in the Spatial Exercises condition. In addition, the effect size for this predictor was too small to have practically useful predictive power. With a standardized regression coefficient of 0.20 for pretest score, an increase of four standard deviations (24.52 points on a test with a maximum score of 30 points) on the pretest corresponds to an increase of just 1 point on the enjoyment rating scale (out of 7 points).

Thus, I found no evidence that demographics influenced training effects (**RQ3**) or enjoyment of training intervention (**RQ4**) in any practically significant way. However, this finding may simply be due to the fact that there was not enough time in the 70 minute training period I used to produce measurable training effects for any participants. More research with a longer intervention time is needed to determine if any effects of demographics might become visible in the longer term for either training effectiveness or enjoyment. In addition, future work should investigate other demographic variables that might affect students' enjoyment of a particular spatial skill training intervention, such as socioeconomic status and preferred types of video games, both of which have been shown to be related to spatial skill in previous work [108, 32, 27, 110].

7.4.3 RQ5: Certain *Homeworld Bound: Redux* Levels Tap Into Players' Spatial Skill

My regression analysis revealed that most Construction Mode levels in *Homeworld Bound: Redux* tapped into players' spatial skills. The relationship between in-game performance and spatial skill was particularly strong in the B2 and B3 levels for number of attempts and completion time, where the largest effect sizes were found (see Table 7.1). For instance, every increase of one standard deviation in number of attempts (2.06 attempts) corresponded to a decrease of 2.31 points on pretest score in the B3 level. Likewise, every increase of one standard deviation in completion time (6.76 minutes) corresponded to a decrease of 2.81 points on pretest score in the B3 level. Thus, the game mechanics of *object rotation* and *object alignment* appear to be spatial in nature for the B2 and B3 levels. However, performance in the B1 level had a weaker association with spatial skill, and performance in the B4 level had no significant association with spatial skill. The B1 level may have been a little too easy to adequately tap into players' spatial skills given that performance in this level seemed to be better than in B2 and B3 (see Table 7.2). It is also possible that the learning curve involved in simply learning how to play affected performance more than spatial skill since B1 was the first Construction Mode level. On the other hand, performance on the B4 level was not noticeably better than on B2 and B3, so this explanation is insufficient for B4.

Another more likely possibility is that players are using the shapes of the 2D attachment areas on object parts to align parts rather than the 3D shapes of the parts themselves. In the B1 and B4 levels, these 2D attachment areas are generally rectangular or triangular, whereas in the B2 and B3 levels, the shapes are more complex (see Figure 7.7). When determining how a part should be rotated to line up its attachment area with the corresponding area on another part, the more complex attachment shapes in B2 and B3 likely require much more mental rotation effort than the simpler and more symmetrical shapes found in B1 and B4.

Performance in Exploration Mode levels, on the other hand, was not predictive of spatial skill. Given that several games successful at training spatial skills require the player to navigate a 3D environment in first person (e.g. *Unreal Tournament, Portal 2, Medal of*

Honor: Pacific Assault), it is a bit surprising that no Exploration Mode levels tap into spatial skills given its emphasis on first person navigation. It seems it is not enough for a game to give players a navigation task in first person. It may be that the navigation task in the first four Exploration Mode levels was too easy, so players could simply wander around until they found all the parts, eliminating the need to use *landmark orientation*.

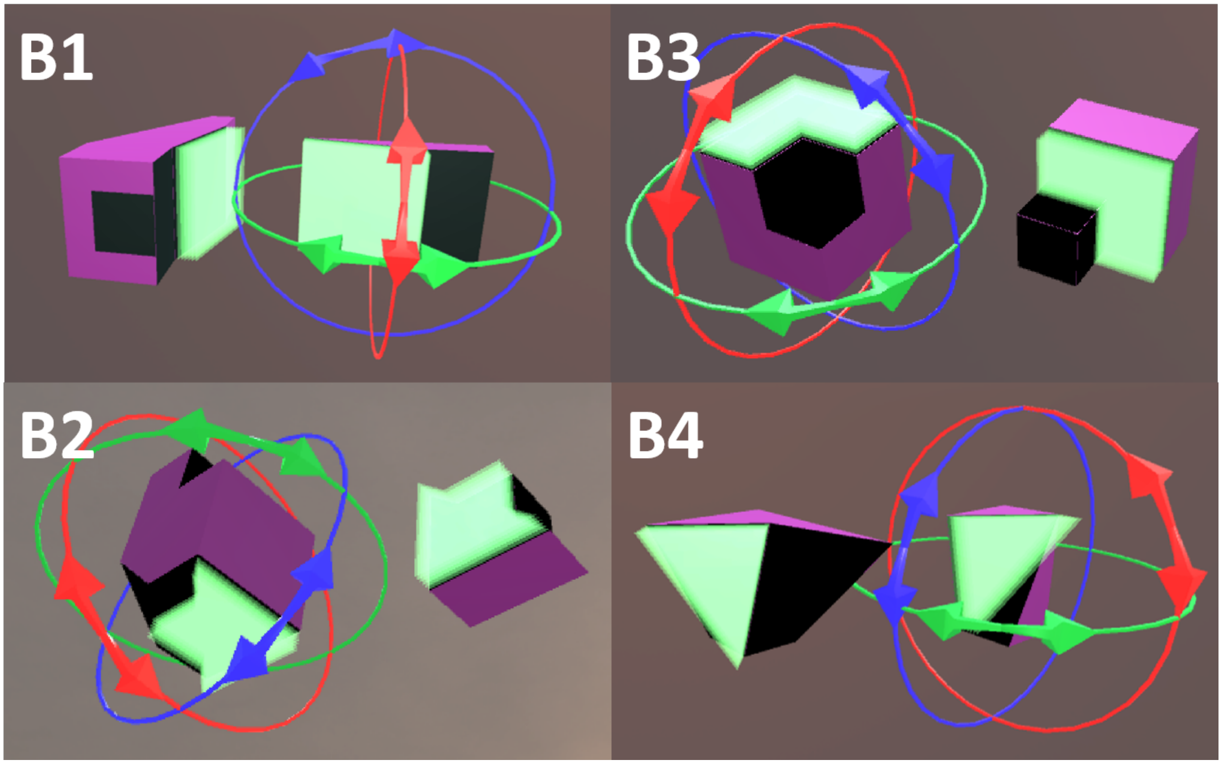


Figure 7.7: Examples of 2D attachment area shapes in levels B1-B4 in *Homeworld Bound: Redux*'s Construction mode. Glowing green areas indicate the attachment areas the player is currently trying to align by rotating one of the parts. In levels B1 and B4, the 2D shapes are simpler and more symmetric, potentially reducing the need to employ mental rotation skills to complete the level.

Another possibility is that the environment had enough nearby landmark cues that players' need to engage their mental rotation skills was reduced, a phenomenon observed in prior work on navigation-based games utilizing such *proximal* landmark cues [155]. A third possibility is that the presence of a map may be crucial for ensuring players engage mental rotation skills. First person shooters like *Unreal Tournament* as well as other navigation-based games shown to train spatial skills like the *Segway Circuit* minigame in *Wii Fit* have minimaps the player must use to orient themselves in the environment.

Some later Exploration Mode levels designed to tap more into players' *navigation visualization* skills include a static map of the environment and a time limit, which may force players to be more strategic in their navigation. The last level in the current version of the game also requires players to collect clues that depict the location of parts. Each of these features may require players to engage in more explicit *landmark orientation* that is likely to engage their mental rotation skills more fully. Future work is needed to explore the relationship between performance and spatial skill on these later levels to test this hypothesis.

In summary, I found that the *object rotation* and *object alignment* features as implemented in Construction Mode levels with more complex, asymmetrical attachment areas were able to tap into players' spatial skills more effectively than the *object rotation* and *object alignment* features found in other Construction Mode levels, while the *landmark orientation* and *navigation visualization* features as implemented in all Exploration Mode levels analyzed were not effective at tapping into players' spatial skill (**RQ5**).

7.4.4 Design Recommendations

Given my findings with regard to **RQ5**, I recommend increasing the visual complexity and asymmetry of simpler attachment shapes in *Homeworld Bound: Redux's* Construction Mode levels so that mentally rotating them requires more effort. Another solution could be to make all the shapes and sizes of attachment areas in a given level identical so that the player is forced to rely on cues other than the shape of attachment areas (e.g. the picture of the finished object and the shape of the 3D object itself) to determine where to attach parts and how to rotate them. A more general design solution for spatial skill training games involving *object rotation* and *object alignment* is to ensure that all shapes being rotated or aligned (whether 2D or 3D) have some degree of asymmetry and that the level design does not enable any circumvention of mental rotation via alternate strategies that may have been unintentionally built into the game by its designers.

Since I found no evidence that the first four levels of Exploration Mode in *Homeworld Bound: Redux* tapped into players' spatial skills, I recommend that they be revised to increase the difficulty of finding parts and to incorporate some sort of map use to prevent players from engaging in nonstrategic wandering that does not tap into their spatial skills. These revisions could also be applied to any spatial skill training game where the player has to explore in first person and collect items in the environment. Pictorial clues showing a screenshot or drawing of an item's location and immediate surroundings could be used in combination with environmental elements that make it impossible or very difficult to find items without using *landmark orientation* to figure out where to go to see the view of the item shown in the clue, as in the Ruined City level in the original *Homeworld Bound* (see Chapter 3). Parts could be hidden behind or inside of objects, or buried in the ground so the player must be in the exact right spot to dig them up.

Incorporating map use into a first person navigation game would likely need to be combined with some sort of time pressure in order to motivate players to use the map and engage their *landmark orientation* skills instead of wandering around randomly until they find whatever they are looking for. Time pressure could be achieved by having the player be chased by enemies, an environmental hazard (like a flood, earthquake, fire, or the rolling stone in *Indiana Jones*) or by simply giving the player a fixed amount of time to navigate through the environment. Designers of spatial skill training games incorporating *landmark orientation* and *navigation visualization* tasks should also limit the presence of proximal landmark cues in the game's environment since their presence may reduce the need for players to exercise their spatial skills while navigating [155].

It is possible that the ability to see relationships between performance in *Homeworld Bound: Redux* and spatial skill was limited by my decision to use completion time as the sole measure of performance in Exploration Mode. Differences in performance might have become evident if more sophisticated metrics such as path traces were employed to analyze how strategic and efficient each player's search strategy for parts was, as was done in DeCastell et al.'s study with a water maze task [155]. Future work could implement automatic path traces in *Homeworld Bound: Redux* or other navigation-based games to enable detailed and scalable analysis of the link between in-game performance and spatial skill.

Since *Homeworld Bound: Redux* is an open source game available on Github (`https://github.com/hwauck/homeworld-bound`), I encourage future work from interested research teams around the world to revise the game based on my recommendations in order to improve its ability to tap into players' spatial skill across all game levels, which may in turn help it to become more effective at training students' spatial skills. My findings can also be used to inform the design of other game-based or virtual spatial skill training interventions, removing some of the guesswork from the design process and potentially increasing the efficiency of game-based spatial skill training.

Chapter 8: Discussion and Future Directions

Spatial skills, although critical for success in STEM disciplines, have long been ignored in formal education in favor of focusing exclusively on developing students' mathematical and verbal skills [216, 217]. Fortunately, the importance of spatial skills in promoting access to STEM majors and careers is beginning to be recognized, as evidenced by the emergence over the last couple decades of a substantial body of literature analyzing the effectiveness of various spatial skill training interventions both in the laboratory [151, 124, 112, 115, 27, 3, 4, 30, 5] and in the classroom [218, 15, 128, 129, 131, 133]. A large portion of such studies analyze the training effectiveness of existing commercial games. Games have great potential for not only developing learners' spatial skills, but harnessing the power of intrinsic motivation - the whole point of game-based interventions - to encourage learners to persist and stay engaged with the intervention. However, the potential of game-based spatial skill training interventions cannot be realized if we study only existing commercial games and do not have the knowledge necessary to build our own effective game-based interventions that are intrinsically motivating for their target audience.

The central contribution and goal of this dissertation is twofold. First, I provide researchers, educators, and game designers with a data-driven, player-centric approach to designing and evaluating games for training spatial skills that can be implemented to produce fun and effective spatial skill training games to help students succeed in STEM majors and careers. Second, I provide an open source, free, and easily accessible game designed using this approach for use and modification by researchers, game designers, and educators as a spatial skill training intervention and a testbed for future spatial skill training game research. In the following sections, I summarize the major findings of this dissertation and discuss potential directions for future research.

8.1 A THEORETICAL FRAMEWORK FOR MAPPING SPATIAL GAME FEATURES TO SPATIAL SKILLS

In Chapter 3, I presented a theoretically grounded approach to developing games and game features that tap into players' spatial skills. This approach was grounded in a theoretical framework I developed to map different dimensions of spatial skill (*intrinsic-dynamic*, *extrinsic-static*, and *extrinsic-dynamic*) to different game features (*object rotation*, *object alignment*, *landmark orientation*, and *navigation visualization*). I developed this theoretical framework based on my analysis of both the existing research literature on spatial skill typology and existing empirical evidence of which video games are effective at training players' spatial skills. I then demonstrated the implementation of my theoretical framework in a functional, engaging children's computer game: *Homeworld Bound*. My small scale training study found that *Homeworld Bound* improved children's scores on a test of perspective-taking (an *extrinsic-static* spatial skill) over the course of four 45 minute sessions, indicating that a game designed using this framework can be an effective spatial skill training intervention. In addition, my analysis of the relationship between spatial skill and level-by-level performance for *Homeworld Bound* in Chapter 5 and its revised version, *Homeworld Bound: Redux* in Chapter 7 provided evidence that each game feature in my theoretical framework, as implemented in the games, tapped into players' spatial skills in some fashion.

However, the results of these studies reveal that my framework may be a bit too generic for consistently effective application across different player ages. While I was able to show that most levels in *Homeworld Bound* tapped into children's spatial skills, including both Construction Mode (*object rotation* and *object alignment* features) and Exploration Mode (*landmark orientation* and *navigation visualization* features), no Exploration Mode levels and only about half of Construction Mode levels analyzed in my study of *Homeworld Bound: Redux* with adults were effective at tapping into players' spatial skills. This suggests that certain simpler spatial features, such as navigating through a 3D environment by walking and jumping in Exploration Mode, while adequate for tapping into children's spatial skills, may be too simple to tax adults' spatial skills.

Thus, it might be helpful to have separate sets of guidelines for games designed for children and for adults, with the adult version specifying more complex features. Based on my finding from Chapter 7 that performance in Exploration Mode levels without maps or a time limit was not associated with spatial skill, orientation based on translating between a 2D map and a 3D environment may be a better way of tapping into players' *extrinsic-static* spatial skills than *landmark orientation*, and using time-constrained navigation tasks in combination with a map (as in certain Exploration Mode levels in *Homeworld Bound: Redux* and commercial games shown to train spatial skills like *Crazy Taxi*, *Segway Circuit*, and *Medal of Honor*) may do a better job of tapping into adults' *extrinsic-dynamic* spatial skills than the more generic feature *navigation visualization*. Similarly, asymmetrical shape rotation and asymmetrical shape alignment may be better features for tapping into adults' *intrinsic-dynamic* spatial skill than simply *object rotation* or *object alignment*. Future work should investigate the viability of this new proposed model for adults with studies analyzing the relationship between performance and spatial skill across more levels of *Homeworld Bound: Redux* than I was able to and with a variety of different types of games containing these hypothesized spatial features.

Another limitation of my existing framework is that most of my studies tested the ability of game features to tap into or train only one dimension of spatial skill, *intrinsic-dynamic* spatial skills. Only one of them analyzed a second dimension, *extrinsic-static* spatial skill, and none of them tested my games' ability to tap into players' *extrinsic-dynamic* spatial skill. This was largely due to the desire to not induce mental fatigue in my participants by administering many pretests and post-tests and the constraints of the classroom contexts where I conducted almost all of my studies. However, in order to truly test whether the mapping I have proposed between spatial game features and different dimensions of spatial skill is accurate, future work using tests of each of the spatial skill dimensions in my framework is needed to investigate the relationship between player performance and spatial skill as well as training effects in *Homeworld Bound: Redux* and other games with the spatial game features in my framework.

8.2 A DATA-DRIVEN APPROACH TO DESIGNING AND EVALUATING SPATIAL GAME FEATURES

In Chapter 4, I took a data-driven approach to studying what game features may contribute to the effectiveness of a game at training spatial skills, using my spatial skill training game *Homeworld Bound* as a testbed. I presented the first empirical study analyzing the potential of different game features to tap into children's spatial skills, how this relationship differs by gender, and gave practical recommendations for implementing these features in games to assess or train players' spatial skills based on the complex relationship between gender, spatial skill, and in-game behavior my results revealed. My data-driven, level-by-level player behavior analysis revealed that most levels in *Homeworld Bound* tapped into children's spatial skills via some dimension of player performance (number of rotation operations and errors in Construction Mode, completion time in both Construction Mode and Exploration Mode).

These findings gave me important insights about which levels of the game were not adequately taxing children's spatial skills and allowed me to hypothesize potential reasons for their inadequacy. These reasons, in turn, allowed me to make the revisions to the game described in Chapter 6 that could be tested in a future study to see if they strengthened the relationship between spatial skills and player performance in each problematic level. Since the revised version of the game, *Homeworld Bound: Redux*, was designed to appeal to low spatial skill high school students and college-age adults (based on the player research I conducted with this target population in Chapter 5, the revisions I made could not be tested with children, *Homeworld Bound*'s original target population. Future research should test

Homeworld Bound: Redux, the revised version of the game (or perhaps an easier version of it), with children to see if these revisions were successful.

My similar analysis of the relationship between player performance and spatial skill with adults playing *Homeworld Bound: Redux* in Chapter 7 showed that no Exploration Mode levels and only about half of Construction Mode levels were effective at tapping into players' spatial skills. As in my study with children, getting these insights allowed me to hypothesize potential reasons for the lack of association between player performance and spatial skill in various levels of the game and suggest potential fixes that might be implemented and tested in future work. While this future work lies outside the scope of this dissertation, I encourage other researchers to conduct studies with *Homeworld Bound: Redux* with adult low spatial skill populations after implementing these fixes to see if they strengthen the relationship between spatial skill and player performance in problematic levels. I also encourage future work analyzing the relationship between player performance and spatial skill using the present version of *Homeworld Bound: Redux* with a longer play time than was possible in my evaluation study in Chapter 7. Since the time participants had to play in my study was 70 minutes and most players only got through the first four levels of Exploration Mode and the first four levels of Construction Mode, doubling or tripling the intervention time would likely give players enough time to play through the later levels of Construction and Exploration Mode that have as-yet untested features.

The research community could employ approaches similar to the data-driven approach described in this dissertation to diagnose problems with other spatial skill training games. This approach may not be feasible with existing commercial games unless game modding software allows for the implementation of detailed level-by-level player performance data collection. If conducting this kind of level-by-level performance and spatial skill analysis with other noncommercial games, care must be taken that they are designed not only to implement detailed telemetry data but also to have modular level structure like *Homeworld Bound: Redux*. Different levels should contain different combinations of spatial features so that it is possible to determine which combinations of features might be tapping into players' spatial skills.

8.2.1 How much level modularity is needed to evaluate specific features?

One criticism of my approach to designing for modularity of level structure in *Homeworld Bound* and *Homeworld Bound: Redux* is that while different combinations of spatial features are present in each level, no feature can be tested in complete isolation. Future work could tackle this problem in two ways. First, creating simpler, more bare-bones games could help isolate individual low-level spatial features by virtue of the game simply possessing fewer potentially relevant game features. However, creating games that are simple enough to test individual spatial features might eliminate some important player experience features, decreasing player motivation and making the training intervention less game-like. Furthermore, the kinds of games that have proven most effective at far transfer spatial skill training are not simple and contain a large variety of different spatial game features that in turn produce a large variety of different spatial tasks for the player. As Spence and Feng as well as Green and Bavelier note, a training intervention designed to target a very specific cognitive function (using much more homogeneous selection of spatial tasks) is likely to produce training effects that do not transfer more generally and remain task-specific [219, 59]. Thus, I argue that the kind of game complexity that makes commercial video games such interesting, immersive, engaging, and varied experiences for players should be preserved in game-based spatial skill training interventions.

Another potential concern with the use of complex games like *Homeworld Bound* and *Homeworld Bound: Redux* as training interventions for spatial skills is that such complex games may contain features that, while enhancing the player experience, are irrelevant to the development of spatial skill and distract the player from learning. Such *seductive details* have been shown in the education research literature to hinder learning in a variety of multimedia contexts [220, 221, 222] (see [223] for a review). While the decision to make *Homeworld Bound* and *Homeworld Bound: Redux* complex games with large numbers of features risks introducing seductive features that distract from spatial learning, I believe potential benefits of increased transfer effects as noted by Green and Spence and improved player experience are worth the risks given the crucial advantages both present compared to more targeted and non-game spatial skill training interventions.

Furthermore, the intentional, data-driven design process I used when adding spatial and experiential game features to *Homeworld Bound: Redux* reduces the risk of introducing seductive features unrelated to spatial skill. *Homeworld Bound* and *Homeworld Bound: Redux* were designed carefully to make sure all game features, even those incorporated for enhancing player experience, were tied into the spatial aspects of gameplay. For instance, the visuals of each level's environment (rock colors and shapes in Exploration Mode, part colors and shapes in Construction Mode) were selected to serve as cues for the spatial tasks that form the core of its gameplay (*landmark orientation, navigation visualization, object rotation,* and *object alignment*). The game's narrative, conveyed via brief dialogue, contained the minimum amount of text necessary to convey how to play the game and to give the player a sense of purpose behind the spatial tasks they were performing - for example, why collect batteries? Why is there a time limit for certain levels? Why is the player collecting parts and using them to build items? This same pattern is seen in commercial games shown to be effective at training spatial skills like *Medal of Honor*, *Super Mario 64*, *Portal 2*, *Crazy Taxi*, and *Unreal Tournament*. These complex games may have a large number of features, but they limit the number of game features unrelated to spatial game mechanics (such as narrative dialogue) and tie their other game features (e.g. game mechanics, visual environments) in with spatial tasks.

Rather than reducing the number of game features to isolate them, a better solution would be to conduct comparative studies using two versions of a game that keep all features of each level the same except for one. For example, one could test versions of the same Exploration Mode level in *Homeworld Bound: Redux* with and without a time limit, or with or without a map. Construction Mode levels could be manipulated to allow unlimited rotations or not, or to change the shape of fusing areas on parts to be symmetrical or not. Such comparative studies focusing on the relationship between player performance and spatial skill could make changes in each level of the game to allow for multiple features to be tested at once. After iterating on the game's design based on the results of these studies, comparative studies of the training effects of two versions of the same game could be conducted in which only one feature common to many levels is manipulated at a time in order to see the influence of an individual feature on training effects.

Overall, my data-driven approach to diagnosing problems with individual levels of spatial skill training games seems to be an effective way of catching problems and identifying potential fixes before (or while) a game is deployed in a controlled training study, based on my findings in Chapters 4 and 7. Catching problems before investing in a costly and time-consuming training study is clearly preferable, as it can help researchers iterate on a game's design multiple times before testing its training effects to increase its chances of being a successful training intervention. However, I still recommend that researchers analyze the relationship between spatial skill and player performance during training studies so that there is additional data to help explain potential reasons for the intervention's success (or lack thereof) as a training intervention and encourage future work aimed at establishing the accuracy of these reasons.

8.2.2 Stealth Assessment

A final area of future work utilizing my data-driven approach to analyzing the relationship between spatial skill and player behavior in games is in the domain of stealth assessment. Using an iterative design and evaluation process, researchers and game designers could revise a game until the relationship between a specific dimension of spatial skill and player performance is very strong in each level. The game could then be used as a *stealth assessment* of spatial skill by researchers and educators as an alternative to standardized paper and pencil tests of spatial skill that are widely used in the spatial skill training research literature.

Stealth assessment is designed to promote a more immersive and positive experience for the learner by placing them in a state of *flow* as described by Csikszentmihalyi - a state in which the learner becomes completely absorbed in an activity, losing track of time and engaging in the activity due to its intrinsic appeal rather than extrinsic incentives like grades or other rewards [224]. Another purpose of stealth assessment is to reduce test anxiety or remove it altogether [225], making estimates of spatial skill more accurate for those with test anxiety. In addition, by concealing that spatial skills are what is being evaluated, stealth assessment may produce more accurate estimates of spatial skill for women and other low spatial skill populations, who tend to perform worse on spatial tasks when the tasks are explicitly framed as spatial [103, 104, 107, 226, 227, 228, 106].

Stealth assessment is often implemented using Bayesian networks or other machine learning methods that can use more complex methods of recognizing learner's skill levels than simple level-by-level quantitative performance metrics, such as learning specific sequences of actions that are relevant to a learner's skill level. These more complex methods can be used to identify specific student misconceptions for the purposes of providing targeted feedback or tailored content [225]. I encourage future work exploring how to implement stealth assessment in *Homeworld Bound: Redux* and other spatial skill training games.

8.3 A PLAYER-CENTRIC APPROACH TO DESIGNING AND EVALUATING SPATIAL SKILL TRAINING GAMES

In Chapter 5, I presented a player-centric approach to designing spatial skill training games. I conducted a survey study across three diverse populations of high school students and college-age adults to gain a deeper and more complex understanding of how the following demographic and prior gaming experience variables predicted pre-existing spatial skill: gender, socioeconomic status (SES), gaming habits, gaming preferences, and gaming motivations. I found that the only predictors of spatial skill were gender and population (males and students at a selective, high SES high school scored higher on a test of mental rotation than females and students at a non-selective, low SES public high school and college-age adults recruited online and around the Champaign-Urbana area). While this finding was consistent with prior work, the lack of relationship between preference for action video games and spatial skill was surprising given the consistent relationship between action gaming and spatial skill found elsewhere in the literature [112, 113, 114, 32, 115, 116, 27, 117, 143, 31].

I used the findings of this study to develop a set of design recommendations for spatial skill training games that align with the preferences of the low spatial skill populations I identified in my results (college-age women and girls from the non-selective high school): *facilitate short gameplay sessions*, *promote simple fun and thrill*, and *focus on adventure and puzzle genres*. I then continued this player-centric approach by implementing these recommendations in my own spatial skill training game, *Homeworld Bound*, to produce a revised version of the game, *Homeworld Bound: Redux*, designed to be more appealing to low spatial skill young adults (Chapter 6). When I evaluated *Homeworld Bound: Redux* in a controlled study conducted in the context of an introductory programming course, my analysis of both its training effects and its intrinsic appeal relative to other interventions focused on low spatial skill students since they were the target audience of the game. In addition, I looked at how other demographic factors such as gender and prior gaming experience influenced training effects and intrinsic appeal.

The results of the study showed that gender and prior gaming experience had no effect on training effects or intrinsic appeal. This is not surprising given that the 70 minute training period was not sufficient to produce detectable training effects for *Homeworld Bound: Redux* or either of the other interventions tested (online spatial workbook exercises and the active control group intervention, a python graphics exercise). However, I did find that low spatial skill students rated their intrinsic motivation to play *Homeworld Bound: Redux* higher than the spatial workbook exercises, indicating that the game succeeded in its goal of being more intrinsically motivating for low spatial skill students than non-game spatial skill training interventions. Of course, there is plenty of room for improvement given that low spatial skill students still rated *Homeworld Bound: Redux* a modest 4.2 (average) out of 7 on intrinsic enjoyment. Future work should conduct training studies with *Homeworld Bound: Redux* that give students the opportunity to provide not just enjoyment ratings but more detailed qualitative feedback about their experience with the game to inform future player experience improvements.

8.3.1 Working with the Target Audience

Overall, my findings point to the importance of recruiting from the correct target audience when conducting studies to inform the design of or to evaluate a game-based spatial skill training intervention. While it may be much easier to recruit convenience samples, such as students required to participate in research studies as part of their introductory psychology courses, research studies using more generic subject pools cannot tell us if the game being studied would ultimately be effective for those it is being designed for, and if the game is actually fun for them - the main advantage of using game-based training interventions. To ensure the effectiveness and intrinsic appeal of a game-based training intervention for low spatial skill students and underrepresented minorities in STEM fields, it is critical that we go out and find participants for our studies from these populations.

While public schools and even some afterschool programs have many more restrictions for how students' time is spent, many will work with researchers if researchers come prepared to invest in a relationship with the school and have something of value to offer teachers and students, such as educational content that can fit in with a school's existing curriculum. In addition, with the increasing interest in scaling university education to larger and larger class sizes, introductory STEM courses at universities provide an excellent opportunity for implementing online game-based spatial skill training interventions that have been designed, as *Homeworld Bound: Redux* was, to be easy for educators to deploy in their classrooms. Engineering programs at various universities have been implementing spatial skill training programs for first year students for a while now, some of which are standalone courses that allow for more long term training effects to be observed [128, 129, 131, 229, 215]. Thus, we should not limit ourselves to laboratory studies when designing and evaluating game-based spatial skill interventions, as the overwhelming majority of the spatial skill game training literature has, if we want to see these games eventually have a measurable long term impact on STEM proficiency and retention in the communities they are designed to serve. Future work should bring together researchers, teachers, and students in a collaborative, mutually beneficial process to design and evaluate spatial skill training games in classroom contexts with low spatial skill populations.

8.3.2 Beyond the Classroom

While classroom contexts are critical for achieving more ecological validity in spatial skill training programs aiming at improving students' STEM efficacy and retention, the ideal context for evaluating game-based training interventions is "in the wild" - in a setting where participants are given the freedom to play the game or not as they choose without any extrinsic incentives. The lack of such "in the wild" studies in the spatial skill training game literature, while understandable given the need for experimental control of some sort when evaluating training effects - should not be ignored. The entire purpose of using games as training tools instead of more traditional classroom instruction or workbook exercises is that they provide the kind of intrinsic motivation that could potentially get people to play them in their leisure time. Children spend only about 18.5% of their waking hours in school [230], so a learning activity that could take place outside of school - and that students would actually be willing to sacrifice leisure time for - would be a huge win for educators and students alike, allowing educators more time to focus on other academic subjects and students more hours in the day to practice spatial skills in a way that's more fun for them. This kind of *informal learning* [208] of spatial skills through leisure activity is currently thought to be one of the main reasons behind the gender disparity in spatial skills that emerges in childhood [100, 161, 73, 102, 231, 77], and so might provide a way to give girls equal and long term access to spatial activities that simply would not be possible with classroom interventions alone.

8.4 HOMEWORLD BOUND: REDUX AS A RESEARCH AND DESIGN TOOL FOR SPATIAL SKILL TRAINING GAMES

In Chapters 3 and 4, I described the design and evaluation of a spatial skill training game, *Homeworld Bound*, that I designed as a testbed for analyzing the effectiveness of different game features in tapping into children's spatial skills. In Chapter 6, I described the design process I used for a new version of *Homeworld Bound* I built for a target audience of high school students and college-age adults. The new version, *Homeworld Bound: Redux*, is open-source, online, and easily scalable so that it can be modified, used for future research studies, and set up in classrooms for training interventions by the larger community of cognitive training game designers, researchers, and educators. The design of *Homeworld Bound: Redux* combines the theoretically grounded, data-driven, and player-centric approaches to game design I used in Chapters 3, 4, and 5. The design of the improved spatial game features in *Homeworld Bound: Redux* was informed by my data-driven analysis of the relationship between spatial skill and performance in Chapter 4, which allowed me to identify potential reasons for certain levels' ineffectiveness at tapping into players' spatial skills. For the new version of the game, I also incorporated the game features that I found low spatial skill young adults preferred in my Chapter 5 survey study to improve the game's intrinsic appeal to this critical target audience for spatial skill training interventions.

There are many future directions for modifying and improving *Homeworld Bound: Redux* for future research studies and educational interventions. In the following sections, I describe a few ideas. I encourage any researchers, game designers, or educators who find these ideas compelling to access the game's source code at `https://github.com/hwauck/homeworld-bound` and make the suggested modifications.

8.4.1 New Game Features to Try

While *Homeworld Bound: Redux* already contains game features relevant to three dimensions of the spatial skill typology I use in this dissertation (*intrinsic-dynamic*, *extrinsic-static*, and *extrinsic-dynamic*), the game does not as of now attempt to incorporate features meant to tap into players' lower level spatial skills. One such low level spatial skill is spatial attention, which concerns the ability to attend and react to multiple moving targets in one's environment. Spatial attention is frequently required in first person shooters and other action games, and training on an action game has been shown to improve players' spatial attention skills [112, 115, 27, 3, 141]. It has been hypothesized that this lower level spatial skill may underlie all the higher level spatial skills in the typology [3, 59, 136]. Future work could investigate adding some sort of moving targets to certain *Homeworld Bound: Redux* levels that the player has to aim at or react to quickly to in order to tap into more of this low level spatial attention skill.

However, I recommend caution if implementing these kinds of features, since research has shown that those with low spatial skill are less likely to play first person shooter games and other action games frequently than those with high spatial skill [112, 113, 114, 32, 115, 116, 27, 117, 143, 31]. Conducting more player research studies with low spatial skill populations in the future could help to clarify whether there are any specific characteristics of action games that they find unappealing. If so, game features that tax players' spatial attention like an action game could be added to *Homeworld Bound: Redux*, but omitting the unappealing aspects of the genre for this target population. For instance, if low spatial skill players are not as interested in combat and shooting weapons, this could be replaced with a game where players need to aim at stimuli for a different reason, such as in the commercial game *Pokémon Snap*, a rail shooter where players take pictures of wild Pokémon while moving through a 3D environment and progress by taking higher quality pictures of their targets.

Another feature I did not analyze in *Homeworld Bound: Redux* is the social aspect of gaming. Many games are played with friends, family, and strangers on the internet. Adding multiplayer and social features, such as collaborative in-game tasks and chat capability, may increase engagement with the game for those who prefer multiplayer experiences. Social aspects of gaming also relate to the third aspect of need satisfaction in Deci and Ryan's self-determination theory: *relatedness*, or the need to feel connected to other people [61]. Future work could investigate ways of implementing multiplayer functionality in *Homeworld Bound: Redux*, either locally or online, and the effects of this *relatedness*-promoting game feature on the player experience for low spatial skill populations.

A final idea is to implement an intelligent system in *Homeworld Bound: Redux* to automatically detect a player's skill level based on their in-game behavior and adjust the game's content to tailor it to the appropriate difficulty for their skill level. Machine learning algorithms, such as the Bayesian networks used by Shute in game-based stealth assessments of various skills [225], could be employed to detect the player's level of spatial skill and present a harder or easier version of the next level based on prior patterns of behavior. This approach has been used successfully in intelligent tutoring systems [232, 233, 234, 235] and has been applied in a number of educational games as well [236, 69, 237, 176] to ensure the content is neither so difficult that learners give up nor so easy that they do not learn anything from it.

8.4.2 New Platforms to Try

I chose to implement *Homeworld Bound* and its successor, *Homeworld Bound: Redux* for personal computers due to their general availability in schools, public libraries, and homes. However, with mobile devices taking the majority of the computing market share as of 2019 [238] and the unique functionalities smartphones afford, a mobile implementation of *Homeworld Bound: Redux* could improve the game's fit with its target audience and the overall player experience in a few ways. First, more natural gesture controls could be used for Construction Mode, such as tapping and dragging to move parts and pinching in or out to change the zoom level of the camera. In addition, the game could take advantage of the gyrometers and accelerometers already present in mobile devices to allow players to rotate their device in order to rotate a part on the screen. This more embodied, spatial gesture-based control, when used in virtual environments and simulations, has been shown to improve students' mental models and general understanding of science concepts [239, 240, 241] and therefore may benefit the development of spatial skills as well.

Furthermore, children are now more familiar with smartphones than computers based on my observations during my classroom studies, so a mobile version of the game may be easier for them to learn the controls for than a game involving mouse and keyboard. And, of course, mobile phones are more portable, enabling people to play anywhere when they have a spare moment. In my online survey study in Chapter 5, I found that pastime (playing just to pass the time) was one of the stronger motivations for play for low spatial skill young adults, and low spatial skill adults also reported playing in relatively short sessions, most frequently between 15-60 minutes. Playing the game on a phone allows low spatial skill players to play to pass the time, such as on the bus, train, or plane, or in the waiting room for a doctor's appointment - situations in where shorter play sessions are the norm and in fact required. Future work should investigate the extent to which a mobile version of *Homeworld Bound:*

Redux could provide a better player experience for children and low spatial skill young adults and facilitate "in the wild" gameplay more easily.

Augmented reality (AR) and virtual reality (VR) simulations and games have recently garnered a lot of attention in the education research community [242, 243, 244], and present a method for creating spatial skill training interventions that are more immersive, more embodied through the use of gesture-based controls, and more closely align with how a person perceives the world around them spatially. Embodied controls and hand gestures have demonstrated success at improving students' mental models in science domains by helping students concretize abstract concepts [239] and make sense of them [240, 241]. It would be interesting to see if a version of *Homeworld Bound* implemented in either AR or VR, perhaps where players could manipulate 3D virtual representations of object parts in Construction Mode directly with their own hands, might be more engaging for players or more effective at training spatial skills.

The only two noncommercial spatial skill training games that have been studied thus far in the research literature, apart from *Homeworld Bound: Redux*, are implemented in either VR [153] or use embodied controls [152], and show some preliminary evidence of success as training interventions relative to non-immersive interventions. In addition, VR and AR software have been studied in combination with workbook exercises and sketching as a way to train engineering students' *intrinsic-dynamic* and *extrinsic-static* spatial skills, such as mental rotation, spatial visualization, and spatial orientation [133]. Unfortunately, VR and AR technology is not as scalable as online or mobile game interventions for large classrooms due to the need for specialized hardware and software for each student. Therefore, I recommend that future work focuses primarily on implementing *Homeworld Bound: Redux* in ways that allow the game to be more accessible to public schools and more scalable to large classrooms or in the wild studies. However, with the advent of more affordable and mobile phone-compatible hardware for VR like Google Cardboard, VR-based, gesture-controlled implementations of spatial skill training games like *Homeworld Bound: Redux* are still worth considering as a potential future work direction.

8.4.3 Children as a Target Audience

While the latter portion of this dissertation focused on designing and evaluating a spatial skill training game for young adults (Chapters 5, 6, and 7), children are still a very important target population for spatial skill training interventions. Training children's spatial skills before low spatial skill becomes a barrier to the pursuit of STEM in secondary and tertiary education is the ideal for spatial skill training interventions since performance and experience in early STEM courses in middle school or high school predicts students' decision of whether or not to pursue STEM majors and careers [245, 246, 247], especially among women [248]. To facilitate future work investigating the design of spatial skill training games for children, *Homeworld Bound: Redux* could be modified to make an easier version suitable for children. For example, in Construction Mode, less complex shapes, auto-alignment of selected fuse areas (as in the final version of *Homeworld Bound*), and increasing the number of rotations allowed per level could make levels easier. In Exploration Mode, giving more time on the timed levels and possibly making the level map draw more attention to key landmarks could make levels easier for children. I invite future work on creating an easier version of *Homeworld Bound: Redux* for children to enable the continuation of the line of research on designing spatial skill training games for children I started in Chapters 3 and 5.

8.5 CONCLUSION

This dissertation has made several contributions to the research literature aimed at combining the approaches of the psychology, education, and human-computer interaction domains to produce novel insights about best practices for designing spatial skill training games to improve efficacy and retention in STEM majors and careers. First, I presented a theoretically grounded approach to developing games and game features that tap into players' spatial skills and implemented these features in a functional, engaging children's computer game that has demonstrated potential to train spatial skills: *Homeworld Bound*. Using *Homeworld Bound* as a testbed for a data-driven approach to evaluating specific game features, I presented the first empirical study analyzing the potential of different game features to tap into children's spatial skills and gave practical recommendations for implementing these features in games to assess or train players' spatial skills.

I then turned to a player-centric approach to the design of spatial skill training games, looking at what game features would promote a more positive, more intrinsically motivating player experience among the important target audience of low spatial skill populations. This investigation contributed to the spatial skill game training research literature via a deeper and more complex understanding of the relationship between pre-existing spatial skill and video game play habits, preferences, and underlying gaming motivations and a set of design recommendations for spatial skill training games that align with the preferences of low spatial skill populations of high school students and college-age adults from diverse backgrounds.

Combining the results of my data-driven and player-centric approaches, I redesigned *Homeworld Bound* to improve its ability to tap into players' spatial skills and its motivational appeal for low spatial skill young adults. In doing so, I demonstrated how to implement the spatial features derived from my theoretical framework and the player experience features I recommended for increasing the game's appeal to low spatial skill young adults in a real game, *Homeworld Bound: Redux*. *Homeworld Bound: Redux* is an open-source and free-to-use spatial skill training game scalable to large class sizes that can be modified, used for future research studies, and set up in classrooms by the larger community of cognitive training game designers, researchers, and teachers.

I evaluated *Homeworld Bound: Redux* in the context of a large introductory STEM course, contributing the first study evaluating the spatial skill training effectiveness of a non-commercial game that 1) analyzed not just training effects but also the intrinsic appeal of the intervention for students, 2) focused analysis of both training effects and intrinsic appeal on low spatial skill students, and 3) used a sample size large enough to detect differences in training effects reliably. In addition, my evaluation of *Homeworld Bound: Redux* in a classroom setting contributes the most detailed analysis to date of the effectiveness of multiple types of game features at tapping in to players' spatial skills, providing researchers and game designers with insights that can be used to increase a spatial skill training game's likelihood of actually achieving training effects and deepen the research community's theoretical understanding of the mechanisms behind spatial skill use in digital games.

These contributions have broader impacts beyond the research literature. Understanding not just which game-based spatial skill training interventions work, but why, can help researchers and game designers design interventions that are more likely to be effective. In addition, focusing the design of spatial skill training games on low spatial skill populations as the target audience and understanding their gaming preferences can help researchers and game designers create interventions that are more likely to interest them enough to persist in playing, achieve noticeable training effects, and increase their chances of succeeding in and pursuing STEM majors and careers.

While *Homeworld Bound: Redux* was designed as a testbed for investigating how to make spatial skill training games more effective, it is my hope that with future revisions and evaluation studies, the game can also serve as an effective, accessible, and scalable training intervention to be used by schools, teachers, and students to improve access to STEM someday. However, I hope *Homeworld Bound: Redux* can be more than a training intervention as well: that it can serve a more general purpose as a tool for researchers, game designers, and educators around the world to investigate the best ways to help students acquire the spatial skills necessary to succeed in STEM coursework and careers using game-based interventions.

References

[1] D. Uttal and C. Cohen, "Spatial Thinking and STEM Education: When, Why, and How?" in *The Psychology of Learning and Motivation.* Elsevier, 2012.

[2] J. Wai, D. Lubinski, and C. P. Benbow, "Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance." *Journal of Educational Psychology*, vol. 101, no. 4, pp. 817–835, 2009. [Online]. Available: http://doi.apa.org/getdoi.cfm?doi=10.1037/a0016127

[3] J. Feng, I. Spence, and J. Pratt, "Playing an Action Video Game Reduces Gender Differences in Spatial Cognition," *Psychological Science*, vol. 18, no. 10, pp. 850–855, oct 2007. [Online]. Available: http://pss.sagepub.com/lookup/doi/10.1111/j.1467-9280.2007.01990.x

[4] M. S. Terlecki, N. S. Newcombe, and M. Little, "Durable and generalized effects of spatial experience on mental rotation: gender differences in growth patterns," *Applied Cognitive Psychology*, vol. 22, no. 7, pp. 996–1013, nov 2008. [Online]. Available: http://doi.wiley.com/10.1002/acp.1420

[5] V. J. Shute, M. Ventura, and F. Ke, "The power of play: The effects of Portal 2 and Lumosity on cognitive and noncognitive skills," *Computers and Education*, vol. 80, 2015.

[6] S. Fayer, A. Lacey, and A. Watson, "BLS Spotlight on Statistics: STEM Occupations - Past, Present, and Future," Bureau of Labor Statistics, Washington, D.C., Tech. Rep., jan 2017.

[7] Y. Xue and R. C. Larson, "STEM Crisis or STEM Surplus: Yes and Yes," *U.S. Bureau of Labor Statistics Monthly Labor Review*, vol. 138, 2015. [Online]. Available: http://heinonline.org/HOL/Page?handle=hein.journals/month138{&}id=268{&}div={&}collection=

[8] L. S. Liben, "The STEM Gender Gap: The Case for Spatial Interventions," *International Journal of Gender, Science and Technology*, vol. 7, no. 2, pp. 133–150, 2015.

[9] L. S. Liben and E. F. Coyle, "Developmental Interventions to Address the STEM Gender Gap: Exploring Intended and Unintended Consequences," *Advances in Child Development and Behavior*, vol. 47, pp. 77–115, jan 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0065240714000111

[10] D. H. Uttal, N. G. Meadow, E. Tipton, L. L. Hand, A. R. Alden, C. Warren, and N. S. Newcombe, "The Malleability of Spatial Skills: A Meta-Analysis of Training Studies." *Psychological Bulletin*, vol. 139, no. 2, pp. 352–402, mar 2012. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/22663761

[11] D. L. Shea, D. Lubinski, and C. P. Benbow, “Importance of assessing spatial ability in intellectually talented young adolescents: A 20-year longitudinal study.” *Journal of Educational Psychology*, vol. 93, no. 3, pp. 604–614, 2001. [Online]. Available: http://doi.apa.org/getdoi.cfm?doi=10.1037/0022-0663.93.3.604

[12] Y. Kali and N. Orion, “Spatial Abilities of High-School Students in the Perception of Geologic Structures,” *Journal of Research in Science Teaching*, vol. 33, no. 4, pp. 369–391, 1996.

[13] N. Orion, D. Ben-Chaim, and Y. Kali, “Relationship Between Earth-Science Education and Spatial Visualization,” *Journal of Geoscience Education*, vol. 45, no. 2, pp. 129–132, mar 1997. [Online]. Available: http://nagt-jge.org/doi/10.5408/1089-9995-45.2.129

[14] R. T. Duesbury and H. F. O’Neil, “Effect of type of practice in a computer-aided design environment in visualizing three-dimensional objects from two-dimensional orthographic projections,” *Journal of Applied Psychology*, vol. 81, no. 3, pp. 249–260, 1996. [Online]. Available: http://doi.apa.org/getdoi.cfm?doi=10.1037/0021-9010.81.3.249

[15] S. Hsi, M. C. Linn, and J. E. Bell, “Role of spatial reasoning in engineering and the design of spatial instruction,” *Journal of Engineering Education*, vol. 86, no. 2, pp. 151–158, 1997.

[16] H. B. P. Gerson, S. A. Sorby, A. Wysocki, and B. J. Baartmans, “The development and assessment of multimedia software for improving 3-D spatial visualization skills,” *Computer Applications in Engineering Education*, vol. 9, no. 2, pp. 105–113, jan 2001. [Online]. Available: http://doi.wiley.com/10.1002/cae.1012

[17] S. P. Lajoie, “Individual Differences in Spatial Ability: Developing Technologies to Increase Strategy Awareness and Skills,” *Educational Psychologist*, vol. 38, no. 2, pp. 115–125, jun 2003. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1207/S15326985EP3802{_}6

[18] M. Hegarty, M. Keehner, C. Cohen, D. R. Montello, and Y. Lippa, “The Role of Spatial Cognition in Medicine: Applications for Selecting and Training Professionals,” in *Applied spatial cognition: from research to technology*, G. L. Allen, Ed. Mahwah, NJ: Lawrence Erlbaum Associates, 2007.

[19] S. J. Jones and G. E. Burnett, “Spatial skills and navigation of source code,” *ACM SIGCSE Bulletin*, 2007.

[20] J. Parkinson and Q. Cutts, “Investigating the relationship between spatial skills and computer science,” in *ICER 2018 - Proceedings of the 2018 ACM Conference on International Computing Education Research*, 2018.

[21] M. C. Linn and A. C. Petersen, "Emergence and Characterization of Sex Differences in Spatial Ability: A Meta-Analysis," *Child Development*, vol. 56, no. 6, p. 1479, dec 1985. [Online]. Available: http://www.jstor.org/stable/1130467?origin=crossref

[22] Y. Maeda and S. Y. Yoon, "A Meta-Analysis on Gender Differences in Mental Rotation Ability Measured by the Purdue Spatial Visualization Tests: Visualization of Rotations (PSVT:R)," *Educational Psychology Review*, vol. 25, no. 1, pp. 69–94, 2013.

[23] S. Neuburger, P. Jansen, M. Heil, and C. Quaiser-Pohl, "Gender differences in pre-adolescents' mental-rotation performance: Do they depend on grade and stimulus type?" *Personality and Individual Differences*, vol. 50, no. 8, pp. 1238–1242, jun 2011. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0191886911000924

[24] T. W. Malone and M. R. Lepper, "Making Learning Fun: A Taxonomy of Intrinsic Motivations for Learning," *Aptitude, learning, and instruction*, vol. 3, pp. 223–253, 1987.

[25] R. M. Ryan, C. S. Rigby, and A. Przybylski, "The Motivational Pull of Video Games: A Self-Determination Theory Approach," *Motivation and Emotion*, vol. 30, no. 4, pp. 344–360, dec 2006. [Online]. Available: http://link.springer.com/10.1007/s11031-006-9051-8

[26] "Parenting in America: Outlook, worries, aspirations are strongly linked to financial situation," Pew Research Center, Tech. Rep., 2015. [Online]. Available: www.pewresearch.org

[27] C. Green and D. Bavelier, "Action-Video-Game Experience Alters the Spatial Resolution of Vision," *Psychological Science*, vol. 18, no. 1, pp. 88–94, jan 2007. [Online]. Available: http://pss.sagepub.com/lookup/doi/10.1111/j.1467-9280.2007.01853.x

[28] S. Kühn, T. Gleich, R. C. Lorenz, U. Lindenberger, and J. Gallinat, "Playing Super Mario induces structural brain plasticity: gray matter changes resulting from training with a commercial video game." *Molecular psychiatry*, vol. 19, no. August 2013, pp. 265–71, 2014. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/24166407

[29] K. Subrahmanyam and P. M. Greenfield, "Effect of video game practice on spatial skills in girls and boys," *Journal of Applied Developmental Psychology*, vol. 15, no. 1, pp. 13–32, jan 1994. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/0193397394900043

[30] I. D. Cherney, K. Bersted, and J. Smetter, "Training Spatial Skills in Men and Women," *Perceptual and Motor Skills*, vol. 119, no. 1, pp. 82–99, aug 2014. [Online]. Available: http://journals.sagepub.com/doi/10.2466/23.25.PMS.119c12z0

[31] G. Sala, K. S. Tatlidil, and F. Gobet, "Video Game Training Does Not Enhance Cognitive Ability: A Comprehensive Meta-Analytic Investigation," *Psychological Bulletin*, dec 2017. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/29239631

[32] C. Quaiser-Pohl, C. Geiser, and W. Lehmann, "The relationship between computer-game preference, gender, and mental-rotation ability," *Personality and Individual Differences*, vol. 40, no. 3, pp. 609–619, 2006.

[33] M. H. Phan, J. R. Jardina, S. Hoyle, and B. S. Chaparro, "Examining the Role of Gender in Video Game Usage, Preference, and Behavior," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 56, no. 1, pp. 1496–1500, sep 2012. [Online]. Available: http://pro.sagepub.com/lookup/doi/10.1177/1071181312561297

[34] N. Yee, "Beyond 50/50: Breaking Down The Percentage of Female Gamers By Genre," 2017. [Online]. Available: http://quanticfoundry.com/2017/01/19/female-gamers-by-genre/

[35] S. Barab and K. Squire, "Design-Based Research: Putting a Stake in the Ground," *The Journal of the Learning Sciences*, vol. 13, no. 1, pp. 1–14, 2004. [Online]. Available: https://www.tandfonline.com/action/journalInformation?journalCode=hlns20

[36] F. Wang and M. J. Hannafin, "Design-based research and technology-enhanced learning environments," pp. 5–23, 2005.

[37] P. Bell, C. M. Hoadley, and M. C. Linn, "Design-Based Research in Education," in *Internet Environments for Science Education*. Routledge, jul 2013, ch. 4, pp. 101–114.

[38] J. Nielsen, "Iterative User-Interface Design," *Computer*, vol. 26, no. 11, pp. 32–41, 1993.

[39] P. Cobb, J. Confrey, A. DiSessa, R. Lehrer, and L. Schauble, "Design Experiments in Educational Research," *Educational Researcher*, vol. 32, no. 1, pp. 9–13, jan 2003. [Online]. Available: http://journals.sagepub.com/doi/10.3102/0013189X032001009

[40] H. Wauck, Z. Xiao, P.-T. Chiu, and W.-T. Fu, "Untangling the Relationship Between Spatial Skills, Game Features, and Gender in a Video Game," in *International Conference on Intelligent User Interfaces, Proceedings IUI*, 2017.

[41] H. Wauck, E. D. Mekler, and W.-T. Fu, "A Player-Centric Approach to Designing Spatial Skill Training Games," in *CHI Conference on Human Factors in Computing Systems Proceedings*. Glasgow: Association for Computing Machinery, 2019, p. 13.

[42] H. Wauck and B. Bailey, "A testbed for fun and effective features in spatial skill training games," in *CHI PLAY 2019 - Extended Abstracts of the Annual Symposium on Computer-Human Interaction in Play*. New York, New York, USA: Association for Computing Machinery, Inc, oct 2019. [Online]. Available: http://dl.acm.org/citation.cfm?doid=3341215.3356298 pp. 763–771.

[43] R. Gorska and S. Sorby, "Testing instruments for the assessment of 3-D spatial skills," in *ASEE Annual Conference and Exposition, Conference Proceedings*, 2008.

[44] M. G. McGee, "Human spatial abilities: Psychometric studies and environmental, genetic, hormonal, and neurological influences," *Psychological Bulletin*, 1979.

[45] A. Chatterjee, "The neural organization of spatial thought and language," *Seminars in Speech and Language*, 2008.

[46] N. S. Newcombe and T. F. Shipley, "Thinking About Spatial Thinking: New Typology, New Assessments," in *Studying Visual and Spatial Reasoning for Design Creativity.* Dordrecht: Springer Netherlands, 2015, pp. 179–192. [Online]. Available: http://link.springer.com/10.1007/978-94-017-9297-4{_}10

[47] M. Kozhevnikov and M. Hegarty, "A dissociation between object manipulation spatial ability and spatial orientation ability," *Memory & Cognition*, vol. 29, no. 5, pp. 745–756, jul 2001. [Online]. Available: http://www.springerlink.com/index/10.3758/BF03200477

[48] J. Huttenlocher and C. C. Presson, "Mental rotation and the perspective problem," *Cognitive Psychology*, 1973.

[49] M. Kozhevnikov, M. A. Motes, B. Rasch, and O. Blajenkova, "Perspective-taking vs. mental rotation transformations and how they predict spatial navigation performance," *Applied Cognitive Psychology*, vol. 20, no. 3, pp. 397–417, apr 2006. [Online]. Available: http://doi.wiley.com/10.1002/acp.1192

[50] M. Hegarty, D. R. Montello, A. E. Richardson, T. Ishikawa, and K. Lovelace, "Spatial abilities at different scales: Individual differences in aptitude-test performance and spatial-layout learning," *Intelligence*, 2006.

[51] M. Kozhevnikov, M. Hegarty, and R. Mayer, "Spatial Abilities in Problem Solving in Kinematics," in *Diagrammatic Representation and Reasoning*, 2002.

[52] M. Kozhevnikov, S. Kosslyn, and J. Shephard, "Spatial versus object visualizers: A new characterization of visual cognitive style," *Memory and Cognition*, 2005.

[53] J. Russell-Gebbett, "Skills and strategiespupils' approaches to three-dimensional problems in biology," *Journal of Biological Education*, vol. 19, no. 4, pp. 293–298, dec 1985. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/00219266.1985.9654755

[54] K. Rochford, "Spatial learning disabilities and underachievement among university anatomy students," *Medical Education*, vol. 19, no. 1, pp. 13–26, jan 1985. [Online]. Available: http://doi.wiley.com/10.1111/j.1365-2923.1985.tb01134.x

[55] M. Hegarty, M. Keehner, P. Khooshabeh, and D. R. Montello, "How spatial abilities enhance, and are enhanced by, dental education," *Learning and Individual Differences*, vol. 19, no. 1, pp. 61–70, jan 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1041608008000289

[56] D. Z. Hambrick and E. J. Meinz, "Limits on the predictive power of domain-specific experience and knowledge in skilled performance," *Current Directions in Psychological Science*, 2011.

[57] D. Z. Hambrick, J. C. Libarkin, H. L. Petcovic, K. M. Baker, J. Elkins, C. N. Callahan, S. P. Turner, T. A. Rench, and N. D. LaDue, "A test of the circumvention-of-limits hypothesis in scientific problem solving: The case of geological bedrock mapping." *Journal of Experimental Psychology: General*, vol. 141, no. 3, pp. 397–403, 2012. [Online]. Available: http://doi.apa.org/getdoi.cfm?doi=10.1037/a0025927

[58] M. Stieff, "Mental rotation and diagrammatic reasoning in science," *Learning and Instruction*, vol. 17, no. 2, pp. 219–234, apr 2007. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0959475207000102

[59] I. Spence and J. Feng, "Video games and spatial cognition." *Review of General Psychology*, vol. 14, no. 2, pp. 92–104, 2010. [Online]. Available: http://doi.apa.org/getdoi.cfm?doi=10.1037/a0019491

[60] L. E. Margulieux, "Spatial encoding strategy theory: the relationship between spatial skill and STEM achievement," in *ICER 2019 - Proceedings of the 2019 ACM Conference on International Computing Education Research*, 2019.

[61] E. L. Deci and R. M. Ryan, *Intrinsic Motivation and Self-Determination in Human Behavior.* Boston, MA: Springer US, 1985. [Online]. Available: http://link.springer.com/10.1007/978-1-4899-2271-7

[62] R. M. Ryan and E. L. Deci, "Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being." *American Psychologist*, vol. 55, no. 1, pp. 68–78, 2000. [Online]. Available: http://doi.apa.org/getdoi.cfm?doi=10.1037/0003-066X.55.1.68

[63] E. L. Deci, *Intrinsic Motivation*, 1st ed. Boston, MA: Springer US, 1975. [Online]. Available: http://link.springer.com/10.1007/978-1-4613-4446-9

[64] R. J. Vallerand and G. Reid, "On the Causal Effects of Perceived Competence on Intrinsic Motivation: A Test of Cognitive Evaluation Theory," *Journal of Sport Psychology*, vol. 6, no. 1, pp. 94–102, mar 1984. [Online]. Available: http://journals.humankinetics.com/doi/10.1123/jsp.6.1.94

[65] E. L. Deci, R. Koestner, and R. M. Ryan, "A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation." *Psychological Bulletin*, vol. 125, no. 6, pp. 627–668, 1999. [Online]. Available: http://doi.apa.org/getdoi.cfm?doi=10.1037/0033-2909.125.6.627

[66] N. J. Whitton, "An investigation into the potential of collaborative computer game-based learning in Higher Education," Ph.D. dissertation, Edinburgh Napier University, 2007.

[67] L. Alfieri, P. J. Brooks, N. J. Aldrich, and H. R. Tenenbaum, "Does discovery-based instruction enhance learning?" *Journal of Educational Psychology*, vol. 103, no. 1, pp. 1–18, 2011. [Online]. Available: http://doi.apa.org/getdoi.cfm?doi=10.1037/a0021017

[68] K. R. Sawyer, *The Cambridge Handbook of the Learning Sciences*, K. R. Sawyer, Ed. Cambridge University Press, 2005.

[69] C. Conati, N. Jaques, and M. Muir, "Understanding attention to adaptive hints in educational games: An eye-tracking study," *International Journal of Artificial Intelligence in Education*, vol. 23, no. 1-4, pp. 136–161, 2013.

[70] H. Wauck and W.-T. Fu, "A Data-Driven, Multidimensional Approach to Hint Design in Video Games," in *International Conference on Intelligent User Interfaces, Proceedings IUI*, 2017.

[71] M. C. Medlock, "Using the RITE method to improve products; a definition and a case study."

[72] E. Andersen, E. O'Rourke, Y.-E. Liu, R. Snider, J. Lowdermilk, D. Truong, S. Cooper, and Z. Popovic, "The impact of tutorials on games of varying complexity," in *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12*. New York, New York, USA: ACM Press, 2012. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2207676.2207687 p. 59.

[73] D. Voyer, C. Nolan, and S. Voyer, "The Relation Between Experience and Spatial Performance in Men and Women," *Sex Roles*, vol. 43, no. 11/12, pp. 891–915, 2000. [Online]. Available: http://link.springer.com/10.1023/A:1011041006679

[74] P. Sweetser and D. Johnson, "Player-Centered Game Environments: Assessing Player Opinions, Experiences, and Issues," in *International Conference on Entertainment Computing*. Springer, 2004. [Online]. Available: http://link.springer.com/10.1007/978-3-540-28643-1{_}40 pp. 321–332.

[75] V. Vanden Abeele, B. De Schutter, L. Geurts, S. Desmet, J. Wauters, J. Husson, L. Van den Audenaeren, F. Van Broeckhoven, J.-H. Annema, and D. Geerts, "P-III: A Player-Centered, Iterative, Interdisciplinary and Integrated Framework for Serious Game Design and Development," in *Serious Games: The Challenge*. Springer, oct 2012, pp. 82–86. [Online]. Available: http://link.springer.com/10.1007/978-3-642-33814-4{_}14

[76] D. Voyer, S. Voyer, and M. P. Bryden, "Magnitude of sex differences in spatial abilities: a meta-analysis and consideration of critical variables." *Psychological Bulletin*, vol. 117, no. 2, pp. 250–270, 1995.

[77] C. A. Lawton, "Gender, Spatial Abilities, and Wayfinding," in *Handbook of Gender Research in Psychology*, 2010.

[78] S. C. Levine, J. Huttenlocher, A. Taylor, and A. Langrock, "Early sex differences in spatial skill." *Developmental psychology*, vol. 35, no. 4, pp. 940–949, 1999.

[79] E. Hampson and D. Kimura, "Reciprocal Effects of Hormonal Fluctuations on Human Motor and Perceptual-Spatial Skills," *Behavioral Neuroscience*, vol. 102, no. 3, pp. 456–459, jun 1988.

[80] E. Hampson, "Estrogen-related variations in human spatial and articulatory-motor skills," *Psychoneuroendocrinology*, vol. 15, no. 2, pp. 97–111, jan 1990.

[81] E. Hampson, "Variations in sex-related cognitive abilities across the menstrual cycle," *Brain and Cognition*, vol. 14, no. 1, pp. 26–43, sep 1990.

[82] M. S. Moody, "Changes in scores on the Mental Rotations Test during the menstrual cycle." *Perceptual and motor skills*, vol. 84, no. 3 Pt 1, pp. 955–61, jun 1997. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/9172209

[83] A. Postma, J. Winkel, A. Tuiten, and J. Van Honk, "Sex differences and menstrual cycle effects in human spatial memory," *Psychoneuroendocrinology*, vol. 24, no. 2, pp. 175–192, feb 1999.

[84] I. Silverman and K. Phillips, "Effects of estrogen changes during the menstrual cycle on spatial performance," *Ethology and Sociobiology*, vol. 14, no. 4, pp. 257–269, jul 1993.

[85] C. Chiarello, M. A. McMahon, and K. Schaefer, "Visual cerebral lateralization over phases of the menstrual cycle: A preliminary investigation," *Brain and Cognition*, vol. 11, no. 1, pp. 18–36, sep 1989.

[86] D. F. Halpern and U. Tan, "Stereotypes and steroids: Using a psychobiosocial model to understand cognitive sex differences," *Brain and Cognition*, vol. 45, no. 3, pp. 392–414, apr 2001.

[87] R. Halari, M. Mines, V. Kumari, R. Mehrotra, M. Wheeler, V. Ng, and T. Sharma, "Sex differences and individual differences in cognitive performance and their relationship to endogenous gonadal hormones and gonadotropins," *Behavioral Neuroscience*, 2005.

[88] R. C. Gur, D. Alsop, D. Glahn, R. Petty, C. L. Swanson, J. A. Maldjian, B. I. Turetsky, J. A. Detre, J. Gee, and R. E. Gur, "An fMRI study of sex differences in regional activation to a verbal and a spatial task," *Brain and Language*, vol. 74, no. 2, pp. 157–170, sep 2000.

[89] B. W. Johnson, K. J. McKenzie, and J. P. Hamm, "Cerebral asymmetry for mental rotation: effects of response hand, handedness and gender," *NeuroReport*, vol. 13, no. 15, pp. 1929–1932, oct 2002. [Online]. Available: http://journals.lww.com/00001756-200210280-00020

[90] K. Hugdahl, T. Thomsen, and L. Ersland, "Sex differences in visuo-spatial processing: An fMRI study of mental rotation," *Neuropsychologia*, vol. 44, no. 9, pp. 1575–1583, jan 2006.

[91] T. Thomsen, K. Hugdahl, L. Ersland, R. Barndon, A. Lundervold, A. I. Smievol, B. E. Roscher, and H. Sundberg, "Functional magnetic resonance imaging (fMRI) study of sex differences in a mental rotation task," *Medical Science Monitor*, vol. 6, no. 6, pp. 1186–1196, 2000.

[92] K. Jordan, T. Wüstenberg, H. J. Heinze, M. Peters, and L. Jäncke, "Women and men exhibit different cortical activation patterns during mental rotation tasks," *Neuropsychologia*, vol. 40, no. 13, pp. 2397–2408, jan 2002.

[93] S. L. Rilea, B. Roskos-Ewoldsen, and D. Boles, "Sex differences in spatial ability: A lateralization of function approach," *Brain and Cognition*, vol. 56, no. 3, pp. 332–343, dec 2004.

[94] D. Voyer and M. P. Bryden, "Masking and visual field effects on a lateralized rod-and-frame test." *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, vol. 47, no. 1, pp. 26–37, 1993.

[95] G. M. Alexander, M. G. Packard, and B. S. Peterson, "Sex and spatial position effects on object location memory following intentional learning of object identities," *Neuropsychologia*, vol. 40, no. 8, pp. 1516–1522, jan 2002.

[96] I. Ecuyer-Dab and M. Robert, "The female advantage in object location memory according to the foraging hypothesis: A critical analysis," *Human Nature*, vol. 18, no. 4, pp. 365–385, oct 2007.

[97] C. M. Jones, V. A. Braithwaite, and S. D. Healy, "The evolution of sex differences in spatial ability," *Behavioral Neuroscience*, vol. 117, no. 3, pp. 403–411, jun 2003.

[98] M. Robert and G. Héroux, "Visuo-spatial play experience: Forerunner of visuo-spatial achievement in preadolescent and adolescent boys and girls?" *Infant and Child Development*, vol. 13, no. 1, pp. 49–78, mar 2004. [Online]. Available: http://doi.wiley.com/10.1002/icd.336

[99] M. S. Terlecki and N. S. Newcombe, "How Important Is the Digital Divide? The Relation of Computer and Videogame Usage to Gender Differences in Mental Rotation Ability," *Sex Roles*, vol. 53, no. 5-6, pp. 433–441, sep 2005. [Online]. Available: http://link.springer.com/10.1007/s11199-005-6765-0

[100] J. S. Kuhlman and P. A. Beitel, "Videogame Experience: A Possible Explanation for Differences in Anticipation of Coincidence," *Perceptual and Motor Skills*, vol. 72, no. 2, pp. 483–488, apr 1991. [Online]. Available: http://journals.sagepub.com/doi/10.2466/pms.1991.72.2.483

[101] S. Ozel, J. Larue, and C. Molinaro, "Relation between sport activity and mental rotation: Comparison of three groups of subjects," *Perceptual and Motor Skills*, 2002.

[102] C. Quaiser-Pohl and W. Lehmann, “Girls’ spatial abilities: Charting the contributions of experiences and attitudes in different academic groups,” *British Journal of Educational Psychology*, vol. 72, no. 2, pp. 245–260, jun 2002. [Online]. Available: http://doi.wiley.com/10.1348/000709902158874

[103] M. J. Sharps, A. L. Welton, and J. L. Price, “Gender and Task in the Determination of Spatial Cognitive Performance,” *Psychology of Women Quarterly*, vol. 17, no. 1, pp. 71–83, mar 1993. [Online]. Available: http://journals.sagepub.com/doi/10.1111/j.1471-6402.1993.tb00677.x

[104] M. J. Sharps, J. L. Price, and J. K. Williams, “Spatial cognition and gender: Instructional and stimulus influences on mental image rotation performance,” *Psychology of Women Quarterly*, vol. 18, no. 3, pp. 413–425, sep 1994. [Online]. Available: http://journals.sagepub.com/doi/10.1111/j.1471-6402.1994.tb00464.x

[105] M. S. McGlone and J. Aronson, “Stereotype threat, identity salience, and spatial reasoning,” *Journal of Applied Developmental Psychology*, 2006.

[106] R. Ariel, N. A. Lembeck, S. Moffat, and C. Hertzog, “Are there sex differences in confidence and metacognitive monitoring accuracy for everyday, academic, and psychometrically measured spatial ability?” *Intelligence*, vol. 70, 2018.

[107] M. Wraga, L. Duncan, E. C. Jacobs, M. Helt, and J. Church, “Stereotype susceptibility narrows the gender gap in imagined self-rotation performance,” pp. 813–819, 2006.

[108] S. C. Levine, M. Vasilyeva, S. F. Lourenco, N. S. Newcombe, and J. Huttenlocher, “Socioeconomic status modifies the sex difference in spatial skill,” *Psychological Science*, vol. 16, no. 11, pp. 841–845, nov 2005.

[109] K. G. Noble, B. D. McCandliss, and M. J. Farah, “Socioeconomic gradients predict individual differences in neurocognitive abilities,” *Developmental Science*, vol. 10, no. 4, pp. 464–480, jul 2007. [Online]. Available: http://doi.wiley.com/10.1111/j.1467-7687.2007.00600.x

[110] M. Carr, N. Alexeev, L. Wang, N. Barned, E. Horan, and A. Reed, “The Development of Spatial Skills in Elementary School Students,” *Child Development*, vol. 89, no. 2, 2018.

[111] B. N. Verdine, R. M. Golinkoff, K. Hirsh-Pasek, N. S. Newcombe, A. T. Filipowicz, and A. Chang, “Deconstructing building blocks: preschoolers’ spatial assembly performance relates to early mathematical skills.” *Child development*, vol. 85, no. 3, pp. 1062–1076, 2014. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/24112041

[112] C. S. Green and D. Bavelier, “Action video game modifies visual selective attention,” *Nature*, vol. 423, no. 6939, pp. 534–537, may 2003. [Online]. Available: http://www.nature.com/doifinder/10.1038/nature01647

[113] S. Green and D. Bavelier, “The Cognitive Neuroscience of Video Games,” *Digital Media: Transformation in Human Communication*, pp. 1–32, 2004.

[114] A. D. Castel, J. Pratt, and E. Drummond, “The effects of action video game experience on the time course of inhibition of return and the efficiency of visual search,” *Acta Psychologica*, vol. 119, no. 2, pp. 217–230, jun 2005. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S000169180500020X

[115] C. S. Green and D. Bavelier, “Enumeration versus multiple object tracking: the case of action video game players,” *Cognition*, vol. 101, no. 1, pp. 217–245, aug 2006. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0010027705001873

[116] C. S. Green and D. Bavelier, “Effect of action video games on the spatial distribution of visuospatial attention.” *Journal of experimental psychology. Human perception and performance*, vol. 32, no. 6, pp. 1465–78, dec 2006. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/17154785

[117] M. W. Dye, C. S. Green, and D. Bavelier, “The development of attention skills in action video game players,” *Neuropsychologia*, vol. 47, no. 8-9, pp. 1780–1789, jul 2009. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0028393209000657

[118] J. Jansz and L. Martens, “Gaming at a LAN event: the social context of playing video games,” *New Media & Society*, vol. 7, no. 3, pp. 333–355, jun 2005. [Online]. Available: http://journals.sagepub.com/doi/10.1177/1461444805052280

[119] I. D. Cherney and K. London, “Gender-linked differences in the toys, television shows, computer games, and outdoor activities of 5- to 13-year-old children,” *Sex Roles*, vol. 54, no. 9-10, pp. 717–726, may 2006.

[120] C. Wilhelm, “Gender role orientation and gaming behavior revisited: examining mediated and moderated effects,” *Information, Communication & Society*, vol. 21, no. 2, pp. 224–240, feb 2018. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/1369118X.2016.1271902

[121] T. Hartmann and C. Klimmt, “Gender and Computer Games: Exploring Females’ Dislikes,” *Journal of Computer-Mediated Communication*, vol. 11, no. 4, pp. 910–931, jul 2006.

[122] L. Cohen, L. Manion, and K. Morrison, *Research Methods in Education*, 2017.

[123] M. Baenninger and N. Newcombe, “The role of experience in spatial test performance: A meta-analysis,” *Sex Roles*, vol. 20, no. 5-6, pp. 327–344, mar 1989. [Online]. Available: http://link.springer.com/10.1007/BF00287729

[124] P. M. Greenfield, C. Brannon, and D. Lohr, “Two-dimensional representation of movement through three- dimensional space: The role of video game expertise,” *Journal of Applied Developmental Psychology*, vol. 15, pp. 87–103, 1994.

[125] I. Spence, J. J. Yu, J. Feng, and J. Marshman, "Women match men when learning a spatial skill." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 35, no. 4, pp. 1097–1103, 2009. [Online]. Available: http://doi.apa.org/getdoi.cfm?doi=10.1037/a0015641

[126] I. D. Cherney, "Mom, Let Me Play More Computer Games: They Improve My Mental Rotation Skills," *Sex Roles*, vol. 59, no. 11-12, pp. 776–786, dec 2008. [Online]. Available: http://link.springer.com/10.1007/s11199-008-9498-z

[127] J. Bisoglio, T. I. Michaels, J. E. Mervis, and B. K. Ashinoff, "Cognitive enhancement through action video game training: Great expectations require greater evidence," *Frontiers in Psychology*, vol. 5, 2014.

[128] S. A. Sorby and B. J. Baartmans, "The Development and Assessment of a Course for Enhancing the 3-D Spatial Visualization Skills of First Year Engineering Students," *Journal of Engineering Education*, vol. 89, no. 3, pp. 301–307, jul 2000. [Online]. Available: http://doi.wiley.com/10.1002/j.2168-9830.2000.tb00529.x

[129] S. A. Sorby, "A Course In Spatial Visualization and its Impact on the Retention of Female Engineering Students," *Journal of Women and Minorities in Science and Engineering*, vol. 7, no. 2, p. 20, 2001.

[130] S. A. Sorby, T. Drummer, K. Hungwe, and P. Charlesworth, "Developing 3-D spatial visualization skills for non-engineering students," in *ASEE Annual Conference and Exposition, Conference Proceedings*, 2005.

[131] S. A. Sorby, "Educational Research in Developing 3D Spatial Skills for Engineering Students," *International Journal of Science Education*, vol. 31, no. 3, pp. 459–480, feb 2009. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/09500690802595839

[132] D. I. Miller and D. F. Halpern, "Can spatial training improve long-term outcomes for gifted STEM undergraduates?" *Learning and Individual Differences*, vol. 26, 2013.

[133] C. Roca-González, J. Martin-Gutierrez, M. García-Dominguez, and M. d. C. M. Carrodeguas, "Virtual technologies to develop visual-spatial ability in engineering students," *Eurasia Journal of Mathematics, Science and Technology Education*, 2017.

[134] R. M. Onyancha, M. Derov, and B. L. Kinsey, "Improvements in Spatial ability as a result of targeted training and computer-aided Design software use: analyses of object geometries and rotation types," *Journal of Engineering Education*, 2009.

[135] Y.-L. Cheng and K. S. Mix, "Spatial Training Improves Children's Mathematics Ability," *Journal of Cognition and Development*, vol. 15, no. 1, pp. 2–11, jan 2014. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/15248372.2012.725186

[136] C. S. Green and D. Bavelier, "Learning, Attentional Control, and Action Video Games," *Current Biology*, vol. 22, no. 6, pp. R197–R206, mar 2012. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0960982212001303

[137] A. C. Oei and M. D. Patterson, "Enhancing perceptual and attentional skills requires common demands between the action video games and transfer tasks," *Frontiers in Psychology*, vol. 6, p. 113, feb 2015. [Online]. Available: http://journal.frontiersin.org/Article/10.3389/fpsyg.2015.00113/abstract

[138] G. Dale and C. Shawn Green, "The Changing Face of Video Games and Video Gamers: Future Directions in the Scientific Study of Video Game Play and Cognitive Performance," *Journal of Cognitive Enhancement*, vol. 1, no. 3, pp. 280–294, sep 2017. [Online]. Available: http://link.springer.com/10.1007/s41465-017-0015-6

[139] D. Bavelier and C. S. Green, "Enhancing Attentional Control: Lessons from Action Video Games," *Neuron*, vol. 104, no. 1, pp. 147–163, 2019.

[140] G. Dale, A. Joessel, D. Bavelier, and C. S. Green, "A new look at the cognitive neuroscience of video game play," *Annals of the New York Academy of Sciences*, 2020.

[141] W. R. Boot, A. F. Kramer, D. J. Simons, M. Fabiani, and G. Gratton, "The effects of video game playing on attention, memory, and executive control," *Acta Psychologica*, vol. 129, no. 3, pp. 387–398, nov 2008. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0001691808001200

[142] K. L. Powers, P. J. Brooks, N. J. Aldrich, M. a. Palladino, and L. Alfieri, "Effects of video-game play on information processing: a meta-analytic investigation." *Psychonomic bulletin & review*, vol. 20, no. 6, 2013.

[143] B. Bediou, D. M. Adams, R. E. Mayer, E. Tipton, C. S. Green, and D. Bavelier, "Meta-Analysis of Action Video Game Impact on Perceptual, Attentional, and Cognitive Skills," *Psychological Bulletin*, 2017. [Online]. Available: http://dx.doi.org/10.1037/bul0000130

[144] E. Adams, *Fundamentals of Game Design*, 3rd ed. New Riders Publishing, 2014. [Online]. Available: https://dl.acm.org/citation.cfm?id=2544002

[145] N. Oxford, "What's the Definition of an Action Game?" 2018. [Online]. Available: https://www.lifewire.com/nintendo-action-game-1126179

[146] "TvTropes: Action Game," 2018. [Online]. Available: https://tvtropes.org/pmwiki/pmwiki.php/Main/ActionGame

[147] H. Choi and S. A. Lane, "Impact of Visuospatial Characteristics of Video Games on Improvements in Cognitive Abilities," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 57, no. 1, pp. 1735–1739, sep 2013. [Online]. Available: http://journals.sagepub.com/doi/10.1177/1541931213571387

[148] P. A. McClurg and C. Chaillé, "Computer Games: Environments for Developing Spatial Cognition?" *Journal of Educational Computing Research*, vol. 3, no. 1, pp. 1–1, oct 1990. [Online]. Available: http://jec.sagepub.com/lookup/doi/10.2190/9N5U-P3E9-R1X8-0RQM

[149] D. Gagnon, "Videogames and spatial skills: An exploratory study," *Educational Communication and Technology*, vol. 33, no. 4, pp. 263–275, 1985.

[150] C. Basak, W. R. Boot, M. W. Voss, and A. F. Kramer, "Can training in a real-time strategy video game attenuate cognitive decline in older adults?" *Psychology and Aging*, vol. 23, no. 4, pp. 765–777, 2008. [Online]. Available: http://doi.apa.org/getdoi.cfm?doi=10.1037/a0013494

[151] L. Okagaki and P. A. Frensch, "Effects of video game playing on measures of spatial performance: Gender effects in late adolescence," *Journal of Applied Developmental Psychology*, vol. 15, no. 1, pp. 33–58, jan 1994. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/0193397394900051

[152] A. Mazalek, S. Chandrasekharan, M. Nitsche, T. Welsh, P. Clifton, A. Quitmeyer, F. Peer, F. Kirschner, and D. Athreya, "I'm in the game: embodied puppet interface improves avatar control," in *Proceedings of the fifth international conference on Tangible, embedded, and embodied interaction - TEI '11.* New York, New York, USA: ACM Press, 2011. [Online]. Available: http://portal.acm.org/citation.cfm?doid=1935701.1935727 p. 129.

[153] J. S.-K. Chang, G. Yeboah, A. Doucette, P. Clifton, M. Nitsche, T. Welsh, and A. Mazalek, "Evaluating the effect of tangible virtual reality on spatial perspective taking ability," in *Proceedings of the 5th Symposium on Spatial User Interaction - SUI '17.* New York, New York, USA: ACM Press, 2017. [Online]. Available: http://dl.acm.org/citation.cfm?doid=3131277.3132171 pp. 68–77.

[154] D. Adams and R. Mayer, "Examining the Connection Between Dynamic and Static Spatial Skills and Video Game Performance," in *Proceedings of the Annual Meeting of the Cognitive Science Society*, 2012.

[155] S. de Castell, H. Larios, and J. Jenson, "Gender, videogames and navigation in virtual space," *Acta Psychologica*, 2019.

[156] Z. Xiao, H. Wauck, Z. Peng, H. Ren, L. Zhang, S. Zuo, Y. Yao, and W.-T. Fu, "Cubicle: An adaptive educational gaming platform for training spatial visualization skills," in *International Conference on Intelligent User Interfaces, Proceedings IUI*, vol. Part F1351, 2018.

[157] R. N. Shepard and J. Metzler, "Mental Rotation of Three-Dimensional Objects," *Science, New Series*, vol. 171, no. 19, pp. 701–703, 1971.

[158] V. K. Sims and R. E. Mayer, "Domain specificity of spatial expertise: the case of video game players," *Applied Cognitive Psychology*, vol. 16, no. 1, pp. 97–115, jan 2002. [Online]. Available: http://doi.wiley.com/10.1002/acp.759

[159] D. R. Montello, "Scale and multiple psychologies of space," in *European Conference on Spatial Information Theory.* Springer, Berlin, Heidelberg, 1993. [Online]. Available: http://link.springer.com/10.1007/3-540-57207-4{_}21 pp. 312–321.

[160] D. R. Montello and R. G. Golledge, "Scale and Detail in the Cognition of Geographic Information," National Center for Geographic Information and Analysis, Tech. Rep., 1998. [Online]. Available: https://escholarship.org/uc/item/1hf6d3fx

[161] M. J. Brosnan, "Spatial ability in children's play with Lego blocks," *Perceptual and Motor Skills*, 1998.

[162] S. V. Coxon, "The malleability of spatial ability under treatment of a FIRST LEGO League-based robotics simulation," *Journal for the Education of the Gifted*, vol. 35, no. 3, 2012.

[163] S. Nath and D. Szücs, "Construction play and cognitive skills associated with the development of mathematical abilities in 7-year-old children," *Learning and Instruction*, vol. 32, 2014.

[164] M. Weinberger, "The one big reason why Minecraft is both super-popular and super-good for kids," 2016. [Online]. Available: http://www.businessinsider.com/the-minecraft-generation-on-why-kids-like-it-so-much-2016-4

[165] J. Schell, *The Art of Game Design: a book of lenses.* CRC Press, 2014.

[166] S. Y. Yoon, "Revised Purdue Spatial Visualization Test: Visualization of Rotations (Revised PSVT:R)," 2011.

[167] Y. Maeda and S. Y. Yoon, "Scaling the Revised PSVT-R: Characteristics of the first year engineering students' spatial ability," in *American Society for Engineering Education (ASEE) Annual Conference and Exposition*, Vancouver, 2011. [Online]. Available: https://peer.asee.org/18522 pp. 22.1273.1 – 22.1273.19.

[168] S. Cooper, K. Wang, M. Israni, and S. Sorby, "Spatial skills training in introductory computing," in *ICER 2015 - Proceedings of the 2015 ACM Conference on International Computing Education Research*, 2015.

[169] S. Sorby, E. Nevin, A. Behan, E. Mageean, and S. Sheridan, "Spatial skills as predictors of success in first-year engineering," in *2014 IEEE Frontiers in Education Conference (FIE) Proceedings*, 2014, pp. 1–7.

[170] S. Y. Yoon and K. H. Min, "College students' performance in an introductory atmospheric science course: associations with spatial ability," *Meteorological Applications*, 2016.

[171] S. Y. Yoon and E. L. Mann, “Exploring the Spatial Ability of Undergraduate Students: Association With Gender, STEM Majors, and Gifted Program Membership,” *Gifted Child Quarterly*, vol. 61, no. 4, 2017.

[172] M. Keehner, M. Hegarty, C. Cohen, P. Khooshabeh, and D. Montello, “Spatial Reasoning With External Visualizations: What Matters Is What You See, Not Whether You Interact,” *Cognitive Science: A Multidisciplinary Journal*, vol. 32, no. 7, pp. 1099–1132, oct 2008. [Online]. Available: http://doi.wiley.com/10.1080/03640210801898177

[173] C. Albers and D. Lakens, “When power analyses based on pilot data are biased: Inaccurate effect size estimators and follow-up bias,” *Journal of Experimental Social Psychology*, vol. 74, pp. 187–195, jan 2018.

[174] K. T. D’Alonzo, “The Johnson-Neyman procedure as an alternative to ANCOVA,” *Western Journal of Nursing Research*, vol. 26, no. 7, pp. 804–812, nov 2004.

[175] A. Field, *Discovering Statistics Using IBM SPSS Statistics*, 4th ed., M. Carmichael, Ed. Sage, 2013.

[176] E. O’Rourke, K. Haimovitz, C. Ballweber, C. Dweck, and Z. Popović, “Brain points: a growth mindset incentive structure boosts persistence in an educational game,” in *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI ’14*. New York, New York, USA: ACM Press, 2014. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2556288.2557157 pp. 3339–3348.

[177] K. Dorst and N. Cross, “Creativity in the design process: Co-evolution of problem-solution,” *Design Studies*, vol. 22, no. 5, pp. 425–437, sep 2001.

[178] So Yoon Yoon, “Psychometric Properties of the Revised Purdue Spatial Visualization Tests: Visualization of Rotations (the Revised Psvt:R),” *Policy*, vol. 9, no. April 2010, pp. 2003–2006, 2008. [Online]. Available: http://docs.lib.purdue.edu/dissertations/AAI3413897/

[179] P. P. Maglio, M. J. Wenger, and A. M. Copeland, “Evidence for the role of self-priming in epistemic action: Expertise and the effective use of memory,” 2008.

[180] P. Jansen, A. Schmelter, C. Quaiser-Pohl, S. Neuburger, and M. Heil, “Mental rotation performance in primary school age children: Are there gender differences in chronometric tests?” *Cognitive Development*, vol. 28, no. 1, pp. 51–62, jan 2013.

[181] B. S. Greenberg, J. Sherry, K. Lachlan, K. Lucas, and A. Holmstrom, “Orientations to Video Games Among Gender and Age Groups,” *Simulation & Gaming*, vol. 41, no. 2, pp. 238–259, apr 2010. [Online]. Available: http://sag.sagepub.com/cgi/doi/10.1177/1046878108319930

[182] E. F. Provenzo, Jr., *Video Kids: Making sense of Nintendo.* Cambridge, MA and London, England: Harvard University Press, jan 1991. [Online]. Available: http://www.degruyter.com/view/books/harvard.9780674422483/harvard.9780674422483/harvard.9780674422483.xml

[183] J. Cassell and H. Jenkins, "Chess For Girls?: Feminism and Computer Games," in *From Barbie to Mortal Kombat: Gender and Computer Games*, J. Cassell and H. Jenkins, Eds. Cambridge, MA: The MIT Press, 1998, ch. 1, pp. 1–32.

[184] J. D. Ivory, "Still a Man's Game: Gender Representation in Online Reviews of Video Games," *Mass Communication and Society*, vol. 9, no. 1, pp. 103–114, feb 2006. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1207/s15327825mcs0901{_}6

[185] A. Canossa and A. Drachen, "Patterns of Play: Play-Personas in User-Centred Game Development," in *Proceedings of DiGRA*, 2009.

[186] A. Bartsch, "Emotional Gratification in Entertainment Experience. Why Viewers of Movies and Television Series Find it Rewarding to Experience Emotions," *Media Psychology*, vol. 15, no. 3, pp. 267–302, jul 2012. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/15213269.2012.693811

[187] N. Lazzaro, "Why We Play: Affect and the Fun of Games," in *Human-Computer Interaction: Designing for Diverse Users and Domains*, A. Sears and J. A. Jacko, Eds. Boca Raton, FL: CRC Press, 2009, ch. 10, pp. 155–175.

[188] E. A. Boyle, T. M. Connolly, T. Hainey, and J. M. Boyle, "Engagement in digital entertainment games: A systematic review," *Computers in Human Behavior*, vol. 28, no. 3, pp. 771–780, 2012.

[189] J. L. G. Sánchez, F. L. G. Vela, F. M. Simarro, and N. Padilla-Zea, "Playability: analysing user experience in video games," *Behaviour & Information Technology*, vol. 31, no. 10, pp. 1033–1054, oct 2012. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/0144929X.2012.710648

[190] M. V. Birk, D. Toker, R. L. Mandryk, and C. Conati, "Modeling Motivation in a Social Network Game Using Player-Centric Traits and Personality Traits." Springer International Publishing, 2015. [Online]. Available: http://link.springer.com/10.1007/978-3-319-20267-9{_}2 pp. 18–30.

[191] R. E. Nisbett and T. D. Wilson, "Telling more than we can know: Verbal reports on mental processes." *Psychological Review*, vol. 84, no. 3, pp. 231–259, may 1977. [Online]. Available: http://doi.apa.org/getdoi.cfm?doi=10.1037/0033-295X.84.3.231

[192] T. D. Wilson and D. T. Gilbert, "Affective Forecasting," *Current Directions in Psychological Science*, vol. 14, no. 3, pp. 131–134, jun 2005. [Online]. Available: http://journals.sagepub.com/doi/10.1111/j.0963-7214.2005.00355.x

[193] K. Reinecke and K. Z. Gajos, "LabintheWild: Conducting Large-Scale Online Experiments With Uncompensated Samples," in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing - CSCW '15*. New York, New York, USA: ACM Press, 2015. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2675133.2675246 pp. 1364–1378.

[194] M. Peters, B. Laeng, K. Latham, M. Jackson, R. Zaiyouna, and C. Richardson, "A Redrawn Vandenberg and Kuse Mental Rotations Test - Different Versions and Factors That Affect Performance," *Brain and Cognition*, vol. 28, no. 1, pp. 39–58, jun 1995. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0278262685710329

[195] F. De Grove, V. Cauberghe, and J. Van Looy, "Development and Validation of an Instrument for Measuring Individual Motives for Playing Digital Games," *Media Psychology*, vol. 19, no. 1, pp. 101–125, jan 2016. [Online]. Available: http://www.tandfonline.com/doi/full/10.1080/15213269.2014.902318

[196] F. De Grove, J. Breuer, V. Hsueh Hua Chen, T. Quandt, R. Ratan, and J. Van Looy, "Validating the Digital Games Motivation Scale for Comparative Research Between Countries," *Communication Research Reports*, vol. 34, no. 1, pp. 37–47, jan 2017. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/08824096.2016.1250070

[197] J. A. Bopp, E. D. Mekler, and K. Opwis, "Negative Emotion, Positive Experience?: Emotionally Moving Moments in Digital Games," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*. New York, New York, USA: ACM Press, 2016. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2858036.2858227 pp. 2996–3006.

[198] H. Wauck, G. Lucas, A. Shapiro, A. Feng, J. Boberg, and J. Gratch, "Analyzing the effect of avatar self-similarity on men and women in a search and rescue game," in *Conference on Human Factors in Computing Systems - Proceedings*, vol. 2018-April, 2018.

[199] "BoardGameGeek: Video Game Guide to Genres," 2018. [Online]. Available: https://boardgamegeek.com/wiki/page/Video{_}Game{_}Guide{_}to{_}Genres

[200] C. R. Wilson Van Voorhis and B. L. Morgan, "Understanding Power and Rules of Thumb for Determining Sample Sizes," *Tutorials in Quantitative Methods for Psychology*, vol. 3, no. 2, pp. 43–50, sep 2007. [Online]. Available: http://www.tqmp.org/RegularArticles/vol03-2/p043

[201] M. Böhmer, B. Hecht, J. Schöning, A. Krüger, and G. Bauer, "Falling asleep with Angry Birds, Facebook and Kindle," in *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services - MobileHCI '11*. New York, New York, USA: ACM Press, 2011. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2037373.2037383 p. 47.

[202] H. Wauck, “Game features and individual differences: What makes a spatial skill training video game effective?” in *International Conference on Intelligent User Interfaces, Proceedings IUI*, 2017.

[203] N. Unsworth, T. S. Redick, B. D. McMillan, D. Z. Hambrick, M. J. Kane, and R. W. Engle, “Is Playing Video Games Related to Cognitive Abilities?” *Psychological Science*, vol. 26, no. 6, 2015.

[204] K. S. Button, J. P. A. Ioannidis, C. Mokrysz, B. A. Nosek, J. Flint, E. S. J. Robinson, and M. R. Munafò, “Power failure: why small sample size undermines the reliability of neuroscience,” *Nature Reviews Neuroscience*, vol. 14, no. 5, pp. 365–376, may 2013. [Online]. Available: http://www.nature.com/articles/nrn3475

[205] A. R. A. Conway, M. J. Kane, M. F. Bunting, D. Z. Hambrick, O. Wilhelm, and R. W. Engle, “Working memory span tasks: A methodological review and user's guide,” *Psychonomic Bulletin & Review*, vol. 12, no. 5, pp. 769–786, oct 2005. [Online]. Available: http://www.springerlink.com/index/10.3758/BF03196772

[206] A. S. Göritz, “Using lotteries, loyalty points, and other incentives to increase participant response and completion.” in *Advanced methods for conducting online behavioral research.* American Psychological Association, mar 2010, pp. 219–233.

[207] J. J. Jirout and N. S. Newcombe, “Building Blocks for Developing Spatial Skills: Evidence From a Large, Representative U.S. Sample,” *Psychological Science*, vol. 26, no. 3, 2015.

[208] H. C. Lane, “Enhancing informal learning experiences with affect-aware technologies,” in *Handbook of Affective Computing*, R. A. Calvo, S. D'Mello, Jonathan Gratch, and A. Kappas, Eds. Oxford University Press, 2013.

[209] S. Sorby, B. Casey, N. Veurink, and A. Dulaney, “The role of spatial training in improving spatial and calculus performance in engineering students,” *Learning and Individual Differences*, vol. 26, pp. 20–29, aug 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1041608013000460

[210] S. G. Vandenberg and A. R. Kuse, “Mental rotations, a group test of three-dimensional spatial visualization,” *Perceptual and Motor Skills*, vol. 47, no. 2, pp. 599–604, dec 1978. [Online]. Available: http://journals.sagepub.com/doi/10.2466/pms.1978.47.2.599

[211] D. Kirsh and P. Maglio, “On distinguishing epistemic from pragmatic action,” *Cognitive Science*, vol. 18, no. 4, pp. 513–549, 1994.

[212] R. M. Ryan, “Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory.” *Journal of Personality and Social Psychology*, vol. 43, no. 3, pp. 450–461, 1982. [Online]. Available: http://content.apa.org/journals/psp/43/3/450

[213] E. McAuley, T. Duncan, and V. V. Tammen, "Psychometric Properties of the Intrinsic Motivation Inventory in a Competitive Sport Setting: A Confirmatory Factor Analysis," *Research Quarterly for Exercise and Sport*, vol. 60, no. 1, pp. 48–58, mar 1989. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/02701367.1989.10607413

[214] N. Tsigilis and A. Theodosiou, "Temporal Stability of the Intrinsic Motivation Inventory," *Perceptual and Motor Skills*, vol. 97, no. 1, pp. 271–280, aug 2003. [Online]. Available: http://journals.sagepub.com/doi/10.2466/pms.2003.97.1.271

[215] Z. Xiao, Y. Yao, C.-H. Yen, S. Dey, H. Wauck, J. Leake, B. Woodard, A. Wolters, and W.-T. Fu, "A scalable online platform for evaluating and training visuospatial skills of engineering students," in *ASEE Annual Conference and Exposition, Conference Proceedings*, vol. 2017-June, 2017.

[216] J. Berkowicz and A. Myers, "Spatial Skills: A Neglected Dimension of Early STEM Education - Leadership 360 - Education Week," 2017. [Online]. Available: http://blogs.edweek.org/edweek/leadership{_}360/2017/02/spatial{_}skills{_}a{_}neglected{_}dimension{_}of{_}early{_}stem{_}education.html

[217] J. Wai, D. H. Uttal, and A. E. I. (AEI), "Why Spatial Reasoning Matters for Education Policy," Tech. Rep., 2018.

[218] G. J. Pallrand and F. Seeber, "Spatial ability and achievement in introductory physics," *Journal of Research in Science Teaching*, 1984.

[219] C. S. Green and D. Bavelier, "Exercising Your Brain: A Review of Human Brain Plasticity and Training-Induced Learning," *Psychology and Aging*, 2008.

[220] S. F. Harp and R. E. Mayer, "How Seductive Details Do Their Damage: A Theory of Cognitive Interest in Science Learning," *Journal of Educational Psychology*, vol. 90, no. 3, pp. 414–434, 1998.

[221] R. Garner, M. G. Gillingham, and C. S. White, "Effects of Seductive Details on Macroprocessing and Microprocessing in Adults and Children," *Cognition and Instruction*, vol. 6, no. 1, pp. 41–57, 1989.

[222] A. K. Gardner, J. Clanton, I. I. Jabbour, L. Scott, D. J. Scott, and M. A. Russo, "Impact of seductive details on the acquisition and transfer of laparoscopic suturing skills: Emotionally interesting or cognitively taxing?" in *Surgery (United States)*, vol. 160, no. 3. Mosby Inc., sep 2016, pp. 580–585.

[223] G. D. Rey, "A review of research and a meta-analysis of the seductive detail effect," pp. 216–237, dec 2012.

[224] M. Csikszentmihalyi, "Literacy and Intrinsic Motivation," *Daedalus*, vol. 119, no. 2, pp. 115–140, 1990. [Online]. Available: http://www.jstor.org/stable/20025303

[225] V. Shute, "Stealth assessment in computer-based games to support learning," in *Computer games and instruction*, S. Tobias and J. D. Fletcher, Eds. IAP Information Age Publishing, 2011, ch. 20, pp. 503–523.

[226] A. M. Ferguson, E. A. Maloney, J. Fugelsang, and E. F. Risko, "On the relation between math and spatial ability: The case of math anxiety," *Learning and Individual Differences*, vol. 39, 2015.

[227] S. Neuburger, V. Ruthsatz, P. Jansen, and C. Quaiser-Pohl, "Can girls think spatially? Influence of implicit gender stereotype activation and rotational axis on fourth graders' mental-rotation performance," *Learning and Individual Differences*, vol. 37, 2015.

[228] M. R. Tarampi, N. Heydari, and M. Hegarty, "A Tale of Two Types of Perspective Taking," *Psychological Science*, vol. 27, no. 11, pp. 1507–1516, nov 2016. [Online]. Available: http://journals.sagepub.com/doi/10.1177/0956797616667459

[229] C. R. González, J. Martín-Gutiérrez, M. G. Domínguez, A. S. Hernanpérez, and C. M. Carrodeguas, "Improving spatial skills: An orienteering experience in real and virtual environments with first year engineering students," in *Procedia Computer Science*, 2013.

[230] P. Bell, B. Lewenstein, A. W. Shouse, and M. A. Feder, *Learning Science in Informal Environments: People, Places, and Pursuits*, P. Bell, B. Lewenstein, A. W. Shouse, and M. A. Feder, Eds. The National Academies Press, 2009. [Online]. Available: http://www.nap.edu/catalog.php?record{_}id=12190

[231] S. Ozel, J. Larue, and C. Molinaro, "Relation between sport and spatial imagery: Comparison of three groups of participants," *Journal of Psychology: Interdisciplinary and Applied*, 2004.

[232] K. Koedinger, J. Anderson, W. Hadley, and M. Mark, "Intelligent Tutoring Goes To School in the Big City," *Human-Computer Interaction Institute*, 1997. [Online]. Available: http://repository.cmu.edu/hcii/5

[233] I. Arroyo, W. Burleson, M. Tai, K. Muldner, and B. P. Woolf, "Gender differences in the use and benefit of advanced learning technologies for mathematics," *Journal of Educational Psychology November 2013*, vol. 105, no. 4, pp. 957–969, 2013.

[234] T. Lazar and I. Bratko, "Data-driven program synthesis for hint generation in programming tutors," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8474 LNCS. Springer Verlag, 2014, pp. 306–311.

[235] V. Rus, R. Banjade, N. Niraula, E. Gire, and D. Franceschetti, "A Study On Two Hint-level Policies in Conversational Intelligent Tutoring Systems." Springer Singapore, 2017, pp. 171–181. [Online]. Available: http://link.springer.com/10.1007/978-981-10-2419-1{_}24

[236] C. Conati and X. Zhao, "Building and evaluating an intelligent pedagogical agent to improve the effectiveness of an educational game," *Proceedings of the 9th international Conference on Intelligent User Interfaces*, pp. 6–13, 2004. [Online]. Available: http://portal.acm.org/citation.cfm?id=964442.964446

[237] E. O'Rourke, C. Ballweber, and Z. Popovií, "Hint systems may negatively impact performance in educational games," in *Proceedings of the first ACM conference on Learning @ scale conference - L@S '14*. New York, New York, USA: ACM Press, 2014. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2556325.2566248 pp. 51–60.

[238] "Statcounter Global Stats." [Online]. Available: https://gs.statcounter.com/platform-market-share/desktop-mobile/worldwide/

[239] I.-C. Hung, L.-I. Lin, W.-C. Fang, and N.-S. Chen, "Learning with the Body: An Embodiment-Based Learning Strategy Enhances Performance of Comprehending Fundamental Optics," *Interacting with Computers*, vol. 26, no. 4, pp. 360–371, jul 2014.

[240] R. Lindgren, M. Tscholl, S. Wang, and E. Johnson, "Enhancing learning and engagement through embodied interaction within a mixed reality simulation," *Computers and Education*, vol. 95, pp. 174–187, apr 2016.

[241] N. Mathayas, D. E. Brown, R. C. Wallon, and R. Lindgren, "Representational gesturing as an epistemic tool for the development of mechanistic explanatory models," *Science Education*, vol. 103, no. 4, pp. 1047–1079, jul 2019. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/sce.21516

[242] M. Billinghurst and A. Dünser, "Augmented reality in the classroom," *Computer*, vol. 45, no. 7, pp. 56–63, 2012.

[243] S. Cuendet, Q. Bonnard, S. Do-Lenh, and P. Dillenbourg, "Designing augmented reality for the classroom," *Computers and Education*, vol. 68, pp. 557–569, oct 2013.

[244] J. N. Oigara, "Integrating Virtual Reality Tools Into Classroom Instruction," in *Virtual Reality in Education*, 2019.

[245] E. J. Shaw and S. Barbuti, "Patterns of Persistence in Intended College Major with a Focus on STEM Majors," *NACADA Journal*, 2010.

[246] X. Wang, "Why Students Choose STEM Majors: Motivation, High School Learning, and Postsecondary Context of Support," *American Educational Research Journal*, vol. 50, no. 5, pp. 1081–1121, oct 2013. [Online]. Available: http://aer.sagepub.com/cgi/doi/10.3102/0002831213488622

[247] E. Lichtenberger and C. George-Jackson, "Predicting High School Students' Interest in Majoring in a STEM Field: Insight into High School Students' Postsecondary Plans," *Journal of Career and Technical Education*, 2012.

[248] C. A. Shapiro and L. J. Sax, “Major selection and persistence for women in STEM,” *New Directions for Institutional Research*, 2011.