A Data-Driven, Player-Centric Approach to Evaluating Spatial Skill Training Games

Helen Wauck
SIFT, LLC
Minneapolis, MN
hwauck@sift.net

Brian S. Woodard
University of Illinois at Urbana-Champaign
Urbana, IL
bswoodrd@illinois.edu

Ziang Xiao
University of Illinois at Urbana-Champaign
Urbana, IL
zxiao5@illinois.edu

Tiffany Wenting Li
University of Illinois at Urbana-Champaign
Urbana, IL
wenting7@illinois.edu

Brian P. Bailey
University of Illinois at Urbana-Champaign
Urbana, IL
bpbailey@illinois.edu

ABSTRACT

Certain commercial games train spatial skills, a subset of cognitive skills used in school, the workplace, and everyday life. However, it is difficult to design spatial skill training games without knowing why some games are effective. In addition, existing game training studies do not analyze motivational factors for critical target populations. We conducted a study comparing spatial skill training effects and enjoyment for low spatial skill students on three training interventions: the computer game Homeworld Bound: Redux, digital spatial workbook exercises, and an active control group. We found no training effects for any intervention, but performance in certain game levels predicted spatial skill, and low spatial skill students enjoyed the game more than the workbook exercises. We provide three design recommendations for spatial skill training games based on our findings: use asymmetry for object manipulation tasks, require explicit 2D to 3D representation translation, and employ time pressure for navigation tasks.

CCS CONCEPTS

• **Human-centered computing** → **User studies**; **Usability testing**; **Empirical studies in HCI**; • **Applied computing** → *Interactive learning environments.*

KEYWORDS

cognitive training games; intrinsic motivation; spatial skills; player experience; STEM education

ACM Reference Format:
Helen Wauck, Brian S. Woodard, Ziang Xiao, Tiffany Wenting Li, and Brian P. Bailey. 2020. A Data-Driven, Player-Centric Approach to Evaluating Spatial Skill Training Games. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '20), November 2–4, 2020, Virtual Event, Canada.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3410404.3414255

CHI PLAY '20, November 2–4, 2020, Virtual Event, Canada

ACM ISBN 978-1-4503-8074-4/20/11…$15.00
https://doi.org/10.1145/3410404.3414255

1 INTRODUCTION

Spatial skills are a subset of cognitive skills essential to performing a wide range of tasks in everyday life. They determine an individual's ability to understand the spatial relationships both between and within objects in their environment [74] and are essential for remembering important locations like stores, schools, and residences and routes to and from them. Spatial skills are also required for policy decisions at all levels of government, such as determining borders, organization of infrastructure, and allocation of funds for public transit systems [22]. A wide variety of careers require the use of spatial skills; they are one of the strongest predictors of success in STEM majors and careers independent of math and verbal ability [77], and are essential in non-STEM disciplines like art and design as well. More broadly, spatial skills are necessary for the interpretation of diagrams, maps, graphs, or other visualizations that aid in visual problem-solving [50].

Despite their importance in everyday life and many career paths, spatial skills have been largely neglected in the development of standardized school curriculae [15, 78]. However, video games have shown promise as tools for training players' spatial skills. A variety of commercial video games have demonstrated empirical effectiveness in spatial skill training, including Medal of Honor: Pacific Assault [18, 20], Super Mario 64 [38, 52], Marble Madness [3, 69], and Tetris [55, 57, 71]. However, other commercial games, such as the cognitive training game *Lumosity* [43], have failed to show training effects [63].

While non-game spatial skill training interventions have also demonstrated empirical effectiveness in classroom environments [32, 65, 66], video games have an advantage as training intervention in that they are designed to harness the power of intrinsic motivation to keep players engaged without any extrinsic incentives like cash payment or course credit [44, 61] typically used in deployments of non-game spatial skill training interventions. Thus, games are likely to be more appealing to students than non-game spatial training interventions, which typically utilize multiple choice questions and sketching exercises [32, 64].

However, it is not well understood why some video games are effective at training spatial skills and some are not since the vast majority of games being studied are commercial games, in which it is very difficult or impossible to isolate and test the efficacy of specific game features. This, 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 video 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.

Research on game-based spatial skill training also tends to focus exclusively on training effects without considering how well a game works for its target audience - both in terms of training and motivation to play. Students with low spatial skill stand the most to gain from spatial training interventions given that their spatial skill becomes a barrier to STEM majors as they progress from high school to college [73, 80], and women consistently make up a disproportionate number of those with low spatial skill [5, 40, 42]. These important demographic groups may not enjoy the kinds of games that are most frequently studied in the spatial skill training literature [58, 80, 85], 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 and evaluated with them as well.

To fill these gaps in the research literature, we conducted a controlled study evaluating the training effectiveness and motivational benefits of a non-commercial spatial skill training game designed around a set of theoretically grounded spatial game features: *Homeworld Bound: Redux* [79]. The game was evaluated in comparison to a set of more traditional workbook-based spatial training exercises and an active control group that completed a graphics programming assignment. As part of the evaluation, we conducted a *data-driven* investigation of the extent to which different combinations of spatial game features were able to tap into players' spatial skills via a regression analysis of the relationship between initial level of spatial skill and in-game performance in each level. Our study was conducted in the context of an introductory course in computer science designed for non-STEM majors (a low spatial skill population) in order to evaluate the intervention in a *player-centric* manner [70, 76] - with the population that stands the most to benefit from spatial skill training interventions. The study utilized a pre/post-test structure; students completed a standardized test of spatial skills, spent about 70 minutes on their assigned spatial training intervention, completed a short questionnaire about their experience with the intervention, and then completed the standardized test again as part of their coursework.

We found that low spatial skill students enjoyed the game-based intervention more than the non-game spatial exercises and that performance in specific levels of *Homeworld Bound: Redux* predicted students' current levels of spatial skill. While we found no evidence that any of the three tested interventions improved low spatial skill students' performance on the test of spatial skills and no practically significant differences in improvement or enjoyment based on demographic factors or initial level of spatial skill, our data-driven analysis of each level of the game provided insights about which parts of the game were most effective at tapping into players' spatial skills and which were least effective. These findings allow us to provide specific design recommendations for improving the efficacy of game-based spatial skill training interventions: 1) use asymmetry in object manipulation tasks, 2) require explicit translation from 2D to 3D environmental representations, and 3) employ time pressure for navigation tasks.

2 RELATED WORK

2.1 Methods of Spatial Skill Training

Spatial skills are somewhat unique among cognitive skills in that training often transfers to a variety of different tasks. The two most common methods of training spatial skills are spatial task training programs conducted in the context of a course [49, 59, 65, 66] and digital games. A meta-analysis by Uttal et al. showed both to be equally effective with an overall moderate effect size ($g = 0.47$) [74]. However, game-based interventions provide an advantage in the form of a motivational element [44, 61].

Action video games, characterized in the cognitive training literature by high speed gameplay and the requirement to quickly attend to multiple moving objects on the screen [26], have been distinguished as one of the more effective genres of games for training spatial skills. Action games such as first person shooters (the *Medal of Honor* series [1, 18] and *Unreal Tournament* [19]) and the arcade games *Stellar 7* [17] and *Battlezone* [4] have been shown to train players' spatial skills [7, 20, 21, 23, 24, 48]. Commercial games from other genres have also been shown to train spatial skills in controlled studies, such as *Crazy Taxi* [13, 31], *Portal 2* [63, 75], and *Tetris* [55, 57, 71].

However, *Lumosity* [43], a suite of minigames designed specifically to train a wide range of cognitive skills, including spatial skills, failed to produce training effects in a controlled study [63]. Why games not designed to train spatial skills are effective at training them but *Lumosity* is not remains unclear, and likely will unless some attempt is made to manipulate and test specific game features in isolation, a difficult proposition with pre-built commercial games.

Non-commercial games present a way of overcoming this limitation, but very little research to date has investigated their training effects. Two recent training studies have evaluated non-commercial first person VR games utilizing tangibles and found training effects [10, 46]. However, these studies are preliminary work with small sample sizes focusing more on evaluating the player experience than the game's training effectiveness. In addition, VR and tangibles are difficult to scale across a large population of students and public school districts with tight budgets.

In this work, we contribute theoretical knowledge about the effectiveness of a spatial skill training game designed around a specific set of theoretically grounded spatial game features in a controlled study with a sample size large enough to observe reliable training effects. In addition, we compare the game-based intervention to not only an active control group activity, but also to an existing non-game intervention shown to train students' spatial skills [65] in order to evaluate whether the game-based intervention provides added value in terms of training effectiveness or enjoyment over more traditional spatial training exercises.

2.2 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. Numerous studies have found that those who play action video games frequently tend to have higher spatial skills than those who do not [9, 20, 24]. Adams and Mayer investigated this relationship with a specific action game, *Unreal Tournament 2004* [19], and found that overall performance in *Unreal Tournament* was more related to players' spatial skill than their performance in the non-action game *Tetris* [2, 57].

However, only two studies that we are aware of have 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, 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 [16].

A different study by Wauck et al. analyzed the correlation between spatial skill and performance in different levels of a first person exploration and 3D object construction game, *Homeworld Bound*. They found that in certain object construction levels, fewer rotation operations, fewer errors, and less time to complete were associated with higher spatial skill. In certain exploration levels, completion time was associated with spatial skill in the same way [81]. However, the number of participants in this study was small for a correlation study ($n = 20$) and participants were children, so the findings may not be generalizable to older populations.

We extend this fine-grained analysis of spatial skill and performance to the new and untested version of Wauck et al.'s spatial skill training game, *Homeworld Bound: Redux*, with a large sample of its intended target population of low spatial skill young adults. Our evaluation of the relationship between spatial skill and level-by-level performance contributes theoretical knowledge about what combinations of game features are successful at tapping into players' spatial skills as well as data-driven recommendations about how these features should be implemented in spatial skill training games to improve their training efficacy.

2.3 Spatial Skill Training and Demographic Factors

Given the importance of spatial skills in everyday life, from navigation to interpreting visualizations [22, 50], those with low spatial skill are at a disadvantage compared to those with more developed spatial skills. This is especially true for those considering careers in STEM, as longitudinal research has shown that spatial skills are one of the strongest predictors of success in STEM, independent of math and verbal ability [62, 77]. 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 [67, 77]. Attaining a certain level of spatial skill may be enough to help low spatial skill students succeed in introductory STEM coursework and transition to more domain-specific knowledge [28, 36, 68, 73]. Thus, students with low spatial skill are likely to benefit the most from spatial skill training interventions [45, 73].

Demographic factors such as gender and prior gaming experience affect spatial skill such that women [5, 40, 42] and those who do not play action video games frequently [9, 20, 24, 25] make up a disproportionately large percentage of those with low spatial skill. These demographic factors also interact with each other; prior work has shown that men tend to play more action video games than women [14, 33, 82, 85]. 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 [29, 33]. Therefore, women and those who do not play action games frequently (lower spatial skill populations) may not be as motivated to play the action games that have been shown to be effective at training spatial skills in the research literature, such as *Unreal Tournament* [19, 23] and *Medal of Honor* [1, 7, 18, 20].

Demographic factors may impact training effectiveness as well. The effect of gender in game-based spatial skill training interventions is mixed; some studies have found that women and men improve equally with a game-based spatial skill training intervention [5, 27], while others have found that women improve more than men [12, 13, 20]. In addition, 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 more slowly than men and women with high levels of spatial activities [71].

Given the potential these demographic factors have to influence both the training effectiveness of a spatial skill training game and the motivation to engage with it, it is critical to ensure that such interventions are effective and motivating for low spatial skill demographic groups that may be less interested in playing the existing commercial action games shown to be effective for a general population. Ensuring spatial skill training games are effective and motivating for low spatial skill students may also be one way of reducing the persistent gender gap in STEM [41].

Thus far, little attention has been paid in the research literature to the impact of motivational factors that may influence a game's training effectiveness, particularly for low spatial skill populations [20, 67]. We aim to fill this gap in knowledge by conducting a *player-centric* [70, 76] evaluation of a spatial skill training game with its intended target population of low spatial skill young adults. Our evaluation analyzes not just the game's training effectiveness but also how fun it is for its target population - the main advantage games have over other types of training interventions. In addition, we conduct a separate analysis of the game's training effects and motivational benefits for specific low spatial skill demographics (women and infrequent video game players). Since one of the end goals of our evaluation is to provide design recommendations for spatial skill training games, our player-centric analyses give members of low spatial skill populations a chance to provide feedback on the game and play an active role in informing the design of future spatial skill training games designed for them.

3 RESEARCH QUESTIONS

The research questions we investigate center around evaluating the non-commercial spatial skill training game *Homeworld Bound: Redux* in terms of enjoyment and training effectiveness for its designated target population of low spatial skill undergraduate students. We chose to evaluate this game in particular given that it is designed around a set of varied and theoretically grounded spatial features and thus the game provides an opportunity to evaluate the effectiveness of different combinations of spatial game features at tapping into players' spatial skills.

RQ1a: How enjoyable for low spatial skill undergraduate students is *Homeworld Bound: Redux* compared to non-game spatial training exercises and a placebo activity?

RQ1b: What is the relationship between enjoyment of each training intervention and current spatial skill level, prior gaming experience, and gender?

RQ2a: How effective is *Homeworld Bound: Redux* compared to non-game spatial training exercises and a placebo activity at training the spatial skills of undergraduate students with an initially low level of spatial skill?

RQ2b: What is the relationship between spatial skill improvement in each training condition and current spatial skill level, prior gaming experience, and gender?

RQ2c: What is the relationship between spatial skill and performance in levels of *Homeworld Bound: Redux* with different combinations of spatial features?

4 METHODS

To answer our research questions, we conducted a training study in the context of an introductory computer science course for non-STEM majors. We 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 [77]. Another advantage to conducting the study in this course was the large class size, which enabled us to be more confident in our estimates of effect size and do follow-up subgroup and correlation analyses. We 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 a spatial skill pretest in the first week of class, an online version of the Revised Visualization of Rotations Test (PSVT:R) (Figure 1) [86]. The online version was created by our research team and automatically randomizes the order of questions and answer choices each time the test is taken to reduce test-retest effects. We logged 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. At the end of the test, the platform displayed the number correct out of 30 but did not grade the result.

One week after taking the test, students participated in a training intervention in their lab section. Since there were 15 lab sections, we 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, we 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 our best to ensure that conditions were evenly distributed over different teaching assistants (some teaching assistants taught multiple sections). We 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.

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 of the Intrinsic Motivation Inventory (IMI), a well-validated measure of intrinsic motivation for a task [47, 60, 72] consisting of 7 7-point Likert scale questions. We 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 us to perform more detailed subgroup analyses of training effects.

Our dependent measure for **RQ1a** and **RQ1b** was enjoyment as measured by the IMI, and our dependent measure for **RQ2a** and **RQ2b** was pretest to post-test score improvement. Our independent measures for these research questions included prior gaming experience and gender (**RQ1b** and **RQ2b** only) as well as the training condition (*Homeworld Bound: Redux*, Spatial Exercises, or Python Graphics), which we describe in more detail below. For **RQ2c**, our dependent measure was pretest score, and our independent measures were aspects of performance in *Homeworld Bound: Redux* (level completion time, number of attempts, and number of errors).

4.1 Training Interventions

4.1.1 Homeworld Bound: Redux. *Homeworld Bound: Redux* is a first person, online computer game designed to train spatial skills and to be appealing to low spatial skill young adults [79]. Game features relevant to spatial skills were proposed from a set of features in other video games that have been empirically demonstrated to train players' spatial skills, such as the object rotation required in *Tetris*

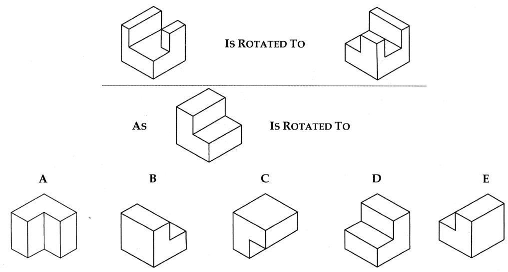

Figure 1: A sample question from the PSVT:R test of mental rotation. Participants are shown an exemplar figure before and after being rotated a certain way and must decide which of the answer choices results from performing the same rotation on the second exemplar figure.

and the first person navigation required in games such as *Portal 2* and *Medal of Honor* [7, 18, 20, 55, 57, 63, 71, 75].

The proposed features were then mapped to an existing taxonomy of spatial skills developed by Chatterjee [11], adopted by other spatial skill researchers [51, 74], and supported by empirical evidence [30, 34, 35, 37]. This taxonomy categorizes spatial skill along two dimensions: *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) [11].

The resulting mapping produced a set of four spatial game features, which 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). Object rotation and object alignment correspond to the *intrinsic-dynamic* quadrant of Chatterjee's taxonomy, while landmark orientation and navigation visualization correspond to the *extrinsic-static* and *extrinsic-dynamic* quadrants, respectively.

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 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, 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.

We chose to study the training effects of *Homeworld Bound: Redux* because while multiple papers have been published about its design and development process [79, 81], 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, we 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.

4.1.2 Spatial Exercises. We included a more traditional non-game spatial training condition to determine if *Homeworld Bound: Redux* does indeed offer advantages in terms of motivation, as claimed by its designers, over more traditional methods of spatial skill training. Our non-game spatial intervention was a subset of the digitized version [84] of Sorby's spatial skill training exercises for first year engineering students [64, 66]. Since these exercises were designed for engineering majors, who typically have higher spatial skill than students in non-STEM majors [77], we 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. The exercises we chose focused on visualizing cross sections of objects, revolving solids about axes, and mental rotation around multiple axes (see Figure 2 for an example of each kind of exercise). The quantity of exercises was chosen so that students would take at least an hour to finish them based on our prior experience using the digital version in engineering classrooms.

4.1.3 Python Graphics. The third training intervention in our study was designed to be more of a regular classroom activity for the course we 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, we considered it unlikely to train the kind of spatial skills we were 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 we 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.

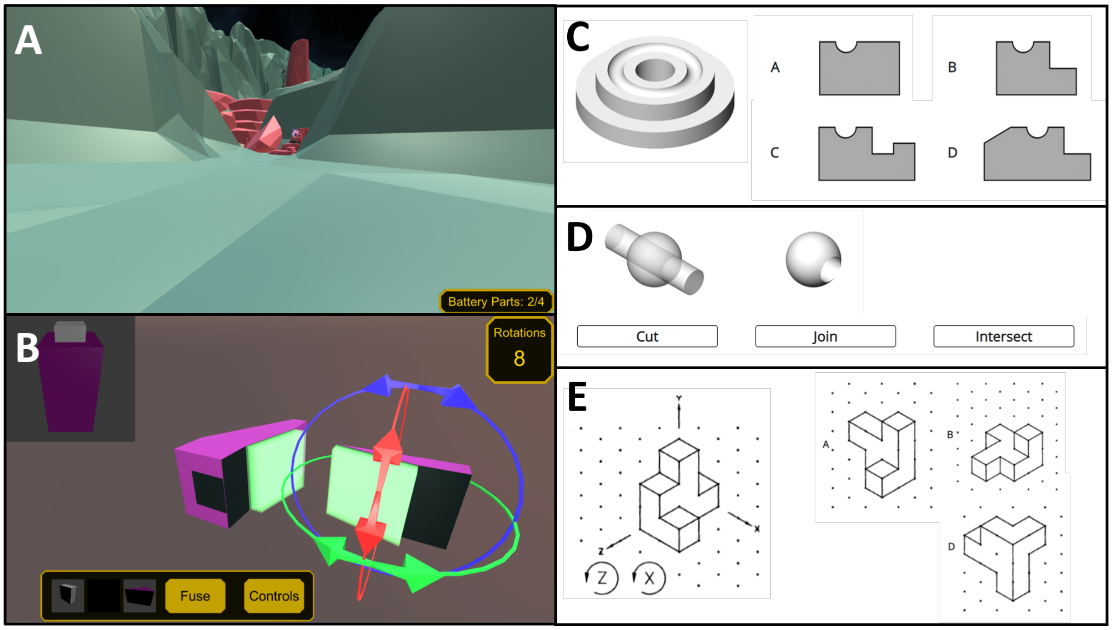

Figure 2: Sample screens from the two spatial training interventions. ***Left:*** **screenshots of Exploration Mode (A) and Construction Mode (B) in** ***Homeworld Bound: Redux. Right:*** **Examples of three question types used in the Spatial Exercises intervention: Choosing the 2D object that produced 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).**

5 RESULTS

5.1 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, we 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, we analyzed the relationship between time spent on the test and test score. We chose time spent versus test score as our metric of test taking effort since students took the test on their own time outside of class, were not graded on their performance on the test, and presumably would not spend much time on each question if they were not taking the test seriously.

Via plot inspection, we observed a positive correlation between time spent and test score for those who spent a low amount of time on the test up until a certain threshold such that no one who spent less time than the threshold scored above around 15 points. For the pretest, this threshold was at about 8 minutes, whereas for the post-test it was about 5 minutes. Therefore, we omitted 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 and received a low score. This criteria eliminated 116 additional students from our dataset (48 in the game condition and 34 in each of the other conditions), leaving us with a total of 343 students in our dataset.

A large number of unanswered questions on the test might also indicate that students had not taken the test seriously if they started it and then let the clock run out while they did something else. Overall, 26 students failed to answer all 30 questions for either the pretest, posttest, or both. The mean number of questions left unanswered in this group was 3.3 and the average test score was 15.67. Given that no students answered less than 63% of test questions and that their average score was well above the level of chance, we did not remove these students from our dataset since it is likely they were taking the test seriously.

5.2 Participant Characteristics

Of the students in our cleaned dataset, 59% identified as male, 40% as female, with the remaining 1% identifying as nonbinary, genderfluid, or 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. Scores on the spatial skill pretest had a roughly normal but left-skewed distribution ranging from 4 to 30 (perfect score), and scores overall (μ = 16.44, median = 17) were lower than the cutoff used in previous engineering research for identifying low spatial skill students (18/30 or 60%) [56, 65].

5.3 Enjoyment (RQ1a and RQ1b)

To determine which spatial training condition was most motivating for low spatial skill students, we analyzed the data of the subset 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). We 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 [56, 65]. As a check, we performed the same analyses using a 21/30 cutoff that is used by engineering professors at our university to identify first year engineering 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, we performed a one-way ANOVA, 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 3). Thus, low spatial skill students reported enjoying the Python Graphics condition the most, but enjoyed *Homeworld Bound: Redux* more than the Spatial Exercises (**RQ1a**).

To determine what individual characteristics and demographic factors may have influenced enjoyment of training interventions, we conducted one linear regression for each training condition, with IMI enjoyment score 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 ($n = 3$), we removed them from the dataset for the purposes of this analysis, but otherwise used our entire dataset ($n = 340$). For each linear regression analysis, we 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, 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 $R^2 = 0.047$, $F(3, 106) = 2.78$, $p = 0.045$). Thus, those with higher pretest scores enjoyed the Spatial Exercises condition more, but there were no other significant predictors of enjoyment in any training condition (**RQ1b**).

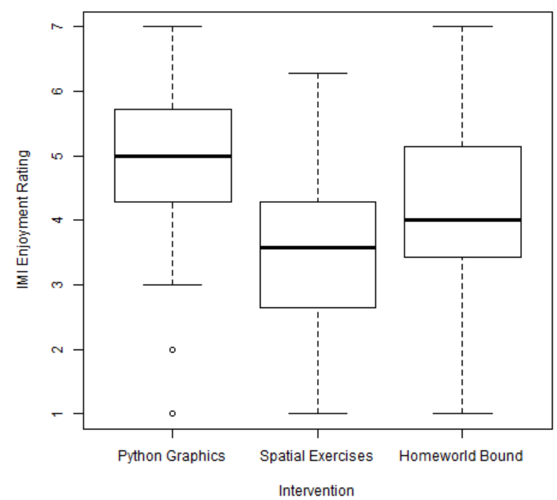

Figure 3: Average Intrinsic Motivation Inventory (IMI) ratings of enjoyment for each training intervention. The scale asked students to indicate the degree of agreement with a set of 7 statements. Higher ratings correspond to more enjoyment.

5.4 Training Effects (RQ2a and RQ2b)

To answer the question of how training effectiveness compared across conditions for low spatial skill students, we 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, using the same low spatial skill subset of the sample used in the enjoyment analysis. 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 ($F(2, 198) = 0.33$, $p = 0.72$) (**RQ2a**).

To determine what individual characteristics and demographic factors may have influenced training effectiveness in each condition, we 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 using the same dataset as in the demographic analysis for enjoyment. 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.

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 $R^2 = 0.16$, $F(3, 106) = 8.15$, $p < 0.0001$). In summary, the only predictor of pre- to post-test spatial skill improvement was pretest score across all three conditions, although the overall model was not significant in the *Homeworld Bound: Redux* condition (**RQ2b**).

5.5 Spatial Skill and In-Game Performance (RQ2c)

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, we conducted a regression analysis investigating the relationship between students' performance in *Homeworld Bound* and their current level of spatial skill, as measured by pretest score, while controlling for prior gaming experience. 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, we were left with $n = 76$ students for the linear regression analysis. For all 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.

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). Our 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. We 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 measure of prior gaming experience) as a covariate. 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 ($\beta = 0.73$, $t = 4.33$, $p < 0.0001$), while weekly video game hours was not ($\beta = -0.03$, $t = -0.061$, $p = 0.95$).

To get a more fine-grained understanding of how performance in *Homeworld Bound: Redux* was related to current spatial skill level, we 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 [83], we analyzed performance for only the first 8 levels (the first 4 levels of Exploration Mode and the first 4 levels of Construction Mode).

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, we 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. We 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 1. Overall, better performance in most Construction Mode levels was associated with higher spatial skill pretest scores even after controlling for prior video game experience, especially for B2 and B3, but performance was not associated with spatial skill pretest score in any Exploration Mode levels (**RQ5**). Weekly hours of video games played was not significantly associated with pretest score in any linear regression. Performance was comparable across levels except for B1 having noticeably fewer average errors and taking less time to complete (see Table 2).

6 DISCUSSION AND FUTURE WORK

6.1 The Game is More Enjoyable than Spatial Exercises

Overall, we found that low spatial skill students enjoyed *Homeworld Bound: Redux* more than the other intervention intended to train spatial skills, Spatial Exercises (**RQ1a**), 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 activity might 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, and students playing *Homeworld Bound: Redux* had enough content to play for the entire 70 minute lab period.

Table 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.

		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	0.00020
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 2: Median and mean performance on first 8 *Homeworld Bound: Redux* levels. Time is measured in seconds.

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

Analyzing potential demographic differences in enjoyment revealed that pretest score was a significant predictor of intervention enjoyment, but 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, we found no evidence that the demographics analyzed in our experiment influenced enjoyment of training intervention (**RQ1b**) 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 we 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 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 [8, 23, 39, 58].

6.2 Short Term Training Yields No Training Effects

We found no evidence of a training effect for any intervention (**RQ2a**). While our training period of 70 minutes was short for a cognitive training study, other controlled studies have found spatial skill training effects using training times of an hour or less [10, 13, 46]. However, one key difference between these other short interval training studies and ours was that each of them conducted pretests, interventions, and post-tests within a very short time period (all in the same day for Mazalek et al. and Chang et al., and within 48 hours for Cherney et al.). Spacing our 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 allowed us to minimize possible test-retest effects and test longer term transferability of our training interventions. Therefore, it is likely that spatial skill training lasting an hour or less results in only short term training effects, whereas longer training periods are necessary to observe lasting training effects. Future work should evaluate *Homeworld Bound: Redux*'s training effectiveness with a longer training period, spaced across multiple weeks as in other spatial skill training studies [6, 7, 20, 23, 24, 63].

Our analysis of demographic predictors of pretest to post-test improvement found that those with lower pretest scores improved more on the post-test in all three conditions. 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. Therefore, as in our analysis of demographic predictors of enjoyment, there were no practically significant effects of demographics on training effectiveness (**RQ2b**). Future work is needed to investigate whether demographic influences on training effectiveness might appear with longer term training interventions.

It is possible that our results may have been affected by our use of time spent taking the test as the sole measure of test effort rather than an explicit attention check question. To see if any students who would have likely failed an explicit attention check question remained in our dataset, we looked for repetitive answer patterns on pretest and posttest data and found 9 students with repetitive patterns of question answering (answering the same choice for almost all questions). Our time spent metric had already removed all but one of these data points from the dataset. Reanalysis of the data showed that none of our results materially changed after removing this additional data point.

6.3 Only Construction Tasks Tap Into Spatial Skills

Our 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 1). Thus, the game mechanics of *object rotation* and *object alignment* appear to be spatial as implemented in 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 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 2). It is also possible that the learning curve involved in learning how to play affected performance more than spatial skill since B1 was the first Construction Mode level.

Another more likely possibility is that players are using the shapes of the 2D attachment areas on parts to align them rather than

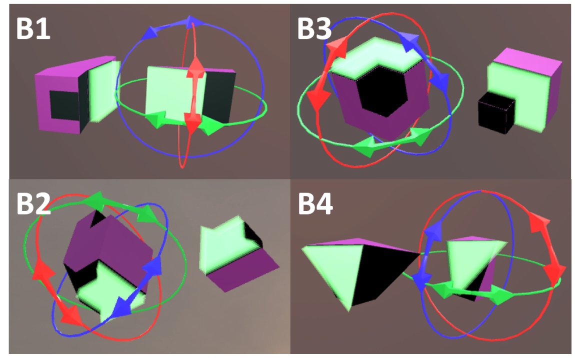

Figure 4: Examples of 2D attachment area shapes in levels B1-B4 in *Homeworld Bound: Redux*'s Construction mode. Glowing green areas indicate the attachment areas the player is currently trying to align by rotating one of the parts. In levels B1 and B4, the 2D shapes are simpler and more symmetric, potentially reducing the need to employ mental rotation skills to complete the level.

the 3D shapes of the parts themselves. In the B1 and B4 levels, these attachment areas are generally rectangular or triangular, whereas in the B2 and B3 levels, they are more complex and contain substantial asymmetry (see Figure 4). 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 more mental rotation effort than the simpler and more symmetrical shapes found in B1 and B4.

Performance in Exploration Mode levels 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 [20, 23, 63] (e.g. *Unreal Tournament* [19], *Portal 2* [75], *Medal of Honor: Pacific Assault* [18]), it is somewhat surprising that no Exploration Mode levels tap into spatial skills given its emphasis on first person navigation. It may be that the navigation task in the first four Exploration Mode levels was too easy, so players could wander around until they found all the parts, eliminating the need to use *landmark orientation*. Another possibility is that the environment had enough proximal landmark cues that players' need to engage their spatial skills was reduced, a phenomenon observed in prior work on navigation games [16].

A third possibility is that the presence of a map may be crucial for ensuring players engage spatial skills. First person shooters like *Unreal Tournament* [19] as well as other navigation-based games shown to train spatial skills like the *Segway Circuit* minigame in *Wii Fit* [53] 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, a time limit, or clues that depict the location of parts, which may force players to be more strategic in their navigation. Future work is needed to explore the relationship between performance and spatial skill on these later levels to test this hypothesis.

7 DESIGN RECOMMENDATIONS

Since our findings revealed that *Homeworld Bound: Redux* players may be using nonspatial strategies for completing certain levels of the game and thus not exercising and developing their spatial skills as intended by the game's designers, our design recommendations for improving the efficacy of spatial skill training games all center around the general principle of incentivizing spatial strategy use.

7.1 Asymmetrical Object Manipulation

First, since our data suggests *Homeworld Bound: Redux* players may have been mentally rotating the simpler 2D shapes on object faces instead of the objects themselves in Construction Mode levels, we recommend using visually complex, asymmetrical shapes for all object rotation and object alignment tasks in spatial skill training games, including any 2D shapes embedded within objects (demarcated by different colors or borders, for example, as in the attachment areas on parts in Construction Mode levels). Simpler, more symmetric shapes like rectangles and triangles that require less effort to mentally rotate should be avoided.

7.2 Explicit Translation from 2D to 3D Representations

Requiring players to perform translations from 2D to 3D representations might be another way to increase the difficulty of mental rotation tasks in games involving construction tasks like *Homeworld Bound: Redux*'s Construction Mode. If the player must rely entirely on a separate 2D diagram showing what the finished construction is supposed to look like to decide what parts to attach where, they will have to engage their spatial skills in a different manner, using mental perspective-taking to convert the parts visible in the 2D diagram (perhaps an isometric or top-down view) to their 3D representations in the game environment. Determining what rotations are necessary to achieve the alignment of parts shown in the diagram could also engage players' mental rotation skills. One way of ensuring the player is required to use the 2D diagram to build an object might be to make the shapes and sizes of all parts identical.

Since we found no evidence that the first four levels of Exploration Mode in *Homeworld Bound: Redux* tapped into players' spatial skills, we recommend incorporating some sort of map use to prevent players from engaging in nonstrategic wandering that does not tap into their spatial skills in games with *landmark orientation* and *navigation visualization* tasks where the player has to explore in first person and collect items or navigate to specific locations. Pictorial clues showing a screenshot or drawing of a destination and immediate surroundings (like those used to locate places where the player can recover memories in *The Legend of Zelda: Breath of the Wild* [54]) could be used in combination with environmental elements that make it impossible or very difficult to find items or locations without using *landmark orientation* to figure out where to go to see the view of the item shown in the clue. 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 [16].

7.3 Employ Time Pressure for Navigation Tasks

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 spatial skills instead of wandering randomly until they find what they are looking for. Time pressure could be achieved by having players be chased by enemies, an environmental hazard (e.g. a flood, the rolling stone in *Indiana Jones*) or by simply giving players a fixed amount of time to navigate through the environment. While the relationship between spatial skill and performance in *Homeworld Bound: Redux's* Construction Mode levels demonstrates that time pressure is not necessary to tap into spatial skills for *object rotation* and *object alignment* tasks, it is possible that time limits could provide a way of increasing the difficulty of a level that is too easy.

8 LIMITATIONS

The chief limitation of this work is that the training interval we used was only 70 minutes, a product of our decision to sacrifice training time for a larger sample of low spatial skill students. In addition, the Spatial Exercises and Python Graphics exercises did not take the full class period as intended for most students, so average training duration was even shorter in these conditions. This also might have affected enjoyment ratings; students who enjoyed the activities less may have rated shorter ones higher. The short training duration we used in our study also limited the number of *Homeworld Bound: Redux* levels students could get through in the allotted training time. Furthermore, our ability to analyze the prior gaming experience of students was limited by our use of one self-report question about average weekly hours of video game play as a proxy for this metric. Adding more detailed prior game experience data, such as preferred game genres, may have yielded different results.

9 CONCLUSION

In this work, we investigated the training effectiveness and enjoyment of a non-commercial computer game designed to train spatial skills, *Homeworld Bound: Redux*, among low spatial skill undergraduate students over a 70 minute training period. Low spatial skill students reported enjoying *Homeworld Bound: Redux* more than the workbook exercises, and neither gender nor amount of gaming experience practically influenced training effectiveness or enjoyment. We found no training effects of the training intervention. However, spatial skill was associated with performance on certain levels of *Homeworld Bound: Redux* emphasizing object manipulation operations, but not on levels emphasizing navigation. Our results provide insights for game designers, educators, and researchers about which game features are better at tapping into players' spatial skills and how these insights can be used to design and revise spatial skill training games to improve their training effectiveness.

ACKNOWLEDGMENTS

The authors would like to thank Professor Craig Zilles as well as his teaching and course assistants for allowing us to conduct this study as part of his introductory computer science course. This research was conducted while the first author was a doctoral student at the University of Illinois Urbana-Champaign.

REFERENCES

[1] 2015, Inc. 2002. *Medal of Honor: Allied Assault.* Game [PC]. Electronic Arts, Los Angeles, California, USA.

[2] Deanne Adams and Rich Mayer. 2012. Examining the Connection Between Dynamic and Static Spatial Skills and Video Game Performance. In *Proceedings of the Annual Meeting of the Cognitive Science Society.*

[3] Atari Games. 1984. *Marble Madness.* Game [Apple II]. Atari Games, Milpitas, California, USA.

[4] Atari, Inc. 1984. *Battlezone.* Game [Apple II]. Atari Inc., Sunnyvale, California, USA.

[5] Maryann Baenninger and Nora Newcombe. 1989. The role of experience in spatial test performance: A meta-analysis. *Sex Roles* 20, 5-6 (mar 1989), 327–344. https://doi.org/10.1007/BF00287729

[6] Chandramallika Basak, Walter R. Boot, Michelle W. Voss, and Arthur F. Kramer. 2008. Can training in a real-time strategy video game attenuate cognitive decline in older adults? *Psychology and Aging* 23, 4 (2008), 765–777. https://doi.org/10.1037/a0013494

[7] Walter R. Boot, Arthur F. Kramer, Daniel J. Simons, Monica Fabiani, and Gabriele Gratton. 2008. The effects of video game playing on attention, memory, and executive control. *Acta Psychologica* 129, 3 (nov 2008), 387–398. https://doi.org/10.1016/j.actpsy.2008.09.005

[8] Martha Carr, Natalia Alexeev, Lu Wang, Nicole Barned, Erin Horan, and Adam Reed. 2018. The Development of Spatial Skills in Elementary School Students. *Child Development* 89, 2 (2018). https://doi.org/10.1111/cdev.12753

[9] Alan D. Castel, Jay Pratt, and Emily Drummond. 2005. The effects of action video game experience on the time course of inhibition of return and the efficiency of visual search. *Acta Psychologica* 119, 2 (jun 2005), 217–230. https://doi.org/10.1016/j.actpsy.2005.02.004

[10] Jack Shen-Kuen Chang, Georgina Yeboah, Alison Doucette, Paul Clifton, Michael Nitsche, Timothy Welsh, and Ali Mazalek. 2017. Evaluating the effect of tangible virtual reality on spatial perspective taking ability. In *Proceedings of the 5th Symposium on Spatial User Interaction - SUI '17.* ACM Press, New York, New York, USA, 68–77. https://doi.org/10.1145/3131277.3132171

[11] Anjan Chatterjee. 2008. The neural organization of spatial thought and language. *Seminars in Speech and Language* (2008). https://doi.org/10.1055/s-0028-1082886

[12] Isabelle D. Cherney. 2008. Mom, Let Me Play More Computer Games: They Improve My Mental Rotation Skills. *Sex Roles* 59, 11-12 (dec 2008), 776–786. https://doi.org/10.1007/s11199-008-9498-z

[13] Isabelle D. Cherney, Kyle Bersted, and Joseph Smetter. 2014. Training Spatial Skills in Men and Women. *Perceptual and Motor Skills* 119, 1 (aug 2014), 82–99. https://doi.org/10.2466/23.25.PMS.119c12z0

[14] Isabelle D. Cherney and Kamala London. 2006. Gender-linked differences in the toys, television shows, computer games, and outdoor activities of 5- to 13-year-old children. *Sex Roles* 54, 9-10 (may 2006), 717–726. https://doi.org/10.1007/s11199-006-9037-8

[15] National Research Council. 2005. *Learning to Think Spatially: GIS as a Support System in the K-12 Curriculum.* National Academies Press. 1–313 pages. https://doi.org/10.17226/11019

[16] Suzanne de Castell, Hector Larios, and Jennifer Jenson. 2019. Gender, videogames and navigation in virtual space. *Acta Psychologica* (2019). https://doi.org/10.1016/j.actpsy.2019.102895

[17] Dynamix, Inc. 1983. *Stellar 7.* Game [Apple II]. Dynamix, Inc., Eugene, Oregon, USA.

[18] EA Los Angeles. 2004. *Medal of Honor: Pacific Assault.* Game [PC]. EA Games, Los Angeles, California, USA.

[19] Epic Games, Digital Extremes. 2004. *Unreal Tournament 2004.* Game [PC]. Epic Games, Cary, North Carolina, USA.

[20] Jing Feng, Ian Spence, and Jay Pratt. 2007. Playing an Action Video Game Reduces Gender Differences in Spatial Cognition. *Psychological Science* 18, 10 (oct 2007), 850–855. https://doi.org/10.1111/j.1467-9280.2007.01990.x

[21] Diana Gagnon. 1985. Videogames and spatial skills: An exploratory study. *Educational Communication and Technology* 33, 4 (1985), 263–275.

[22] Reginald G. Golledge. 2002. The Nature of Geographic Knowledge. *Annals of the Association of American Geographers* 92, 1 (mar 2002), 1–14. https://doi.org/10.1111/1467-8306.00276

[23] C.Shawn Green and Daphne Bavelier. 2007. Action-Video-Game Experience Alters the Spatial Resolution of Vision. *Psychological Science* 18, 1 (jan 2007), 88–94. https://doi.org/10.1111/j.1467-9280.2007.01853.x

[24] C. Shawn Green and Daphne Bavelier. 2003. Action video game modifies visual selective attention. *Nature* 423, 6939 (may 2003), 534–537. https://doi.org/10.1038/nature01647

[25] C Shawn Green and Daphne Bavelier. 2006. Effect of action video games on the spatial distribution of visuospatial attention. *Journal of experimental psychology. Human perception and performance* 32, 6 (dec 2006), 1465–78. https://doi.org/10.1037/0096-1523.32.6.1465

[26] C. Shawn Green and Daphne Bavelier. 2012. Learning, Attentional Control, and Action Video Games. *Current Biology* 22, 6 (mar 2012), R197–R206. https://doi.org/10.1016/j.cub.2012.02.012

[27] Patricia M. Greenfield, Craig Brannon, and David Lohr. 1994. Two-dimensional representation of movement through three- dimensional space: The role of video game expertise. *Journal of Applied Developmental Psychology* 15 (1994), 87–103. https://doi.org/10.1016/0193-3973(94)90007-8

[28] David Z. Hambrick, Julie C. Libarkin, Heather L. Petcovic, Kathleen M. Baker, Joe Elkins, Caitlin N. Callahan, Sheldon P. Turner, Tara A. Rench, and Nicole D. LaDue. 2012. A test of the circumvention-of-limits hypothesis in scientific problem solving: The case of geological bedrock mapping. *Journal of Experimental Psychology: General* 141, 3 (2012), 397–403. https://doi.org/10.1037/a0025927

[29] Tilo Hartmann and Christoph Klimmt. 2006. Gender and Computer Games: Exploring Females' Dislikes. *Journal of Computer-Mediated Communication* 11, 4 (jul 2006), 910–931. https://doi.org/10.1111/j.1083-6101.2006.00301.x

[30] Mary Hegarty, Daniel R. Montello, Anthony E. Richardson, Toru Ishikawa, and Kristin Lovelace. 2006. Spatial abilities at different scales: Individual differences in aptitude-test performance and spatial-layout learning. *Intelligence* (2006). https://doi.org/10.1016/j.intell.2005.09.005

[31] Hitmaker. 2001. *Crazy Taxi.* Game [Gamecube]. Sega, Tokyo, Japan.

[32] Sherry Hsi, Marcia C. Linn, and John E. Bell. 1997. Role of spatial reasoning in engineering and the design of spatial instruction. *Journal of Engineering Education* 86, 2 (1997), 151–158. https://doi.org/10.1002/j.2168-9830.1997.tb00278.x

[33] Jeroen Jansz and Lonneke Martens. 2005. Gaming at a LAN event: the social context of playing video games. *New Media & Society* 7, 3 (jun 2005), 333–355. https://doi.org/10.1177/1461444805052280

[34] Maria Kozhevnikov, Mary Hegarty, and Richard Mayer. 2002. Spatial Abilities in Problem Solving in Kinematics. In *Diagrammatic Representation and Reasoning.* https://doi.org/10.1007/978-1-4471-0109-3_9

[35] Maria Kozhevnikov, Stephen Kosslyn, and Jennifer Shephard. 2005. Spatial versus object visualizers: A new characterization of visual cognitive style. *Memory and Cognition* (2005). https://doi.org/10.3758/BF03195337

[36] Maria Kozhevnikov, Michael A. Motes, and Mary Hegarty. 2007. Spatial Visualization in Physics Problem Solving. *Cognitive Science* 31, 4 (jul 2007), 549–579. https://doi.org/10.1080/15326900701399897

[37] Maria Kozhevnikov, Michael A. Motes, Bjoern Rasch, and Olessia Blajenkova. 2006. Perspective-taking vs. mental rotation transformations and how they predict spatial navigation performance. *Applied Cognitive Psychology* 20, 3 (apr 2006), 397–417. https://doi.org/10.1002/acp.1192

[38] Simone Kühn, Tobias Gleich, Robert C. Lorenz, Ulman Lindenberger, and Jürgen Gallinat. 2014. Playing Super Mario induces structural brain plasticity: gray matter changes resulting from training with a commercial video game. *Molecular psychiatry* 19, August 2013 (2014), 265–71. http://www.ncbi.nlm.nih.gov/pubmed/24166407

[39] Susan C. Levine, Marina Vasilyeva, Stella F. Lourenco, Nora S. Newcombe, and Janellen Huttenlocher. 2005. Socioeconomic status modifies the sex difference in spatial skill. *Psychological Science* 16, 11 (nov 2005), 841–845.

[40] Lynn S. Liben. 2015. The STEM Gender Gap: The Case for Spatial Interventions. *International Journal of Gender, Science and Technology* 7, 2 (2015), 133–150.

[41] Lynn S. Liben and Emily F. Coyle. 2014. Developmental Interventions to Address the STEM Gender Gap: Exploring Intended and Unintended Consequences. *Advances in Child Development and Behavior* 47 (jan 2014), 77–115. https://doi.org/10.1016/BS.ACDB.2014.06.001

[42] Marcia C. Linn and Anne C. Petersen. 1985. Emergence and Characterization of Sex Differences in Spatial Ability: A Meta-Analysis. *Child Development* 56, 6 (dec 1985), 1479. https://doi.org/10.2307/1130467

[43] Lumos Labs Inc. 2007. *Lumosity.* Game [PC]. Lumos Labs Inc, San Francisco, California, USA.

[44] Thomas W. Malone and Mark R. Lepper. 1987. Making Learning Fun: A Taxonomy of Intrinsic Motivations for Learning. *Aptitude, learning, and instruction* 3 (1987), 223–253.

[45] Lauren E. Margulieux. 2019. 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.* https://doi.org/10.1145/3291279.3339414

[46] Ali Mazalek, Sanjay Chandrasekharan, Michael Nitsche, Tim Welsh, Paul Clifton, Andrew Quitmeyer, Firaz Peer, Friedrich Kirschner, and Dilip Athreya. 2011. 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.* ACM Press, New York, New York, USA, 129. https://doi.org/10.1145/1935701.1935727

[47] Edward McAuley, Terry Duncan, and Vance V. Tammen. 1989. Psychometric Properties of the Intrinsic Motivation Inventory in a Competitive Sport Setting: A Confirmatory Factor Analysis. *Research Quarterly for Exercise and Sport* 60, 1 (mar 1989), 48–58. https://doi.org/10.1080/02701367.1989.10607413

[48] Patricia A. McClurg and Christine Chaillé. 1990. Computer Games: Environments for Developing Spatial Cognition? *Journal of Educational Computing Research* 3, 1 (oct 1990), 1–1. https://doi.org/10.2190/9N5U-P3E9-R1X8-0RQM

[49] David I. Miller and Diane F. Halpern. 2013. Can spatial training improve long-term outcomes for gifted STEM undergraduates? *Learning and Individual Differences*

26 (2013). https://doi.org/10.1016/j.lindif.2012.03.012

[50] Marina Milner-Bolotin and Samson Madera Nashon. 2012. The essence of student visual-spatial literacy and higher order thinking skills in undergraduate biology. , 25–30 pages. https://doi.org/10.1007/s00709-011-0346-6

[51] Nora S. Newcombe and Thomas F. Shipley. 2015. Thinking About Spatial Thinking: New Typology, New Assessments. In *Studying Visual and Spatial Reasoning for Design Creativity*. Springer Netherlands, Dordrecht, 179–192. https://doi.org/10.1007/978-94-017-9297-4_10

[52] Nintendo EAD. 1996. *Super Mario 64*. Game [N64]. Nintendo, Kyoto, Japan.

[53] Nintendo EAD. 2008. *Wii Fit*. Game [Wii]. Nintendo, Kyoto, Japan.

[54] Nintendo EPD. 2017. *The Legend of Zelda: Breath of the Wild*. Game [Nintendo Switch]. Nintendo, Kyoto, Japan.

[55] Lynn Okagaki and Peter A. Frensch. 1994. Effects of video game playing on measures of spatial performance: Gender effects in late adolescence. *Journal of Applied Developmental Psychology* 15, 1 (jan 1994), 33–58. https://doi.org/10.1016/0193-3973(94)90005-1

[56] Richard M. Onyancha, Matthew Derov, and Brad L. Kinsey. 2009. 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). https://doi.org/10.1002/j.2168-9830.2009.tb01014.x

[57] Alexey Pajitnov. 1984. *Tetris*. Game [PC]. Spectrum Holobyte, Boulder, Colorado, USA.

[58] Claudia Quaiser-Pohl, Christian Geiser, and Wolfgang Lehmann. 2006. The relationship between computer-game preference, gender, and mental-rotation ability. *Personality and Individual Differences* 40, 3 (2006), 609–619. https://doi.org/10.1016/j.paid.2005.07.015

[59] Cristina Roca-González, Jorge Martin-Gutierrez, Melchor García-Dominguez, and Ma del Carmen Mato Carrodeguas. 2017. Virtual technologies to develop visual-spatial ability in engineering students. *Eurasia Journal of Mathematics, Science and Technology Education* (2017). https://doi.org/10.12973/eurasia.2017.00625a

[60] Richard M. Ryan. 1982. Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of Personality and Social Psychology* 43, 3 (1982), 450–461. https://doi.org/10.1037/0022-3514.43.3.450

[61] Richard M. Ryan, C. Scott Rigby, and Andrew Przybylski. 2006. The Motivational Pull of Video Games: A Self-Determination Theory Approach. *Motivation and Emotion* 30, 4 (dec 2006), 344–360. https://doi.org/10.1007/s11031-006-9051-8

[62] Daniel L. Shea, David Lubinski, and Camilla P. Benbow. 2001. Importance of assessing spatial ability in intellectually talented young adolescents: A 20-year longitudinal study. *Journal of Educational Psychology* 93, 3 (2001), 604–614. https://doi.org/10.1037/0022-0663.93.3.604

[63] Valerie J. Shute, Matthew Ventura, and Fengfeng Ke. 2015. The power of play: The effects of Portal 2 and Lumosity on cognitive and noncognitive skills. *Computers and Education* 80 (2015). https://doi.org/10.1016/j.compedu.2014.08.013

[64] Sheryl A. Sorby. 2001. A Course In Spatial Visualization and its Impact on the Retention of Female Engineering Students. *Journal of Women and Minorities in Science and Engineering* 7, 2 (2001), 20. https://doi.org/10.1615/jwomenminorscieneng.v7.i2.50

[65] Sheryl A. Sorby. 2009. Educational Research in Developing 3-D Spatial Skills for Engineering Students. *International Journal of Science Education* 31, 3 (feb 2009), 459–480. https://doi.org/10.1080/09500690802595839

[66] Sheryl A. Sorby and Beverly J. Baartmans. 2000. The Development and Assessment of a Course for Enhancing the 3-D Spatial Visualization Skills of First Year Engineering Students. *Journal of Engineering Education* 89, 3 (jul 2000), 301–307. https://doi.org/10.1002/j.2168-9830.2000.tb00529.x

[67] Ian Spence and Jing Feng. 2010. Video games and spatial cognition. *Review of General Psychology* 14, 2 (2010), 92–104. https://doi.org/10.1037/a0019491

[68] Mike Stieff. 2007. Mental rotation and diagrammatic reasoning in science. *Learning and Instruction* 17, 2 (apr 2007), 219–234. https://doi.org/10.1016/J.LEARNINSTRUC.2007.01.012

[69] Kaveri Subrahmanyam and Patricia M. Greenfield. 1994. Effect of video game practice on spatial skills in girls and boys. *Journal of Applied Developmental Psychology* 15, 1 (jan 1994), 13–32. https://doi.org/10.1016/0193-3973(94)90004-3

[70] Penelope Sweetser and Daniel Johnson. 2004. Player-Centered Game Environments: Assessing Player Opinions, Experiences, and Issues. In *International Conference on Entertainment Computing*. Springer, 321–332. https://doi.org/10.1007/978-3-540-28643-1_40

[71] Melissa S. Terlecki, Nora S. Newcombe, and Michelle Little. 2008. Durable and generalized effects of spatial experience on mental rotation: gender differences in growth patterns. *Applied Cognitive Psychology* 22, 7 (nov 2008), 996–1013. https://doi.org/10.1002/acp.1420

[72] Nikolaos Tsigilis and Argiris Theodosiou. 2003. Temporal Stability of the Intrinsic Motivation Inventory. *Perceptual and Motor Skills* 97, 1 (aug 2003), 271–280. https://doi.org/10.2466/pms.2003.97.1.271

[73] David Uttal and Cheryl Cohen. 2012. Spatial Thinking and STEM Education: When, Why, and How? In *The Psychology of Learning and Motivation*. Elsevier. https://doi.org/10.1016/B978-0-12-394293-7.00004-2

[74] David H. Uttal, Nathaniel G. Meadow, Elizabeth Tipton, Linda L. Hand, Alison R. Alden, Christopher Warren, and Nora S. Newcombe. 2012. The Malleability of Spatial Skills: A Meta-Analysis of Training Studies. *Psychological Bulletin* 139, 2 (mar 2012), 352–402. https://doi.org/10.1037/a0028446

[75] Valve Corporation. 2011. *Portal 2*. Game [PC]. Valve Corporation, Bellevue, Washington, USA.

[76] Vero Vanden Abeele, Bob De Schutter, Luc Geurts, Stef Desmet, Jeroen Wauters, Jelle Husson, Lieven Van den Audenaeren, Frederik Van Broeckhoven, Jan-Henk Annema, and David Geerts. 2012. P-III: A Player-Centered, Iterative, Interdisciplinary and Integrated Framework for Serious Game Design and Development. In *Serious Games: The Challenge*. Springer, 82–86. https://doi.org/10.1007/978-3-642-33814-4_14

[77] Jonathan Wai, David Lubinski, and Camilla P. Benbow. 2009. Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology* 101, 4 (2009), 817–835. https://doi.org/10.1037/a0016127

[78] Jonathan Wai, David H Uttal, and American Enterprise Institute (AEI). 2018. *Why Spatial Reasoning Matters for Education Policy*. Technical Report.

[79] Helen Wauck and Brian Bailey. 2019. 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*. Association for Computing Machinery, Inc, New York, New York, USA, 763–771. https://doi.org/10.1145/3341215.3356298

[80] Helen Wauck, Elisa D. Mekler, and Wai-Tat Fu. 2019. A Player-Centric Approach to Designing Spatial Skill Training Games. In *CHI Conference on Human Factors in Computing Systems Proceedings*. Association for Computing Machinery, Glasgow, 13. https://doi.org/10.1145/3290605.3300296

[81] H. Wauck, Z. Xiao, P.-T. Chiu, and W.-T. Fu. 2017. Untangling the Relationship Between Spatial Skills, Game Features, and Gender in a Video Game. In *International Conference on Intelligent User Interfaces, Proceedings IUI*. https://doi.org/10.1145/3025171.3025225

[82] Claudia Wilhelm. 2018. Gender role orientation and gaming behavior revisited: examining mediated and moderated effects. *Information, Communication & Society* 21, 2 (feb 2018), 224–240. https://doi.org/10.1080/1369118X.2016.1271902

[83] Carmen R. Wilson Van Voorhis and Betsy L. Morgan. 2007. Understanding Power and Rules of Thumb for Determining Sample Sizes. *Tutorials in Quantitative Methods for Psychology* 3, 2 (sep 2007), 43–50. https://doi.org/10.20982/tqmp.03.2.p043

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

[85] Nick Yee. 2017. Beyond 50/50: Breaking Down The Percentage of Female Gamers By Genre. http://quanticfoundry.com/2017/01/19/female-gamers-by-genre/

[86] So Yoon Yoon. 2011. Revised Purdue Spatial Visualization Test: Visualization of Rotations (Revised PSVT:R). https://www.silc.northwestern.edu/revised-purdue-spatial-visualization-test-revised-psvtr-visualization-of-rotations/