UNIVERSITY OF TWENTE.

Faculty of Electrical Engineering,
Mathematics & Computer Science

Case study of the development of an Academic Task Game for Behavioural- and Neuropsychology

Tim Yeung

M.Sc. Thesis for Interaction Technology

April 18, 2025

Supervisors:
First Supervisor: M.A. Gómez Maureira PhD
Second Supervisor: dr. M.A. Friehs
Acting chair: dr.ir. R.W. van Delden

Abstract

Over the years, many researchers have attempted to apply ludic elements from games to perform experimental tasks. Cognitive psychologists have tested multiple approaches to "gamify" validated tasks. However, there is no consistent framework for the development of such academic task and there is a lack of consensus on how validated tasks can be "gamified". Furthermore, game elements must be applied with care, as ludic games often present elements that conflict with the validity of a task. This thesis investigates how game design can be applied to a game version of a property verification task for cognitive psychology to increase ecological validity while maintaining experimental validity. This game was developed in a collaborative setting with three cognitive psychology researchers with no prior experience with academic games, and two experts of academic games as participants to the collaboration. The game was developed using direct feedback from the participants in four iterative feedback cycles, resulting in a game that satisfied the validity requirements set by the participants. The entire collaborative development was also documented, including design rationale, iterative changes, and collaboration challenges. The discussion of the process and the final result may help the development of future cognitive task games by suggesting strategies to satisfy researcher's expectations for valid cognitive tasks. Adding narrative elements and game-like elements such as 3D environments and game-UI to cogntivie tasks could help generate ecological validity without sacrificing the validity of the experiment. Developers should consider the specific design of visual elements to reduce distractions and make sure that the narrative framing does not interfere with the research goal or introduce biases. Validated academic task games could limit the variability of the game between participants, by making the game state progress independent of the player by applying hidden manipulations to guide players into consistent game states, though this may compromise repeated experiments and the overall image of future academic games. Furthermore, strategies for collaborative development with researchers, such as setting the final design earlier in development, being able to visualize the design on a detailed level through mock-ups, preparing discussions for topics that could potentially overlooked by researchers, and offering multiple levels of "gamification" approaches can help smoothen communication with stakeholders and prevent the concept from being changed throughout implementation.

Acknowledgments

I would like to express my thanks to my initial supervisors Marcello A. ”Maro” Gómez Maureira and Maximillian A. ”Max” Friehs for their tireless fforts, working through holidays and late hours to support my work and ultimately resulting in this thesis, as well as a published paper from my research topics. I am truly blessed to have had you as my supervisors. Additionally, I would like to give my thanks to the chair of my graduation committee, Robby W. van Delden for graciously coming in at the end, providing much needed outside feedback, and allowing me to graduate.

To my collaborators at the CoPla and CogEl, Philip Kuhnke, Gesa Hartwigsen, and Markus Kiefer, thank you for giving me your time, the opportunity to work on such an interesting task, and allowing me to gain such valuable experience.

Special thanks to my family, parents, and grandparents for working hard for many years to support me in my studies.

Finally, thanks to my friends and most of all my other half for supporting me everywhere and always, through my ups and downs.

Contents

Legend of the
Lunchbox

Chapter1

Introduction

Since their inception, video games have grown into a widely popular medium, attracting increasing interest in their applications beyond entertainment. Researchers and developers have increasingly collaborated to design and implement video games as tools for academic research, exploring their potential for applied purposes. This has led to the rise of serious games as effective tools in various domains [1]. One of these domains is using video games as a research tools to study cognitive and motor functions, a practice dating back to at least the 1990s [2].

In research settings, games can be designed to incorporate experimental tasks, allowing participants to engage with research activities in a game-like environment. This approach helps maintain the suspension of disbelief, which is often crucial in such studies [3], [4]. For instance, cognitive psychology researchers have developed video games as alternatives to traditional assessments like the Stop Signal Task, where participants must inhibit a response upon receiving a stop signal [5]. Presenting tasks within a game format can enhance engagement with minimal gameplay elements [6]–[9]. More complex games that incorporate narratives and immersive environments can further improve ecological validity, allow for detailed event logging, and support scalability through modular design [10].

However, designing academic games for experimental research—and applied games more broadly—introduces a tension between game mechanics intended for player engagement and those necessary to meet research objectives [11]. This challenge, referred to as ludo-utilitarian dissonance [12], requires developers to balance engaging gameplay with the rigorous experimental controls needed for valid research outcomes. As a result, trade-offs may be necessary, potentially affecting both participant experience and the integrity of the experimental data.

This dissonance is also reflected in the development process of academic games. Researchers wanting to employ games for their studies often hire internal or external developers to co-develop the game for them or have some sort of experience with game or interaction design themselves. However, this collaboration is often far from smooth due to an imbalance in the understanding of academic games. Researchers are sometimes unable to support or suggest engaging game designs, while developers may lack understanding in how game design affects the research [12].

One such field that particularly struggles with implementing games into research tasks is psychology. Traditional psychological experiments consists of validated tasks that are designed to be highly controlled. However, games are designed to be engaging and fun, which introduces confounding variables and studies that utilize gamified tasks often require comparability with previous studies using the conventional task in order to be relevant to the field [13].

Games that incorporate validated tasks, often do so for the sake of improving engagement and maintaining performance [7], [9], [14]–[18]. However, such games have also been utilized for their capability to increase ecological validity and immerse the player [3], [12], though increasing ecological validity is only explicitly targeted in a handful of validated cognitive tasks [7], [9]. Additionally, the ecological validity offered by the context of a game needs to be balanced very carefully against deviating from the validated task. This leads to the following main research question underlying this current thesis:

How can game design be integrated into the development of an academic task game to create a methodologically valid game acting as a validated task for research?

- What game mechanics can be adapted to enhance ecological validity while maintaining experimental validity?
- What strategies facilitate effective collaboration between cognitive psychologists and game developers to ensure experimental validity of an academic task game?

This report aims to contribute to the development of future academic task games by documenting the design and development of an academic test game. I will focus on how specific design elements were crafted to support research validity, the iterative changes made over time to maintain and enhance that validity, as well as challenges that arose in developing a game in collaboration with academic stakeholders.

To this end, I developed a game in collaboration with two researchers affiliated with the "Cognition and Plasticity research group at the Max Planck Institute for Human Cognitive and Brain Sciences" and one researcher affiliated with the "Section for Cognitive Electrophysiology at Ulm University", these parties hereafter referred to as *CoPla* and *CogEl* respectively.

This game, commissioned by the CoPla for their use in future research, aimed to function as a property verification task within experimental psychology research. I provided designs and implementation of this game, guided by feedback from the two researchers affiliated with CoPla as well as one researcher from the CogEl, and two experts in academic game development, affiliated with the University of Twente. This development was part of a larger multifaceted project aimed at examining both the collaborative dynamics between researchers and developers and the unique design challenges inherent to academic game development. A preliminary interview study was conducted as part of the Research Topics for this Thesis to inform the collaborative development process of this project and also resulted in a full paper to be presented at Foundations of Digital Games 2025. While this report focuses on the documentation and analysis of the game's design, a planned subsequent publication will present a meta-analysis of the collaborative development process.

This game meets the requirements set by the CoPla stakeholders, providing added value to the original task through the enhancement of ecological validity, while maintaining the mechanical validity of the task and is intended for use in future works. The discussion of this process will serve to show the conflicts between ludic games and classical experiments, and suggest strategies of what ludic elements can be incorporated into experiments, and how certain developmental strategies can help smoothen the development process of future academic task games.

During the course of this thesis, the focus of the research has evolved significantly. In the original scope, the collaborative elements of the development process were focused on more, making the design of the game itself secondary to the meta-analysis of the collaborative process. As a result, the interview study conducted during the research topics phase was mainly conducted with the goal of informing the collaborative elements of the thesis. However, the relevant results of the research topic are still presented in this thesis to inform the design of the collaboration, as well as my philosophy for the design of the academic task games.

Chapter2

Related work

Many studies that use games as research tasks primarily focus on the final product and research methodology, offering limited insight into the design process [6], [7], [9], [19], [20]. This lack of transparency regarding the reasoning behind specific design choices can lead to the unintended change of essential elements in future research, such as when adding specific game elements that were originally omitted for experimental control. Consequently, discussions on the challenges of game design and development are often overlooked, particularly in relation to the complexities of interdisciplinary collaboration and its impact on the design process.

2.1 Exploratory interview study

To explore the field using games as tasks, I conducted an interview study that aimed to demystify the challenges of designing academic task games in a collaborative setting. This preliminary study was approved by the EEMCS ethics board at the University of Twente under application number 240322 and acted as part of the research topic for this thesis. Below I will briefly detail the relevant results of that study, though for the interested reader, a pre-print version of that paper can be found in Appendix B: Research Topics, "Mapping Academic Game Development: An Exploratory Investigation".

During the study, a small sample of 9 salaried professionals within academia were interviewed based on their experience using video game-like tools as tasks to conduct research. These participants were questioned on their role within the development process, assuming a distinction between *developers* of the game vs. *researchers* also referred to as *stakeholders* employing the game.

These participants were then interviewed in online semi-structured interviews that provided insights into the purpose of game-like elements within the games of participants, as well as challenges found in maintaining the validity of their research. To help categorize these participants, a 3 axis paradigm was conceived in the form of a canvas that participants could fill in to help categorize

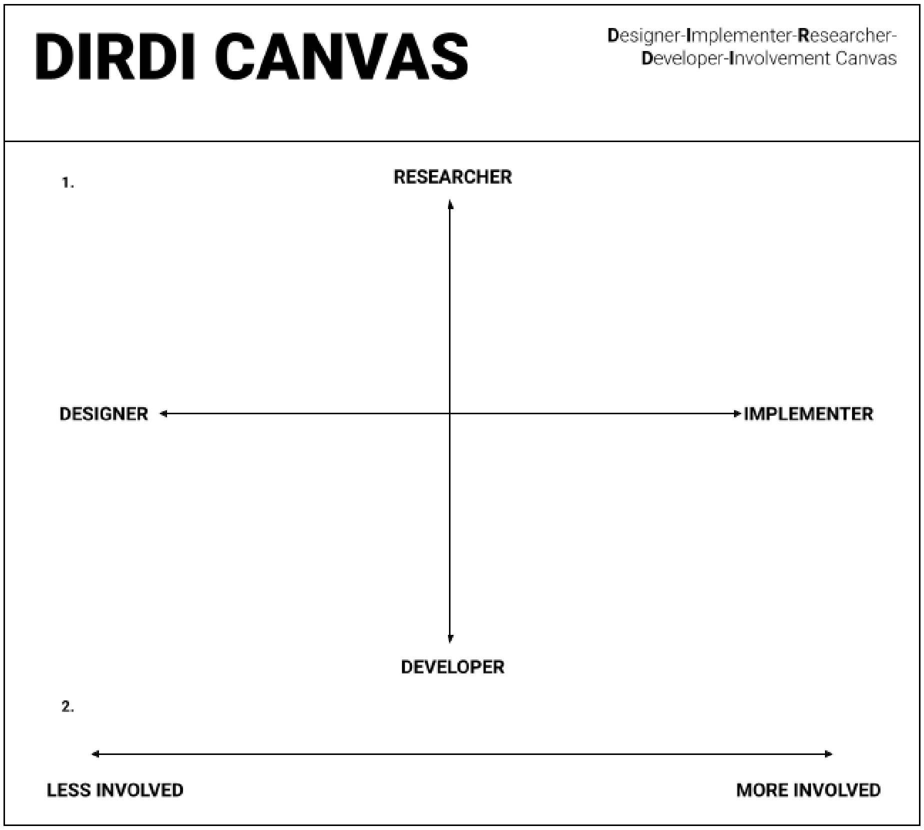


Figure 2.1: DIRDI CANVAS utilized in the exploratory interview study.

themselves based on their role in the academic game development process. The first axis distinguished between participants with expertise in games versus participants with expertise in a research field that aimed to integrate games, while the second axis contrasted those more engaged in the technical implementation of an academic game and those more engaged with the design aspects. The third axis represented the participant's overall involvement in the project relative to other stakeholders of the development. This canvas was aptly named the quick and DIRDI canvas (Designer Implementer Researcher Developer Involvement) and can be found in Figure 2.1.

The results of the interviews were processed with abductive analysis and revealed nine shared themes among participants. Using those themes, a framework was conceptualized, categorizing games that are used as an experimental task into three categories. The most important takeaways for this project are summarized below.

Task games encompass a very broad spectrum of applications. The study conceptualized three different purposes for incorporating games as research tasks. Note that the terms "emulating" and "embedded" games were changed in the latest version of the paper, with the pre-print referring to these as "immerse" and "necessitate" games.

Engaging games are designed with a primary focus on maximizing participant engagement. The incorporation of game elements in these projects requires a high level of control to ensure alignment with the adapted validated task, often following a structured experience. Engagement is typically fostered through narrative components and visual styles that emulate those of traditional video games. This approach has been shown to produce valid results while enhancing participant engagement [6]–[9]. Additional strategies, such as implementing scoring systems [8] and promoting para-telic motivation through enjoyable tasks [21], can further improve motivation. Developers should aim to enhance task satisfaction without introducing mechanics that fundamentally alter the core research task [16].

Emulating games aim to replicate real-world behaviors, often leveraging virtual reality (VR) to enhance realism. While this approach offers the potential for high ecological validity, it also presents challenges in managing participants' playful tendencies, which can compromise behavioral authenticity. To address this, additional game elements are typically minimized to maintain focus on the core task. Researchers in academic game development have highlighted the importance of improving the integration of experimental design elements, such as participant briefings, noting that varying levels of gaming familiarity can affect the clarity of instructions. Moreover, transitions between tasks, such as shifting from gameplay to survey administration can disrupt immersion [22], [23]. This suggests that the entire experimental process, from instructions to evaluations, should be seamlessly embedded within the game environment to preserve participant engagement.

Embedded games integrate tasks within conventional video games to study behavior in specific gaming contexts. However, using commercial games poses challenges, such as limited control over in-game events and mechanics. To address this, re-engineered versions are often created to replicate the original experience while allowing for experimental manipulation. These games, which rely on authentic gaming contexts, face fewer design constraints if replicating a true game environment aligns with the research goals.

Across all task games, incorporating familiar game mechanics or adapting designs from existing games can be used to reduce learning curves and streamline both development and participant on-boarding. Furthermore, some players may be more experienced with games and others less. This skill variation needs to be considered and accounted for when developing the task game, such as by implementing ways of measuring performance beforehand or making games accessible for all skill levels

Regarding the collaboration within academic game development, collaboration between researchers who lack video game or design knowledge and developers may be challenging and requires careful communication to effectively translate the goal of the researcher into an effective and valid game.

Researchers are crucial in identifying validity issues during a game's development, though developers should be cautious of input that may unintentionally compromise the research objectives. However, as the ultimate stakeholders, the requirements of researchers should override recommendations of the developer.

Developers should also investigate the background of the research task to gain a basic understanding of the research objective and task mechanics and researchers may provide relevant works to help prepare the developer. Clear research objectives and requirements should be set before the start of development.

Regular communication is recommended and regular calibration meetings should be used to explain and discuss relevant design decisions and get feedback regarding research validity. Iterative development can help manage expectations and rough prototypes can be used to demonstrate core game mechanics.

Documentation of the design process, including changes, may help also help inform future academic games and continuations of similar research.

These findings acted as the foundations of the general design of the game mechanics and collaboration with the "researchers" in the upcoming design project.

2.2 Validated task games

While the interview study provided broad insights into academic task games, certain challenges arise when integrating cognitive psychology tasks into game-based environments. Validated tasks in traditional experimental settings require careful adaptation, as game elements may introduce confounding variables that may impact task validity [7], [9], [14]–[18], [24].

Several recent works have attempted to adapt established cognitive psychology tasks to "gamified" solutions. Most of these works take an additive approach to "gamification", taking elements from games and applying it to the pre-existing task. In the following paragraphs, I discuss the different game elements that previous works have applied to validated cognitive tasks and what they can add to experimental tasks.

One of the most common applications of game elements in these tasks is using narrative storytelling [7], [9], [16], [24]. Researchers have tried different approaches to implementing this method, some choosing to only show the narrative and the context through visuals and text before presenting the task separately [16], while others may also change the presentation of the task itself, such as replacing the minimalist graphics of the original task with 3D models that support the narrative context [7], [9], [18]. Wiley et al. noted that the first approach was able to temporarily increase enjoyment before and during the game, though enjoyment after playing the game was worse than the original task, which they believe is caused due to a mismatch between the story context and the game mechanics [16]. Other researchers that applied the second approach encountered mixed results. Friehs et al. and Kirsten et al. noted that the narrative and visual context did not affect the performance of participants when compared to the original task [7], [9]. Participants only felt more engagement when having played/performed both the game and the conventional test, but the researchers did claim that the added visual complexity of the game was believed to enhance ecological validity [7]. On the other hand, Lumsden et al. demonstrated that the exact design of the stimuli could negatively influence performance [24]. The game-visual-like stimuli that their participants needed to respond to were found to be less distinct than that of the original task, making the task more difficult for participants and creating more false-positives.

Points have also commonly been applied to validated tasks as a way to add extrinsic rewards to the execution of the tasks [15], [16], [24]. In these cases, points involves the usage of some sort of score that is visible to the player. This score is then increased or decreased based on the player's performance of the task. To accentuate the importance of the points, games using this mechanic also incorporate visual or audio feedback. The implementation of points has primarily served to maintain participant engagement throughout the experiment, and has been confirmed to be able to improve enjoyment [15], [16], [24]. However, researchers have noted different results regarding the effect of points on performance. Wiley et al. (2020) found that points increased reaction time and error rate. In a later study, Wiley et al. (2024) conducted a study with a different validated task, but similar added game mechanics [15]. This study found that overall performance was enhanced, with players reaching peak performance earlier and maintaining peak performance longer. On the other hand, an earlier study by Lumsden et al. found that adding points in their task game did not affect significantly affect reaction times and accuracy at all [24].

The above goes to show that even among multiple games, results can vary depending on the execution of the design. Two games could have have the same points mechanic, but the execution of one game, or even the base task of one game may change the experience to the point of influencing the results. Thus, academic game developers must carefully consider not only what types of mechanics to add to their task, but also the specific design of those elements, in order to ensure that the intended effect on the experiment is accomplished.

2.3 Insights validated task games

These studies highlight the tradeoffs in applying game elements to validated experiments, and serve to help inform the design philosophy that was employed throughout the design of the study. First, the presentation of the game primarily focused on the usage of a narrative context, including visual design that is common in games. Second, mechanics such as points and mechanics that encouraged players to vary their behavior or directly distract from or overlap with the main task were avoided. Third, where possible, other game mechanics that potentially offered more engagement without affecting the primary focus of the study could be explored, but had to be discussed with the collaborators.

Finally, the design of the game took inspiration from existing popular games to ensure that designs were pre-validated (through their popularity or acclaim), and using existing design aimed to maximize accessibility among participants with different levels of familiarity with games. Additionally, the design of the collaboration included an iterative development cycle with multiple calibration meetings. Researchers were asked to set the requirements for a valid game suiting their research objective at the start of the collaboration and provide related materials to help understand the background of the task and study. Finally prototypes and mock-ups were included to aid in the communication with the researchers.

The following sections will describe the the variables of the study and the constraints of the task, followed by a documentation of the development, describing how these insights have shaped the design choices and the resulting game's validity.

Chapter3

Methodology

This section outlines the parameters and methodology of the design case. It first details the participants, data collection, and initial requirements for validity established by the CoPla participants. It then describes the study's methodology, which follows an iterative development cycle. In total, five cycles were completed, incorporating feedback from four semi-structured meetings, conducted during each cycle, with researchers from CoPla, CogEl, and experts from the University of Twente (UT) all of whom also acted as participants to the validation of the game. The iterative design cycle is structured into two key phases: (1) implementation of design modifications following previous feedback and requirements, and (2) evaluation through the meeting discussions. Findings of the meeting discussion are then combined to evaluate the final result with respect to the validation of the requirements.

3.1 Participants

Participants were recruited through professional networks and were affiliated with either the UT, CoPla, or CogEl. The study included two distinct groups based on the second axis of the DIRDI canvas (see Figure 2.1 from the exploratory interview study), distinguishing between cognitive psychology "researchers" who had no prior experience in academic game development, including those from the CoPla and CogEl and academic game "experts" from the UT who possessed relevant expertise in the field.

The "researcher" participants comprised three salaried cognitive psychology researchers affiliated with the CoPla or CogEl, who also served as end-user representatives for the implementation of the final game. The "expert" group consisted of two salaried researchers from the UT, part of the Human Media Interaction and the Psychology of Conflict, Risk, and Safety research groups. In total, the focus group included five participants, two from CoPla, one from CogEl and two from UT.

This composition was intentionally structured to examine ludo-utilitarian dissonance within the development process. By incorporating researchers without prior experience in academic game development, the study allowed for the natural emergence of development challenges, while allowing these researchers to challenge the design by conventional experimental standards within cognitive psychology. Furthermore, the inclusion of "expert" participants ensured that the design process remained informed by established academic game development principles. This balance facilitated a more comprehensive understanding of development from both a research perspective and a game development perspective, contributing to a more multifaceted evaluation of the design.

3.2 Data collection and discussion design

The focus group meetings with the participants were conducted through online MS Teams meetings. Both audio, video, and automated transcripts through the MS Teams platform were recorded of all participants, ranging from 86 to 110 minutes in length. I obtained informed consent for all the data collected, with no incentive other than the developed game offered. The processing of this meeting data was approved by the ethics board at the University of Twente under application number 250417.

The meetings comprising the focus group followed a semi-structured format with ample leeway given to participants for additional discussion amongst each other. Each meeting consisted of the following key components: (1) a summary of development progress since the previous meeting, (2) a presentation of recent design changes, accompanied by visual mock-ups or a working prototype to help researchers understand and envision designs, and (3) a discussion of specific design considerations, which I, as the developer, had prepared in advance. Visual mock-ups were prepared as a strategy to help visualize the game for the "researcher" participants, and specific questions were prepared for the discussion to prompt them on issues that I believed could be overlooked by the researchers due to a lack of familiarity with games. These discussions included an evaluation of project requirements to validate the completion of requirements. Additionally, any emerging topics relevant to the design process were addressed as they arose throughout the meetings and meeting-specific points outside of the validation of requirements, such as discussions regarding deployment of the game within, were added per meeting. The summary per meetings were based on key themes that emerged in responses of the stakeholders. While no formal transcript analysis was conducted, iterative feedback cycles were used to ensure validity through direct stakeholder input. This process is further discussed in *Procedure*.

3.3 Validation requirements

The objective of this project was the development of a video game designed to function as a property verification task, also known as a feature verification task. This task is used to investigate cognition in the field of semantic memory and conceptual knowledge, and participants intended to use this task to specifically investigate the representation of perceptual-motor features, following previous works [25], [26]. This type of task involves assessing a subject's ability to determine whether a specific property (e.g., "is yellow" or "can be thrown") is applicable to a given object or concept. A typical implementation of a property verification tasks involve a sequence of computer-based trials in which participants are presented with a structured series of stimuli. Typically, each trial consists of at least the following stages: (1) a visual or auditory stimulus representing an object, (2) a delay period, and (3) a subsequent visual or auditory stimulus representing a potential feature of the object. Participants must then respond by indicating whether the presented feature is an accurate property of the initial stimulus [27], [28]. Each screen is designed to be as controlled and non-distracting as possible, using monochromatic backgrounds and minimizing the number of visual elements on a screen at any given time. Trials are performed in blocks, with each block allowing for the testing of different experiment conditions. For example, the first block could have the participant only focus on material properties, while the second block only prompts sound qualities. Users are also able to take short breaks between these blocks. A visualization of the screens of a general property verification task is given in Figure 3.1.

The development of the game aimed to enhance the ecological validity of the traditional property verification task by embedding it within a more immersive and contextually relevant environment. To leverage the interactive nature of the game medium, the design prioritized the presentation of physical objects alongside their corresponding auditory and action-pantomime properties. For example, a hammer would first be visually presented, followed by the sound of a hammer striking a nail (sound) or an animation of a person miming the act of swinging it (action-pantomime).

Additionally, to facilitate validation studies, the game incorporated an option to replace visual and auditory stimuli with plain text representations. This feature enabled direct comparisons between word-based and sensory-based object-property pairings, as well as cross-modal trials where a textual object description was paired with a visual or auditory property, or vice versa.

Standard practice for obtaining participants at the CoPla was to use financial incentives. Thus, while increased participant engagement was recognized as a potential advantage of the game format, it was not a primary objective. Instead, CoPla emphasized maintaining the core structure of the original task to ensure methodological rigor and preserve its validity, thereby mitigating potential critiques in future research employing the game.

In summary, the following requirements were taken into account.

- **R1**: The game must present a task that is mechanically analogous to a property verification task.
  - Specifically, it should contain the following sequences: presenting stimuli, presenting a property, having the user evaluate the property.

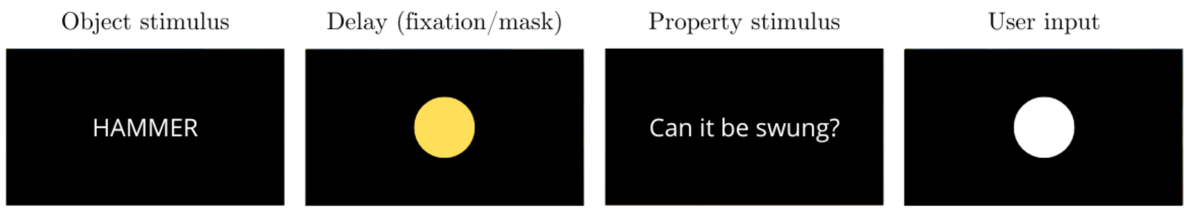


Figure 3.1: Example of a basic property verification task.[1]

- **R2**: The game must enhance the ecological validity of conventional property verification tests.
- **R3**: The game must maintain experimental validity of the task to ensure its suitability for future research applications and comparability with previous works.

Other functional and technical requirements were also prioritized but not specifically specified to support the game's use in experimental settings. These included incorporating design features compatible with external measurement tools such as EEG and fMRI, the support of localization (i.e. translation and implementation of text) for German language, and being able to output logs that would be compatible with the CoPla's existing methods.

3.4 Procedure

The development of the game was initially scheduled to span four months, from September 2024 to December 2024, but was extended to January 2025 due to scheduling conflicts with the participants and challenges in the implementation phase.

Throughout this period, a total of five online meetings were conducted with the participants. These meetings included an initial meeting to define the project's scope and requirements, followed by four meetings acting as the combined focus group, designed to facilitate progress reviews, gather iterative feedback, and discuss necessary adjustments to the game's design.

Development followed an iterative cycle, visualized in Figure 3.2, with each cycle starting with refinement of the design and further implementation based on the previous meeting or initial design. The four meetings at the end of each cycle are referred to as a milestone meeting from here on out, with the number indicating the current cycle.

For Milestone 1, three initial prototypes were presented, and stakeholders provided feedback to determine the preferred direction. The following cycle focused on refining this concept, resulting in a visual mock-up consisting of storyboards that illustrated gameplay flow and task integration. At Milestone 2,

[1] Inspired by design from Liuzzi et al. [28]

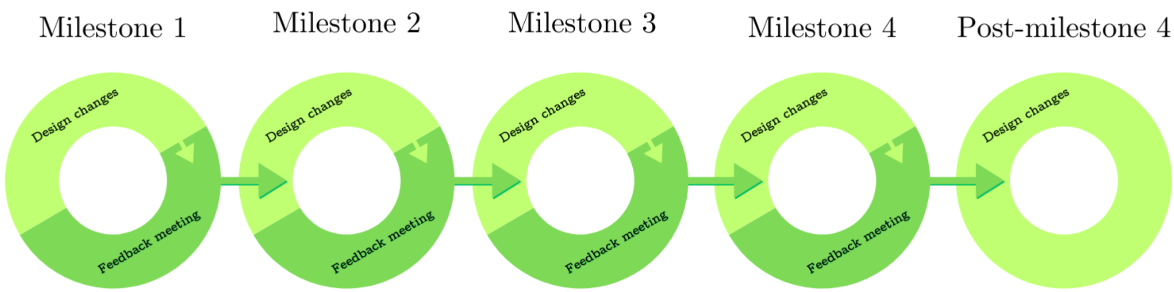


Figure 3.2: Overview of the 4 milestone cycles and post-milestone 4.

stakeholders reviewed a video mock-up of the main interaction, addressing potential conflicts with research objectives before approving further development.

The next cycle emphasized implementation, leading to a high-fidelity prototype by Milestone 3. This iteration incorporated aesthetics, functionality, and logistical considerations such as data input and localization. Stakeholder feedback guided final refinements before the last cycle, culminating in a fully functional version presented at Milestone 4.

After milestone 4, additional changes were made to the game based on the feedback in the final milestone meeting, although these changes were not further discussed by participants. For completeness they have nonetheless been included in the documentation under a section called "Post-milestone 4.

This iterative approach, was informed by the interview study conducted as a preamble to this study [12] and aimed to ensure continuous refinement while addressing stakeholder challenges in conceptualizing game design and maintaining research validity.

The game was developed using Unity 2022, chosen for its ease of use, ability to build to multiple platforms and relatively low performance cost, which aids the performance of the game on lower spec machines.

At the end of the development period, the discussions from the focus group were summarized and used to validate the satisfaction of requirements at the end of the development.

Although the entire development process has been thoroughly documented, the in-depth analysis of the collaboration dynamics between the CoPla/CogEl stakeholders, the University of Twente supervisors, and the author falls beyond the scope of this report. These aspects will be explored in greater detail in a future publication focusing on the meta-analysis of this development case.

Chapter4

Result

The following section details the chronological development of the audio, visual, and mechanical elements of the final property verification task game, titled *Legend of the Lunchbox*. The section is sub sectioned by the iterative cycles of the project, starting with the development of three initial game concepts and then showcasing the design considerations and feedback throughout the cycles. These sections are organized into further sub sections for each distinct aspect to the design, i.e., audio, visual, mechanical elements. This section provides short descriptions of actionable feedback from the focus group meetings that was taken into account for the subsequent iterations of design changes. The full findings of all combined meetings are then given at the end of the section.

Milestone 1

1. Developed four initial game concepts (A1-A4).
2. Combined A2 and A3 and polished and presented the remaining three concepts (A1, A3, A4).
3. A3 selected by stakeholders.

Milestone 2

1. Re-contextualized throwing mechanic of A3 to reduce cognitive bias, removed combat elements and focused on real-world knowledge.
2. Changed stimuli design of properties.
3. Re-designed the friendly encounters to a different mini-game.
4. Added post-trial and post-encounter feedback.

Milestone 3

1. Changed presentation order and timings of stimuli that the user responds to.
2. Implemented support for external assets to allow for validation studies of stimuli.
3. Shift from third- to first-person perspective.
4. Re-contextualized health bar.
5. Removed progress bar and health bar during trials.
6. Re-designed the timer during trials.
7. Changed visual elements to be less distracting.
8. Added environmental sound and effects.

Milestone 4

1. Added additional options for researchers to add or remove specific game elements.
2. Finalized UI elements for encounters.
3. Finalized design of level environments.
4. Incorporated input features for stimuli, timings.
5. Implemented tutorial and opening and ending narrative cut scenes.

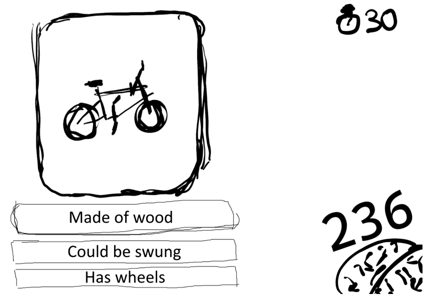
Post-milestone 4

1. Enabled researchers to select the number of levels and which background environment and sound is used for each level.
2. Added introductory location text at start of level.
3. Removed letter grade from level-over screen.

4.1 Milestone 1

During milestone 1, initial concepts for a possible game that could satisfy the requirements of the participants were explored. To this end, I developed four different game concepts with different levels of abstraction from the original property verification task to gauge to what extent the mechanics of the original task could be modified while retaining the core cognitive principles of the task,

(a) Screen capture from Big Brain Academy: Brain vs Brain [29] showing a mini-game where the player needs to match and identify shadows with their colored form.[1]



(b) Concept A1 storyboard sketch. The main mini-game interface where the player is given a multiple choice prompt to assign a property

Figure 4.1: Comparison between the inspiration for Concept A1 and a sketch visualizing the proposed design.

.

These concepts, designated A1, A2, A3, and A4 (in increasing order of deviation from the original task), were intended to assess the optimal balance between experimental fidelity and engaging gameplay. Concepts were obtained by taking inspiration from existing games with similar interactions to the mechanics of the conventional task and adapting elements to better fit the task or leaving out elements that were not directly relevant to the task. This practice was adopted following the results of the preliminary study [12], which indicated two possible benefits. First, using existing game elements from games that are widely known or are embedded in popular culture may help reduce the learning time for participants. Second, taking inspiration from existing games could help streamline the design process and reduce the time spent on the design.

4.1.1 Concept A1

Concept A1 retained the original task structure while embedding it within a "brain training" game format, inspired by titles such as Big Brain Academy [29] and Brain Age (marketed as Dr. Kawashima's Brain Training in PAL regions) [30]. In these games, players complete mini-games that include puzzles that test specific skills like math, pattern recognition, and logical capabilities. The main game loop of these games consist of the player undertaking tests, in which multiple of these mini-games must be solved consecutively.

The design of these games already provides little distraction, with screen designs for the mini-games consisting of a consistently colored background with the main content of the mini-game, including the puzzle itself and buttons for the multiple choice answers, being placed square in the center of the screen,

[1] Image source: [31]

capturing the player's full attention. Additional information, such as the prompt for the puzzle, the "brain age" and "big brain brawn" (tangible representations of the player's skill at solving the puzzles), the player avatar, the timer, are all placed outside of the area of the puzzle. This information is also not necessary when solving the puzzle, allowing the user to fully focus their attention on the puzzle itself during gameplay. The design itself is somewhat similar to the presentation of the original task, giving the users puzzles to solve in a very systematic way with minimal distracting components, which made it very suitable to embed the property verification task into. For an example of a mini-game screen, please refer to Figure 4.1a.

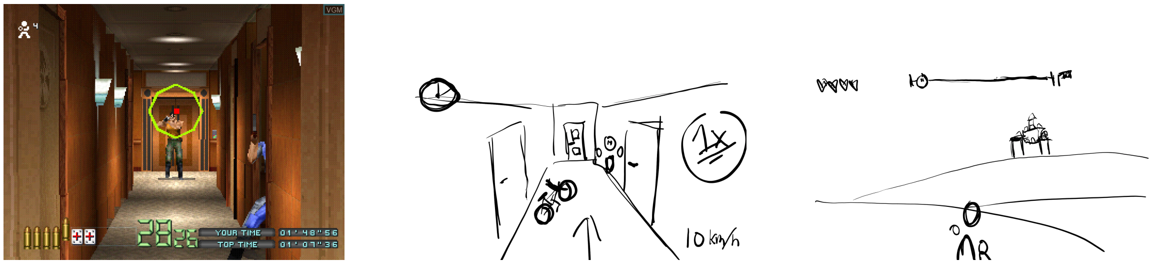
A1's design replaced the various puzzles with a presentation of the conventional property verification task that was slightly modified to fit the format of the original game. Rather than first presenting the object stimulus, having it disappear, and then showing the property, the trials would retain the image of the object on the screen and include text and icon multiple choice prompts, see Figure 4.1b.

To retain the competitive element and sense of self challenge found within the original game, high scores and rankings were simplified. Only a single numerical score was displayed to the user, their "brain" score and a textual rank would be displayed next to it which compared the intelligence of the user to a specific age or education group (e.g., "smarter than a kindergartner," "high school graduate," etc.). This simplification aimed to reduce the complexity of the scoring system to make it more accessible to people with less game experience. Since many trials would have to be performed, trials were subdivided over levels which helped give the player a sense of progression, and also allowed researchers to include multiple conditions within their study design by splitting conditions over levels, similar to the blocks of the conventional test. Additionally, by dividing trials into discrete levels, this approach allowed for natural break points, reducing cognitive fatigue while maintaining a sense of achievement and progression.

4.1.2 Concept A2

Building upon Concept A1, Concept A2 introduced a more immersive and organic environment while maintaining the core property verification task. In this concept the player navigated an on-rails experience, where the player would automatically move through the game environment and stop at specific points prompting the player with some sort of challenge or interaction. Inspired by arcade "on-rails" light gun games like Time Crisis [32] or interactive shooting theme park rides like the Buzz Lightyear Laser Blast in Disneyland Paris [33], the design of the concept facilitated a more complex looking game by allowing the player to traverse an environment without having to provide any additional input, something that less experienced gamers may struggle with.

Taking Time Crisis as an example of the mechanics, the game features two distinct phases within the game loop. First the camera, representing the player's point of view, automatically traverses part of an area such as a warehouse. Then



(a) Screen capture from Time Crisis Project Titan [34] showing the player in a hall way during an encounter section where the player is shooting at baddies.[2]

(b) Concept A2 storyboard sketch. A passive on-rail section is shown where the player traverses a hallway and sees a bike and distressed person

(c) Concept A3 storyboard sketch. The player avatar traverses an open nature environment during an "on-rails" section.

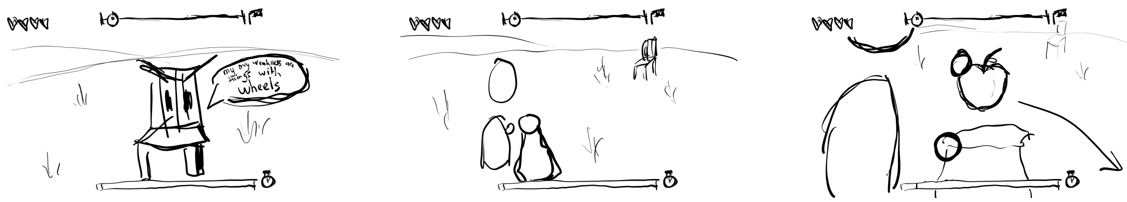
Figure 4.2: Comparison between the inspiration for Concept A2 and A3 and sketches visualizing the proposed design of A2 and A3.

at a certain point enemies will appear within frame, halting the progress of the player. The player must then shoot at the enemies to clear the area, allowing the player to continue, with the goal of completing the level in as little time as possible. I found these two alternating sections where the player is either passive or active to be a very interesting way to embed the task in a way that makes it feel as if the player is doing a lot, while mechanically they are only shooting at enemies and taking cover. The game itself takes care of the traversal and automatically moves the player towards their final goal, the completion of the level and the story. This traversal of an environment also allowed the player to visually perceive the progress that they were making towards the end of the level. For a screenshot showing an active section where the player encounters and shoots at enemies, see Figure 4.2a

A2 replaced the shooting and ducking during the active sections with a more realistic portrayal of the property verification task. The game simulated a commute from home to the workplace, with the player automatically traversing the environment during the passive gameplay section, aiming to reach their work in time and as quickly as possible. During the active section, the camera would stop moving and rather than finding a screen full of enemies, the player would encounter an object along the path. Then after a certain amount of time, a non-playable character (NPC) would appear, stating they had lost an item and describing one of its properties. The player's would then have to determine whether the previously seen object matched the described property and provide a response accordingly, representing a possible real life version of the original task. See Figure 4.2b for a sketch visualizing a passive "on-rails" section just before encountering an object and a person.

Similar to Time Crisis, poor performance such as providing incorrect answers resulted in temporary penalties, with NPCs obstructing the player's path and

[2] Image source: [35]



(a) 1. The player encounters an enemy (a chair), prompting the player with a vulnerability the enemy has to a specific property

(b) 2. The camera transitions from showing the enemy to showing the player's bag

(c) 3. The player avatar pulls out an apple from the magic bag, with the enemy in the background of the screen.

Figure 4.3: Concept A3 storyboard sketches.

delaying progress. To balance out the negative feedback that the player received from incorrect answers, a positive reward was added. Correct responses contributed to a combo mechanic, rewarding consecutive accuracy with increased movement speed. Different levels within the game would also portray locations closer and closer to the office, giving the player a sense of progression.

4.1.3 Concept A3

Concept A3 re-imagined the "on-rails" experience of Concept A2 by placing the player in a fantasy setting, where they embarked on a quest to recover a stolen lunchbox.

The active sections of the "on-rails" loop were distinct from the previous concept, taking further inspiration from RPG's such as the Pokemon [36] series where players would battle enemies in turn based battles. In these games, enemies typically have a designed weakness to a type of attack or item that would deal damage to them. The first person perspective was also replaced with a third person perspective with an avatar, which was more similar to the perspective offered in games like Pokemon as shown in Figure 4.2c.

Similar to the mechanics of A2, concept A3 would have the player encounter enemies in the form of a possessed object and engage them in battles that visually took inspiration from early generation Pokemon games, where the player avatar is shown from a third person perspective with the enemy in the background The enemy would first taunt the player by revealing a specific vulnerability within a speaking bubble (see Figure 4.3a), such as a weakness to "something that could be swung". This weakness was the representation of a property of an object from the conventional task. The camera perspective would then change to show the player holding a magic bag (see Figure 4.3b). The player would then repeatedly draw objects from this bag, holding an infinite amount of predetermined objects (see Figure 4.3c). By selecting the right objects that would match the property prompted by the enemy and throwing them at the enemy, the player could damage the enemy and ultimately defeat them.

The change to a battle driven context with enemies necessitated the re-implementation of a health system. Taking inspiration again from Time Crisis which used a discrete health system, a health system was added with a number of hearts representing the player health. Hearts were also chosen as they are commonly used all sorts of games ranging from popular classic titles such as the original Legend of Zelda [37], to more recent mobile games such as Candy Crush [38], making the meaning of hearts more understandable for many people. During the battle, if the player would take too long to decide what to do with an object, the enemy would counter-attack, dealing damage to the player. The battle would continue until either the enemy or the player sustained a certain amount of damage, leading to either victory and progression or failure and a restart.

Players started with four hearts, losing one per incorrect decision. Losing all hearts through failing encounters with enemies would result in a restart of the level, similar to the original Time Crisis. However, practically this would mean that players would have to repeat trials, and given that the encounters would have to be designed specifically with the different objects and properties, this would lead the players to have to repeat the same encounters. This would be quite inefficient for researchers as they would already have captured the data from those trials, making the restarted section obsolete in terms of data collection and only cause the experiment and game time to become longer.

To retain the illusion of progress while avoiding repeating encounters, the game would play an animation that communicated a restart, while resetting visual elements like progress bars to make the player think they had started over.

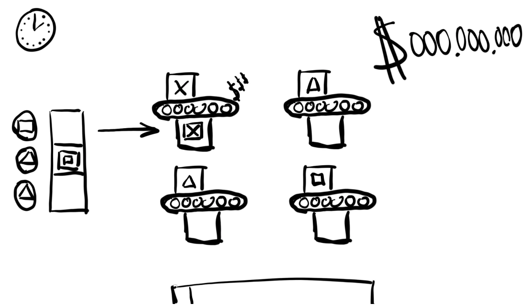
Notably, encounters did not align one-to-one with the conventional trials. Multiple trials could occur within a single encounter as opposed to in concept A2, presenting a greater challenge to the player, by needing to provide multiple correct answers to defeat an encounter.

Beyond combat, the game featured multiple biomes, acting as multiple levels to give a sense of progression. Additionally, to give the players a chance to recoup health, players could encounter neutral NPCs which granted the player an additional heart. For the sake of simplicity in the interactions and minimizing different interactions, these NPCs could be befriended through an inverted version of the combat mechanic. Rather than exploiting a weakness, players strengthened relationships by offering objects that did not match the NPC's stated vulnerability. As these NPC encounters differed in presentation from enemy encounters, it was suggested that they could serve as dummy trials, excluded from data collection in potential studies to ensure methodological consistency.

4.1.4 Concept A4

Concept A4 took a fundamentally different approach to embedding the game mechanics by integrating a minigame within a larger macro-game. This concept attempted to explore how the task mechanic could be embedded as a smaller part of a larger game. In other words, creating a more covert method of gauging

(a) Screen capture from Diner Dash [39] showing the main screen of the game where the player seats customers and serves them.[3]



(b) Concept A4 storyboard sketch. The main screen of the macro-game view overlooking multiple conveyors in the middle with robots represented with boxes

Figure 4.4: Comparison between the inspiration for Concept A4 and a sketch visualizing the proposed design.

the user's performance of the task, while presenting a fully fledged game that on the surface would look like a normal conventional game. This would also put the task in an overt gaming context, which I thought could be an interesting presentation for ecological validity.

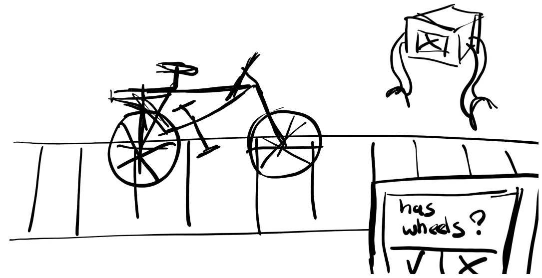
To this end, I took inspiration from a combination of two games that on the surface look widely dissimilar, but contain very similar mechanics: Diner Dash [39] and Yakuza 0 [40].

Diner Dash was a widely popular strategy and time management game franchise that saw the protagonist open a restaurant and manage shifts. During these shifts, players would have to assign guests to tables, and have the protagonist take orders, deliver food, etc. Guests will pay for their meal once they are done, and the total revenue will determine the score of the player, with a threshold serving as the success condition for the level. Different guests will order different items that will take longer to cook but may yield more money. Additionally, taking too long to seat or serve guests will result in the guests leaving, withholding their money, so players are encouraged to quickly and carefully assign guests based on their revenue and patience. A screen capture of the main game screen is shown in Figure 4.4a.

With the above, the design was set for an overarching macro-game. However, the embedding of the task was still unclear. The second inspiration, Yakuza 0 provided a solution to this issue. While Yakuza 0 on the surface appears like an action brawler game set in the criminal underworld of Japan, it contains multiple mini-games that can be accessed from the game's overworld. One of these is a management game where the player manages a cabaret club. Similar to Diner Dash, the player must assign employees to guests at different tables. However, unlike Diner Dash, guests generate revenue every few seconds while

[3]Image source: [41]

(a) Screen capture from Yakuza 0 [40] showing the mini-game where the player must recognize the right hand signal and choose the appropriate action.[4]



(b) Concept A4 storyboard sketch. Mini-game screen showing a trial where the player needs to decide whether the object matches the property on the tablet on the bottom right.

Figure 4.5: Left: screenshot from Yakuza 0. Right: Concept A4 sketch.

seated at the table. Guests and employees have distinct types, and placing employees at the same table as guests with the same type will result in a higher revenue throughout the guest's stay. However, the interesting part about this mini-game is that employees will occasionally require assistance, causing a pop-up to appear next to the table. When the player clicks on the pop-up, the camera transitions to a closeup of the table, showing the employee making a sign with their hand and a multiple choice prompt containing service actions, such as bringing more glasses, bringing a towel, bringing a menu, etc. The user is then tasked with interpreting the hand sign of the employee to determine which prompt should be pressed, with a correct answer providing a temporary boost in revenue production. See Figure 4.5a

This last mechanic introduces a test of the user's perceptive abilities, and could somewhat be likened to an experimental task. By combining this mechanic with the main mechanic from A2, seeing an object "in the wild" and having to determine whether it was the object described by the NPC, I could embed the task within the game in a way that made sense within the environment of the game, but was also distinct in its presentation, separating the macro-game and the task mechanic visually by having different user interfaces and camera angles.

To facilitate these mechanics with the task, I created a story which positioned the player as a factory overseer managing multiple conveyor belts to maximize revenue, instead of tables. In the macro-game, players assigned robots, similar to guests within Diner Dash and Yakuza 0, to conveyor belts, with each robot and conveyor having specific types like in the latter game. Matching a robot to a conveyor of the same type increased productivity and revenue. However, robots deteriorated over time and disappeared after being placed on a conveyor for a certain period. New robots continuously appeared on the side of the screen, requiring players to manage their available resources dynamically. See Figure

[4]Image source: [42]

4.4b for a visualization of the main screen for this concept.

For the context of the mini-game, robots on the conveyor belts would malfunction, temporarily halting production. To resolve this, players first had to examine an object that was produced on the conveyor. They were then shown a property of the object on an in-game tablet and had to approve or disapprove of it. Incorrect responses or failing to respond within a set time resulted in penalties, further delaying production. Conversely, correct responses temporarily increased the conveyor's efficiency, incentivizing accuracy and quick decision-making. For a visualization, see Figure 4.5b

Player progression was structured through levels, with higher levels introducing more conveyor belts and robots to oversee. Additionally, a high-score leaderboard encouraged competition, allowing players to compare their performance with previous participants.

4.1.5 Pre-focus group changes

These initial four game concepts were discussed with the two "expert" participants from the UT to evaluate their validity as a property verification task and identify key discussion points. The informal feedback consisted of the following points:

Concept A1

- Lack of ecological validity due to the abstract game environment.
- Lack of added value compared to the original task presentation.

Concept A2 and A3

- Mechanically similar.
- Potential inefficiency in presentation of task due to frequent "passive" sections, which could prolong the experiment.

Concept A4

- Player goals and stakes were more connected to the overarching macro-game and could take away attention from the mini-games, and thereby the experimental trials.

Overall, the "experts" noted that all of the concepts had merit and were suitable for the task structure. Subsequently, A2 was dropped due to the similarity to A3, limiting the possible directions that the stakeholder participants could chose from to just 3 options. Concept A3 was ultimately chosen over A2 because its context and setting more closely aligned with existing action-adventure games such as The Legend of Zelda: Breath of the Wild, which made it easier for people with game experience to relate the concept to conventional games.

4.1.6 Results meeting 1

Reference	Feedback	Changes in
RM1-1	Real world vs. made up properties	4.2.1 Mechanical design changes
RM1-2	Temporal separation of stimuli	4.2.1 Mechanical design changes
RM1-3	Conceptual bias with ”throwing”	4.2.1 Mechanical design changes
RM1-4	Clarifying property presentation	4.2.1 Mechanical design changes
RM1-5	Friendly encounters interfering with measurements	4.2.1 Mechanical design changes

During the first meeting, the ”researcher” participants noted similar concerns expressed during the pre-meeting discussion with the ”experts”. They concluded that Concepts A1 and A4 were less suitable for the task due to the lack of ecological validity and a fear of the macro-game overshadowing the task elements respectively. Subsequently concept A3 was eventually chosen by the participants to be continued in development. The following comments were given regarding the design of A3, which is from here on referred to as *the* game and named *Legend of the Lunchbox*, referring back to the story.

- **RM1-1**: The design did not clarify whether an enemy's vulnerability was logically derived from the object it represented or if it was randomly assigned. For example, if an enemy took the form of a chair, it was unclear whether its weakness should relate to its real-world properties or be arbitrarily assigned.
- **RM1-2**: The initial concept envisioned both the enemy (acting as the object within the task) and the property (drawn from the bag) appearing on screen simultaneously, albeit with one in the foreground and the other in the background. However, the ”researcher” participants raised concerns that this presentation could interfere with the necessary temporal separation required for EEG and fMRI analyses.

- **RM1-3**: The framing of the game's core mechanic required revision due to conceptual biases associated with the act of "throwing" objects from the bag. Certain objects naturally linked to throwing actions could unintentionally influence trial results. The same was said of "dealing damage" to the enemy, as this could introduce biases towards objects that could be naturally linked to dealing damage.
- **RM1-4**: Further clarifications were given on how sound and action-pantomime properties were presented. "Researcher" participants recommended that sound properties be represented by actual object sounds, while action-pantomimes should be depicted through animated avatars or hand movements rather than describing the action or sound in words.
- **RM1-5**: Friendly encounters could potentially evoke theory of mind processes, which is caused by the player associating with the player character/avatar and overlaps with the semantic networks that are studied in the task.

4.2 Milestone 2

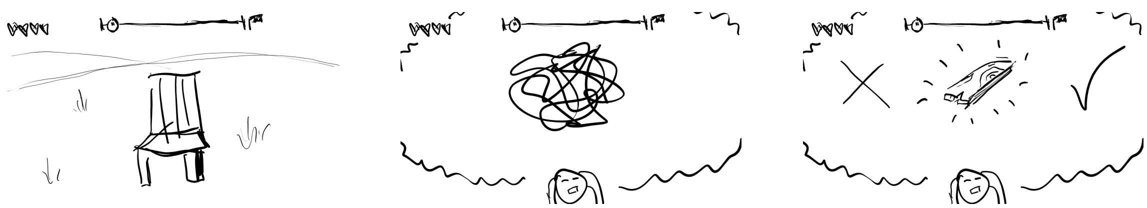
During the subsequent milestone, the design of the chosen concept was further expanded and a visual mock-up was created with the goal of presenting the most critical interaction of the game: the trials. The visual mock-up was presented as a low fidelity version of the game with temporary assets and showcased a passive "on-rail" section and the entirety of an enemy encounter.

4.2.1 Mechanical design changes

To better align the game with the original task, a structural change was implemented. Rather than assigning properties to enemies, the enemies themselves would represent objects, while properties would be drawn from the magic bag (**RM1-1**). If a property from the bag matched the enemy's object identity, the player would "throw" that property at the enemy.

The game's design was also revised to introduce temporal separation between object and property presentations using camera angles and transitions (**RM1-2**). The player would first be presented in close-up with the enemy object, then the camera would move to show the bag in close-up while omitting the enemy object in the background.

To mitigate the conceptual bias of the throwing mechanic and dealing damage to the enemy (**RM1-3**), the framing of the main mechanic was restructured, necessitating a narrative revision. The new storyline positioned enemies as objects "trapped" by the game's antagonist, a wizard. Instead of throwing objects,



(a) 1. The player encounters an enemy (a chair)

(b) 2. A gray thinking cloud appears on screen, and the player avatar is shown thinking of a property

(c) 3. The player avatar thinks of a property (in this case wood) and must choose whether it fits the object or not.

Figure 4.6: Updated storyboard sketches of *Legend of the Lunchbox* following the narrative framing changes.

the player would use "the power of their mind" to reconstruct the object by identifying its correct properties. The bag-drawing mechanic was re-framed as the player attempting to recall a property, aligning the game's mechanics more closely with the cognitive processes under investigation. To reinforce this shift, an animated panel first appears to symbolize the player character contemplating a property, which then disappears to reveal the identified attribute (see Figure 4.6 for sketches showing the updated sequence).

By correctly identifying properties, the player would restore the object and "free" it. Incorrectly identifying multiple properties or answering too slowly in enough trials would lead to a "loss" of the object. This change also led to the planned boss encounter at the end being removed, as it did not directly contribute to the experimental data and added unnecessary scope.

The above changes in the story made the mechanic of enemies attacking the player feel out of place. Thus, the mechanics were adjusted so that the depletion of the timer would not immediately damage the player. Instead, the player would continue seeing properties and determining whether they fit until all properties for that enemy were played through. Afterwards, the game would count the total number of correct or false properties and determine whether the player had met a threshold of correct answers to win the encounter. The game would then show different animations of the object depending on the outcome (see 4.2.2).

Explicit textual prompts for properties were removed (**RM1-1, RM1-4**). Instead, properties were represented with either an icon of a speaker accompanied by the sound of a property or a miniature of the player avatar miming the action.

The mechanics of friendly encounters were reworked after considering the intended length of the game. Players would likely have to play through many trials during the game, with one of the participants indicating roughly 300 trials, and previous studies using the conventional task contain around 192 to 256 trials [27], [28], [43] Given the high number of trials, I believed that

(a) Combat sequence from Wii Sports Resort [47] that influenced early the game's art style in early stages of development[5].

(b) Open-world landscape from The Legend of Zelda: Breath of the Wild (Nintendo EPD, 2017) serving as inspiration for the game's environmental design and later art style.[6]

Figure 4.7: Left: screenshot from Wii Sports Resort. Right: The Legend of Zelda: Breath of the Wild.

some variation was necessary to sustain engagement. To address this, a more distinct design of the friend encounter was introduced. Rather than increasing the number of different "alternate encounters," I chose to redesign the friend encounter itself to offer greater contrast to standard trials.

To accommodate the limited control scheme, inspiration was taken from button-mashing mini-games commonly found in party games such as the Mario Party series [44] and the "Test Your Might" mini-game in Mortal Kombat [45]. The revised mechanic required players to rapidly press the left and right keys to fill a progress bar. Once the bar was fully charged, the friend would be successfully befriended, providing a more dynamic alternative to the standard encounter format.

4.2.2 Visual design changes

Following the selection of A3 as the main concept, several visual components were further developed.

The game's art style was initially influenced by the simplistic yet quasi-realistic aesthetics of the Wii Sports series [46] (see Figure 4.7a), which was widely popular during the Wii's lifespan and accessible to players of all skill levels. This visual style complemented the embodied nature of those games, enhancing player immersion, which aligned with the aim of this project.

However, as development progressed, the art style evolved to reflect more modern design trends. The original aesthetic was deemed outdated and misaligned with the game's fantastical setting. Instead, inspiration was drawn from contemporary adventure titles, particularly The Legend of Zelda: Breath of the Wild [50] and Tears of the Kingdom [51], which feature stylized fantasy worlds

[5] Image source: [48]
[6] Image source: [49]

similar to the game's envisioned setting (see Figure 4.7b).

Additionally, several changes were made in response to feedback from focus group meetings.

Following the change in the story, the visuals for the enemy objects presented were changed to to be grayscale during encounters, but would regain their colors after the encounter had ended if the player successfully completed the encounter. This change was made to further reduce distractions during the trial, and support the updated story framing, with the returning of color to the object representing its reconstruction from its "trapped" form.

The visual representation of encounters and the player's cognitive shift was further developed. A graphic overlay was added during encounters to distinguish the "passive" sections of the game from the "active" sections. This element took the form of a gray cloud expanding from the player's head, resembling a thought bubble commonly seen in cartoons.

Two graphics with a check mark and cross were also added to the sides of the graphic for the property to help remind the player of the decision corresponding with each input. For example, to the left of the property was a cross corresponding to the choice to reject the property using the left key, while to the right was a check mark corresponding with the choice to accept it. To further strengthen this symbolism, properties would fly off-screen to the left or right based on the key that was pressed.

A visual feedback element was also added to encounters to indicate the player's responses to properties. After each trial, the game would show a green check mark for correct answers and a red check mark for incorrect ones. This provided direct visual feedback and gave users a sense of how they were performing trial by trial.

After encounters were completed, the game would show two different feedback animations based on the success state of the encounter. For won encounters, the object would twirl in the air and eventually fly up and away, leaving behind sparkles (see Figure 4.11c). Lost encounters would result in a camera shake, with the object falling out of frame onto the floor.

Three distinct environments were chosen for the backgrounds of different levels: a grassy meadow, a bustling castle city market, and the interior of a castle. During the development of the first biome, additional random events were proposed, such as discovering a campsite or passing through a village.

To reinforce the sense of progress and goal-oriented movement, the environment was designed so that during on-rail sections, the player could consistently see the distant castle they were striving to reach. This approach parallels the use of tall, visually striking landmarks in modern open-world games, such as recent Legend of Zelda titles, which naturally draw players' attention toward key destinations. As the game progresses, the castle gradually appears larger and closer, subtly signaling the player's advancement.

A progress bar was also displayed at the top of the screen to overtly indicate the player's location within the current level. This ensured that all players, regardless of prior experience, could easily understand their progression. The progress bar followed a conventional design, with a marker representing the

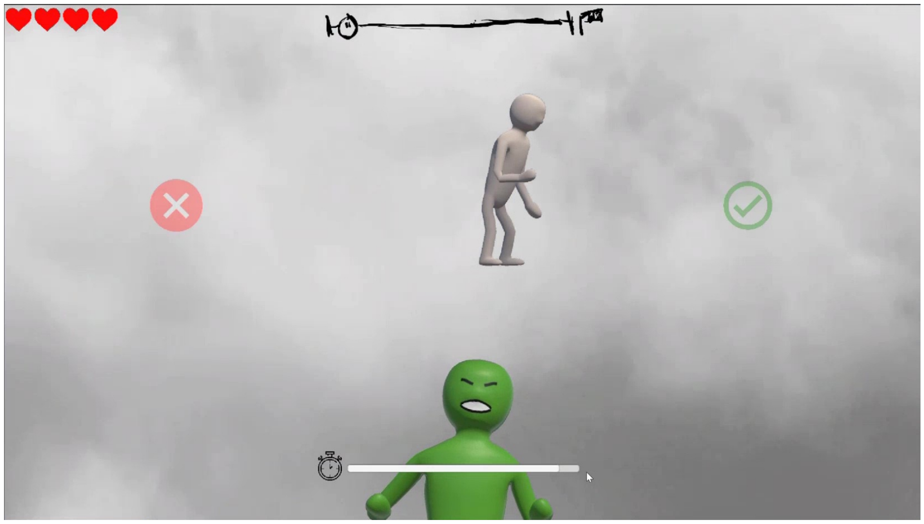


Figure 4.8: Screen capture from the low fidelity visual mock-up of the redesigned encounter analogous to the sketch shown in Figure 4.6c. An action property is shown where the player character is shown looking at a miniature version of the player character miming a hammering motion inside of a thought bubble represented by the gray cloud.

player's current position moving from left (start) to right (end). To further reduce ambiguity, a label reading "Progress" was added, reinforcing the bar's purpose and making it immediately interpretable.

After each level, the player was presented with a summary screen displaying their performance. The statistics shown were intentionally abstract and lacked detailed context to prevent players from focusing too analytically on their performance while reinforcing positive motivation. For example, it displayed the number of objects the player had freed without showing the total number of objects. Additional stats included the longest streak of consecutively freed objects and completion time. The player was also given a letter grade that started at A and could scale up to S or S+, which was originally inspired by tier lists and ranking from games that in hindsight are primarily found in Japanese games such as Metal Gear Solid [52] and the Dance Dance Revolution series [53].

4.2.3 Results meeting 2

Reference	Feedback	Changes in
RM2-1	Third person perspective invoking theory of mind	4.3.2 Visual design changes
RM2-2	Player avatar association effects	4.3.2 Visual design changes
RM2-3	Losing hearts and player motivation	4.3.1 Mechanical design changes
RM2-4	Direct feedback after trials	4.4.1 Mechanical design changes
RM2-5	Timer design during trials	4.3.2 Visual design changes
RM2-6	Distracting user interface during trials	4.3.2 Visual design changes
RM2-7	Extending graphics	4.3.1 Mechanical design changes

- **RM2-1**: The third person perspective and the inclusion of a player avatar of the game was theorized to engage theory of mind processes. The network of areas within the brain that process these processes could potentially overlap with the semantic areas that would be studied using the game. While this could have been acceptable given the objective of ecological validity, other issues with the usage of the third person perspective arose.

- **RM2-2**: The inclusion of a player avatar could create potential positive or negative associations between the player and the character, generating potential effects on player performance related to gender, age, or other personal characteristics.

- **RM2-3**: Concerns were raised about the impact of losing hearts and the health mechanic on player motivation.

- **RM2-4**: Direct feedback after each trial was nice and was accepted and used in conventional property task presentations. However, for cognitive-level analysis, which includes fMRI and EEG measurements, feedback could have an influence on the response of those measurements. Therefore, if you would have a trial directly after the feedback, there would need to be a significant delay until the start of the next trial. This did not apply for the feedback after the encounters, as those were followed by the passive section, allowing ample time for the response signal to reset before the start of the next encounter. Implementing a toggle was recommended to allow researchers to test the effect of the feedback.

- **RM2-5**: The timer used during trial presentations was too distracting as it appeared quite large in frame and was located below the property graphic, drawing attention away from the graphic during trials.

- **RM2-6**: In general, the user interface during trials was considered to be too populated, while those visible elements such as the health bar and progress bar were not directly necessary to the player during the trial screen.
- **RM2-7**: Validation studies would be required for the graphics of the objects and properties. It would be best to have some set of graphics that would be included with the game, while also being able to add new graphics to the game.

4.3 Milestone 3

Presenting the participants with the low-fidelity mock-up allowed me to collect their discussion on the mechanics of the enemy encounters. The following milestone would expand upon the overall graphics, the design of the game outside of the enemy encounters, and the technical implementation of the game.

4.3.1 Mechanical design changes

As the previous milestone focused on presenting the main interaction during "enemy" encounters, the feedback regarding the friendly encounter mechanic was updated in this milestone. To reduce the perceived severity of player punishment and enhance coherence with the updated framing, the health system, previously represented by hearts, was replaced with a more abstract, continuous bar symbolizing the player's imaginative abilities. In this revised system, failing encounters depleted the bar and subtly de-saturated the environment's colors, reinforcing the game's thematic elements in a less punitive manner (**RM2-3**). Similarly, friend encounters were redesigned to align with this new presentation. Friendly objects were replaced with vials containing "liquid imagination," which, when successfully opened by rapidly pressing input buttons, restored a portion of the player's imagination bar (**RM1-5**).

Concerns about punishment negatively impacting player motivation also led to modifications in the mechanics of the friend/vial encounter itself (**RM2-3**). Originally, if players failed to press the buttons quickly enough, the vial would remain sealed, preventing them from restoring their imagination bar. However, differences in skill level and dexterity among participants could create disparities in player confidence, leading to inconsistent experiences between more and less skilled players. To address this, a subtle design adjustment was implemented: regardless of actual input speed, the game provided auditory and visual feedback during button mashing and always rewarded the player by opening the vial. This ensured a more balanced and encouraging experience while preserving engagement.

During the previous meeting, participants discussed which objects and properties, and indirectly which encounters, should be included. Initially, the game was designed to be static, with 3D objects, animations, and sounds pre-generated and integrated into the game. These assets were meant to match the game's art style, and researchers could then specify which objects appeared as encounters, which graphics or sounds were presented as properties, and how many encounters and properties were included in each session.

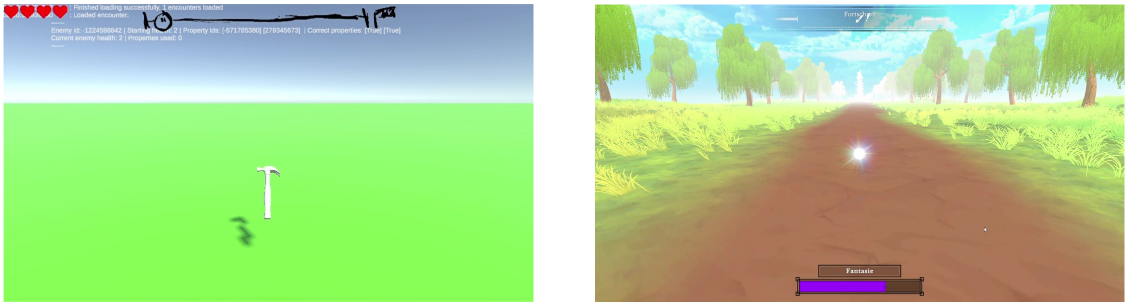
However, as development progressed, it became evident that both the objects and property presentations needed validation studies before they could serve as a reliable substitute for the property verification task. This posed a significant challenge: conducting a validation study within the scope of development was impractical. Additionally, restricting the game to a predefined set of objects and properties would limit researchers' ability to adapt it for different studies requiring alternative stimuli.

To address these issues (**RM2-7**), the game was redesigned to support the loading of external visual and audio assets. Users could copy flat image and audio data into a fixed folder within the root folder of the game, allowing them to be loaded into the game at runtime. The game would then match the file names with CSV files in which the user could specify the encounters' objects and properties. Recurring objects and properties could also use the same asset without requiring duplicates within the assets folder.

Given the need to represent action-pantomime properties, these visual assets had to accommodate both static visual properties and animated sequences. This requirement ruled out the direct use of 3D models, as integrating them would significantly increase the technical complexity of the project and require the future researchers to undertake 3D animation tasks, which would be difficult without prior expertise. Instead, the implementation was adapted to display objects and properties as either static images or sequences of images for animations. This approach preserved the flexibility to incorporate 3D elements when necessary, as researchers could hire third parties to render and export 3D animations externally before importing them as image sequences. As a result, the game maintained a 3D visual perspective while providing researchers with greater flexibility in adding new objects and properties. Audio assets for sound properties could also be included by providing WAV format audio files to the game.

To fill out the game with temporary assets, images and videos including 95 objects and 17 action-pantomime properties from a previous study from the participants were graciously provided by the "researcher" participants to be used for the purpose of this thesis. These assets were processed to remove the black background and turned into image sequences to match the game's input format.

Following an informal discussion with the "expert" participants, additional concerns were raised regarding the presentation order of the trials. Initially, the game displayed the object only at the beginning of the encounter, after which only its properties were shown sequentially. This posed potential issues for extended encounters, where players might need to complete five or more trials, requiring them to recall the original object across all trials. More critically, from an experimental standpoint, the varying time intervals between the object's ini-



(a)

(b)

Figure 4.9: Comparison between earlier version (a) of *Legend of the Lunchbox* where the object is already shown during world traversal, versus in the final version (b) where the object is obscured by a shiny graphic.

tial presentation and the subsequent property stimuli introduced inconsistencies in stimulus timing.

To address this, the design was first adjusted so that the object was always shown immediately before its corresponding property. Later, the initial in-environment presentation of the object was delayed entirely, revealing the object only once within the encounter. Instead, a generic shiny placeholder object remained on screen until the player initiated the trial loop, as seen in Figure 4.9. For a full comparison between the base property verification task and the updated gamified version presented in *Legend of the Lunchbox*, see Figure 4.10.

4.3.2 Visual design changes

The concept initially included a third person perspective that was intuitively chosen. This decision was influenced by a two factors. First, first person games are not commonly combined with validated task games, and conventional games that use first person perspectives are typically games that use embodied controls (such as VR) or games in the genre of simulation games or shooters. Second, a third person perspective allows for a virtual avatar that helps convey the idea of a game.

However, due to the concerns relating to the third-person perspective and the player avatar, the game design was modified to adopt a first-person perspective (**RM2-1**). Additionally, the original miniature player avatar, which was intended to mimic the player's actions, was replaced with a more neutral representation (**RM2-2**). A wooden drawing mannequin was selected as a substitute, as it lacked distinct characteristics such as gender or age, ensuring a more inclusive and universally relatable experience.

Despite concerns regarding the timer within the trial mechanic, removing the timer entirely was undesirable, as penalizing players for slow responses without providing a clear time limit could lead to frustration due to a lack of explicit

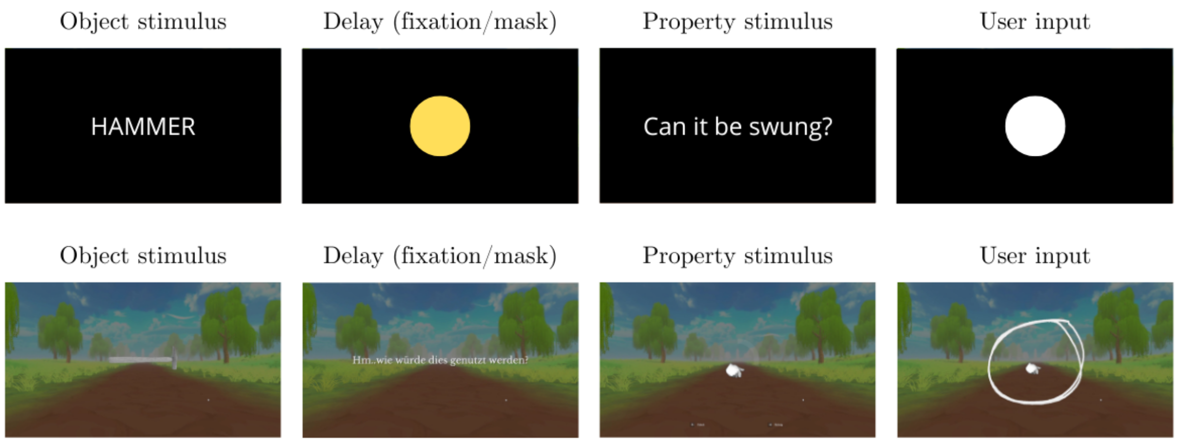


Figure 4.10: The base property verification task compared to the *Legend of the Lunchbox* property verification task.

feedback. To address this issue, the timer was redesigned as a circular element and repositioned in the background behind the property (**RM2-5**). This adjustment allowed players to keep their gaze centered on a single focal point while still perceiving the remaining time (see Figure 4.10).

The health bar was also redesigned from a discrete, heart-based system to a continuous "Imagination" bar to align with the game's revised narrative. This change served two purposes. First, it mitigated the punitive perception of damage, as hearts often symbolize vulnerability in games. Second, by removing discrete health units, the design discouraged players from prematurely restarting levels out of fear of failure. Inspired by mechanics in action games such as Metal Gear Rising [54] and Doom [55], the system could dynamically adjust difficulty by subtly reducing the amount of health lost when players were critically low.

The health bar and progress bar were also removed from the screen during encounters and trials to reduce distractions while players were performing trials (**RM2-6**).

To enhance clarity and reduce visual clutter, the brightly colored direct feedback graphics were recolored to a more neutral white. A short animation featuring splashing lines was added for correct answers to increase player satisfaction. Similarly, the overlay during encounters was revised to a rectangular, dark, transparent screen that flashed onto the display, replacing the previous gray cloud, which negatively impacted contrast. This update ensured that on-screen elements remained visible while maintaining environmental immersion. These updated graphics for the direct feedback can be seen in Figure 4.11.

Informal testing revealed that players needed additional guidance during trials. To address this, a prompt was introduced before each property's presentation, framed as an internal monologue encouraging the player to reflect

(a) (b) (c)

Figure 4.11: *Legend of the Lunchbox* provides players with direct positive (a) and negative (b) feedback after trials and a "reward" feedback (c) after encounters if the player got enough right answers. Direct feedback was changed later in development to be optional depending on research preferences.

on which properties might suit the object. Additionally, the left/right property indicators were removed due to poor readability, replaced by smaller yet clearer button prompts featuring a key icon and single-word instruction.

The encounter reveal sequence was also modified. Initially, encounter objects were hidden off-screen until the camera rotated to reveal them, but this provided no indication of when an encounter would start. To improve anticipation, encounters were redesigned as "collectibles" positioned along the player's path. These collectibles resembled glowing orbs, a common design in games such as in Yakuza 0 [40] or Dark Souls [56], which use neutral representations for discoverable items. When players reached a collectible, they would "pick it up," revealing the object and triggering the encounter.

As development progressed, the art style evolved beyond its initial Wii-inspired aesthetic. The original design felt outdated and misaligned with the game's fantastical setting. Instead, inspiration was drawn from contemporary adventure titles, particularly The Legend of Zelda: Breath of the Wild and Tears of the Kingdom, which feature stylized fantasy worlds similar to the game's envisioned setting.

4.3.3 Audio design changes

During this milestone, the sound design for the game was further developed. A primary concern was ensuring that in-game audio did not interfere with the perception of property-related sounds, thereby avoiding unintended priming effects.

During the on-rails sections, the audio elements featured ambient environmental sounds, the player's footsteps, and subtle whooshing sound effects to emphasize camera animations when encountering objects. Ambient sounds and footstep sounds were changed and audio processing effects were used per environment to fit the physics of the environment per level.

For the trial sections, feedback sounds were deliberately chosen to be abstract and object ambiguous, meaning that sounds were chosen that avoided obvious connotations with specific objects. To this end, sound effects that accentuated feedback consisted musical motifs or sounds found in nature. A soft click signaled the start of each trial, subtly drawing the player's attention. Positive feedback was conveyed through a synthesized rising musical motif, reinforcing a sense of achievement, while negative feedback was represented by a more abstract sound, loosely resembling a drop of water, to soften its connotation. Additionally, a subtle underwater ambient sound was introduced as background audio during encounters, creating a distinct contrast between passive on-rail and active encounter phases, with the underwater sound providing a background to the trials that was monotonous in tone and volume in an attempt to minimize its distracting capabilities.

4.3.4 Results meeting 3

Reference	**Feedback**	**Changes in**
RM3-1	Sound effects and distraction	4.4.1 Mechanical design changes, 4.4.2 Visual design changes
RM3-2	Environment variability	4.4.2 Visual design changes
RM3-3	Inner monologue as prompt	4.4.1 Mechanical design changes
RM3-4	Inter trial interval	4.4.1 Mechanical design changes

- **RM3-1**: During the intended cognitive experiments, researchers will continuously take measurements of players to be able to compare the measurements during the presentation of the task to the measurements outside of the task, known as the control condition. For academic task games, this means that any part of the game that falls outside of the trials within the encounters can be considered the control condition, and this condition must be relatively stable to allow for valid comparisons between these two states. The sound effects during the passive sections of the game were deemed quite loud and potentially too stimulating as a control condition of the study. If possible, it would be nice to be able to test the validity by implementing a toggle to turn it on or off.

- **RM3-2**: While currently not an issue, a discussion about expanding the background environment with more elements and strong topography revealed that the background should remain relatively monotonous to avoid cognitively stimulating the player in between encounters as that would vary the control condition of the study during gameplay.

- **RM3-3**: The overt "inner monologue" prompts seemed unnecessary, though participants agreed to add an option within the game settings to enable or disable the prompts for future testing. The text of the inner monologue was also a bit large, taking up most of

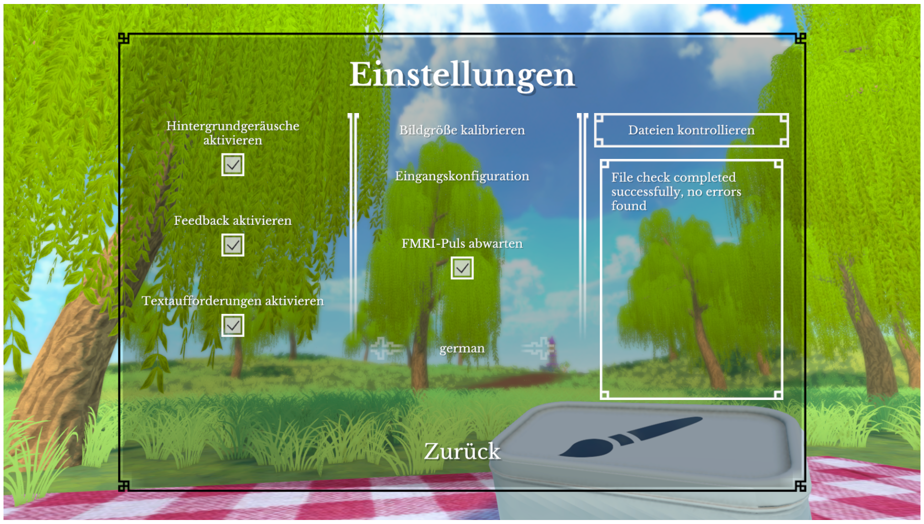


Figure 4.12: The settings menu in *Legend of the Lunchbox*.

the screen while the property graphic took up relatively less space.

- **RM3-4**: The game currently did not support a variable trial interval (delay between the start of one trial/property and the next), which would be desirable for fMRI measurement.

4.4 Milestone 4

The final milestone of the development put the last mechanical changes necessary for validity into effect. The main focus throughout the milestone was the finalization of the technical implementation and features outside of the game design, such localization and accommodating validation studies of the game.

4.4.1 Mechanical design changes

In response to concerns from the past two meetings regarding the duality of experiment validity with additional game elements, a settings menu was added to the design (see Figure 4.12). This menu allowed researchers to toggle specific elements that stakeholder were uncertain about in terms of their potential impact on the game's validity, enabling future validation studies on these elements. These options included the ability to:

- Disable all game audio except for property-related sounds to reduce cognitive load (**RM3-1**).

(a) An evil wizard challenges the player at the start of *Legend of the Lunchbox*.

(b) A tutorial with explanation text is shown in *Legend of the Lunchbox* after the opening cut scene. Here, the game explains the "imagination" bar.

Figure 4.13: Screen captures from the opening cut scene and tutorial section of *Legend of the Lunchbox*.

- Remove direct feedback after trials to mitigate concerns that incorrect answers could negatively affect player motivation (**RM2-4**).
- Omit the internal monologue before presenting properties to streamline the trial flow (**RM3-3**).
- Synchronize the game with external MRI pulses, ensuring that levels would only begin when the fMRI system was ready to capture measurements.

These configurable options provided researchers with greater flexibility, allowing them to tailor the game's settings to better suit the requirements of different experimental conditions.

To accommodate the customization of the levels, encounters, and the necessary localization for the game, a series of spreadsheet templates were designed to allow researchers to input the files in a way. The design of these spreadsheets was based off of spreadsheets provided by the CoPla that were already in use as input to python programs that they would use for their property verification tasks to allow for easy transfer of previous experiments. A functionality was also implemented within the game to check the provided input data from the researchers for common errors that would cause the game to crash or hang.

Variable delays were also implemented into the game logic, allowing researchers to set a specific delay per encounter through the spreadsheet templates (**RM3-4**).

Finally, a keybinding function was implemented, enabling lab technicians to reconfigure the game's controls. This feature allowed for compatibility with external hardware and different input methods for experimental setups.

To further sell the narrative of the story, an introductory and ending cut scene were added, explaining the story of game.

In the introductory cut scene, the player is placed in a field on a picnic blanket where the evil wizard appears. The wizard then challenges the player on their ability to imagine objects and determine whether properties are part of that object (see Figure 4.13a). They then proceed to show that they can turn objects into "stardust" (which is represented by the collectible found during levels), stealing the player's lunchbox in the process. The wizard then taunts the player and departs, with the camera showing the wizard's tower in the distance.

In the final cut scene, the player runs through the halls of the last level to find the wizard in a room at the end of the hall. Shocked by the player's ability, they admit defeat and return the lunchbox.

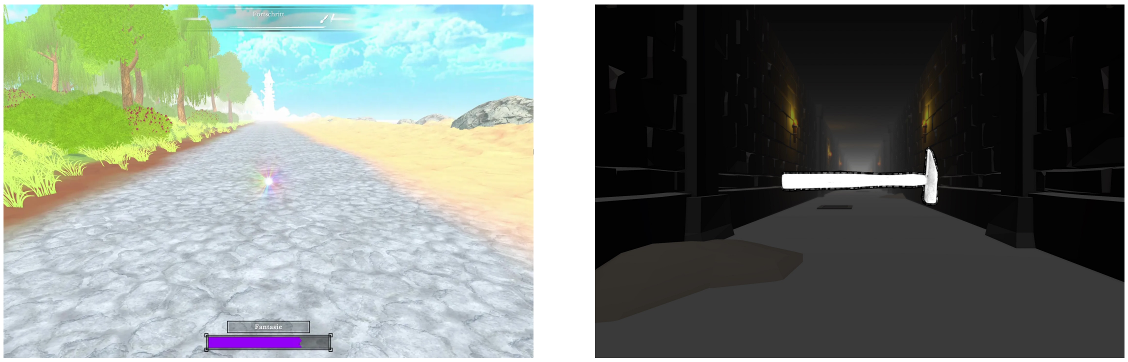
To help teach the players the mechanics of the game, the game also incorporates a tutorial level where the game showcases a few enemy encounters and a vial encounter. During critical moments, the game also pauses and presents the player with explanations for what is happening on screen, and instructing them on what to do (see Figure 4.13b).

4.4.2 Visual design changes

Following the concerns regarding the variability of the background environment, the design of the environments was simplified. While gradual environmental changes were acceptable, it was crucial that the environment within a level remained relatively consistent to maintain experimental control. As a result, the environments were designed to present a largely homogeneous setting, with random variations like tree placement to add subtle diversity (**RM3-2**). Additionally, the second biome underwent changes in response to the project's scope. A realistic city market, requiring numerous animated character models and detailed street elements, proved too complex to implement within the project's constraints. Consequently, the second biome was re-imagined as a lake bank, offering a variation that aligned more closely with the first biome's design. The third biome was also renamed to the wizard's tower to reflect the updated story. For previews of the graphics of these updated biomes, see Figure 4.15

4.4.3 Audio design changes

The volume of the effects during passive sections of the game was considerably lowered (**RM3-1**).



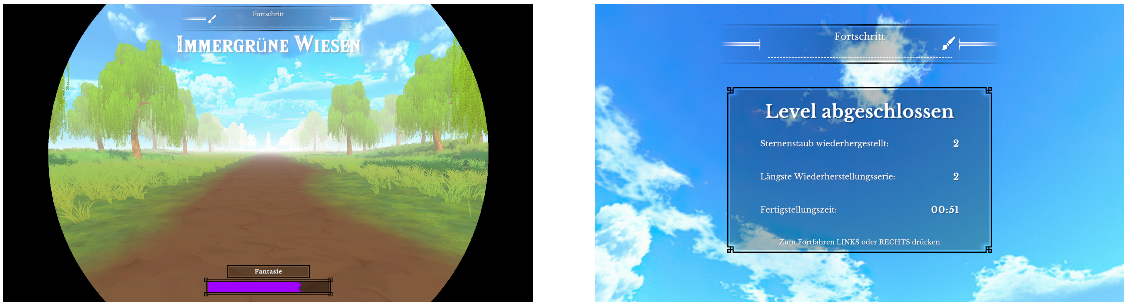
(a) (b)

Figure 4.14: The second (a) and third (a) biomes implemented for *Legend of the Lunchbox*.

4.4.4 Results meeting 4

Reference	Feedback	Changes in
RM4-1	Environments changing in between levels	4.5.1 Mechanical design changes
RM4-2	Choosing environment audio	4.5.1 Mechanical design changes
RM4-3	Location confusion	4.5.2 Visual design changes
RM4-4	Letter grade unclear	4.5.2 Visual design changes

- **RM4-1**: The difference in environment between levels could result in variable control conditions between levels. However, this was not necessarily seen as a problem as it could be balanced out by reversing the level order. Therefore, participants requested a change where the background environments could be swapped out for each level.
- **RM4-2**: One of the participants also commented on the ambient sound and wondered whether it would be possible to have these be chosen by the researchers, like with the audio properties
- **RM4-3**: The sudden environment change between levels caused some confusion on where they were and how this change in location related to the story.
- **RM4-4**: Finally, the letter grade on the end screen was not immediately understood and could be confusing to some.



(a) (b)

Figure 4.15: Final graphical design of user interface in *Legend of the Lunchbox* showing the inclusion of a text showing the current level location at the start of a level (a) and the removal of the letter grade in the level over screen (b).

4.5 Post-milestone 4

4.5.1 Mechanical design changes

During the last meeting, a discussion came up regarding potential variability in results across different biomes, as environmental differences could lead to differences in distraction and performance per biome. However, the participants found it valuable to investigate the potential effects of these variations as a controlled factor in future studies. As a result, an option was implemented to allow researchers to manually set the environment for each level, including the corresponding ambient sound for each biome. This feature provided greater experimental control, enabling studies on the influence of environmental context on player responses (**RM4-1, RM4-2**).

4.5.2 Visual design changes

To reduce player confusion in the first few seconds of a new level, a brief text animation was eventually incorporated at the beginning of each level to introduce the new location (see Figure **??**a), providing players with a smoother and clearer contextual transition while minimizing additional scope (**RM4-3**).

The letter grade from the level debrief screen was also removed (see Figure **??**b) to avoid confusing players unfamiliar with the letter grade system (**RM4-4**).

4.6 Requirements Validity Results

Participants were overall satisfied with the final implementation of the game and the development process. The graphics and faithfulness to the original task were praised in particular, with participants noting that they would provide a more interesting environment for players.

Regarding the fulfillment of the requirements, the following was said.

R1: The game was considered to be capable of acting as a property verification task, presenting the task in an interesting context while essentially retaining the same structure of the original task.

> "now it's ... similar to property verification tasks that we know."

R2: The presentation of the task, with players traversing an environment and completing encounters along the way was considered to adequately add ecological validity for the player.

> "[It] is like an actual interaction game where you move through a world, you encounter real objects and then you interact with them in a way. That also gets closer to kind of the ecological validity you know? ... interacting with that actual object"

> "It gives enough of a game feel to me, and some natural validity, but not too much at the same time."

R3: The validity of task was discussed at length during meetings and many elements of the design's mechanical, visual, and audio elements were updated to finally accommodate a valid presentation of the task.

> "one makes compromises at the end of the day, of course, right. To also get ... the experimental control back that we often want in psychological experiments"

> "If it's like very much like a game, and there's a lot of things that are not controlled, then one has always the problem with ... experimental control later. I think [now] it it looks cool and looks similar to what we were now imagining [not too much like a game]."

Apart from these requirements, visual and auditory stimuli were present within the game, and several changes were suggested by the participants and implemented to ensure that users could post-hoc add custom stimuli, including images, animations, sounds, and text.

The game also exported data in a format that was based on the output of existing programs that implemented the task.

The graphics design and implementation of the game were designed to minimize the performance cost on machines. Though the performance of the game was not tested by the participants before the end of the final meeting, participants informally replied to the performance after the meeting. The game was tested on machines at the CoPla, where researchers noted that the Windows build of the game ran well. The additional Linux build unfortunately performed less well, though the cause has not yet been investigated.

Finally, a localization functionality was present which provided researchers with a German translation and allowed them to edit text elements within the game itself in case they wanted to change the chosen translations.

Chapter5

Discussion

The development of this project has provided valuable insights into the complexities of academic game development, particularly the balance between maintaining experimental integrity and incorporating engaging ludic elements.

Despite these challenges, this project has yielded several noteworthy observations, particularly regarding the role of ludic elements in experimental tasks and the obstacles encountered in collaborative academic game development. The most significant findings are discussed in the following sections.

I would like to preface this discussion by noting that these findings should be interpreted with caution, as replicating the design and development approach of a successful project does not guarantee the same outcome. Player experiences can vary. Factors, such as personality, game preferences, game experiences, and perceived performance can have an impact on the player's experience [57], [58]. Nevertheless, the design, rationale, and challenges of this game may still help future research to better understand academic games as a whole as well as inform design decisions of future works.

5.1 Ludic Elements in Academic Game Development

The design of this game was greatly shaped by the requirement to adhere closely to the mechanics of the original experimental task. Stakeholders frequently requested modifications that limited the implementation of ludic elements commonly found in conventional games. As a result, the final design largely preserved the structure of the original experiment, in which users were sequentially shown object stimuli, property stimuli, and then prompted to provide an input. Due to these constraints, the game relied primarily on visual presentation to create the illusion of a game, a strategy to enhance ecological validity that is validated by other academic task games [6]–[9]. Additionally, aesthetic enhancements, such as additional animations or dramatic visual effects, were deliberately restrained.

This approach focused on adding game elements to the structure of the original task, similar to the works mentioned in 2.2. However, it is important to note that within academic game development, the opposite approach can be used as well. One example of such a work that appeared during the writing of this thesis is Tunnel Runner by Markovitch et al. [59], an endless runner that incorporates four psychology tasks and aims to "*stop imitating cognitive tasks and instead use cognitive tasks as inspiration for gameplay mechanics*" [59, p. 3]. Both approaches have shown results that are capable of gathering valid results, albeit with different issues. Tunnel Runner provides better player experience across participants and produces internally valid results. However, the effect of some of the tasks on response time was different when compared to the conventional task and there were several limitations to the study that could have put the results into question, such as players adopting response strategies like preemptively pressing keys and experimental design potentially influencing the results.

As mentioned in the introduction, this issue is particularly pronounced in cognitive psychology, where task validity is highly scrutinized [7], [9], [15], [16], [18]. Any deviation from the original task presentation may result in skepticism in academic publications, and differences in results compared to prior studies risk being dismissed as artifacts of an altered setup. With this in mind, it further exemplifies why current psychology researchers may be hesitant to adopt full-on ludic games as experimental tasks and may prefer to focus on "turning tasks into games", rather than "turning games into tasks".

5.1.1 Types of experimental concerns

Looking at the type of feedback found throughout the milestone meetings, roughly four types of design concerns were addressed specifically regarding a direct impact on experimental validity. These include: design that deviated from the research goal, design that introduced biases to the task, design that distracted from the task, and design that caused variability within the control condition.

Design that deviated from the research goal included design that deviated either from the structure required for the task or processing of the collected data, or otherwise impacted the player in a way that conflicted with the research object. This was exemplified by the original third person perspective and inclusion of the friend mechanic which could have invoked theory of mind processes which overlap with the semantic networks studied and measured during the game.

- **RM1-2** Temporal separation of stimuli
- **RM1-4** Clarifying property presentation

- **RM1-5** Friendly encounters interfering with measurements
- **RM2-1** Third person perspective invoking theory of mind
- **RM2-4** Direct feedback after trials
- **RM3-3** Inner monologue as prompt
- **RM3-4** Inter trial interval

Design that introduced biases to the task included design that could potentially have introduced biases for the player, impacting the results of the task. An example of this would be the initial framing of approving a correct object by "throwing it" which could create a bias if the object is already associated with a throwing action.

- **RM1-1** Real world vs. made up properties
- **RM1-3** Conceptual bias with "throwing"
- **RM2-2** Player avatar association effects

Design that distracted from the task included design that was found to be too distracting from the main task, potentially impeding performance. This is exemplified by the removed prominent UI elements that took up screen space and could have attracted the gaze of the player while the main task was being performed.

- **RM2-5** Timer design during trials
- **RM2-6** Distracting user interface during trials
- **RM3-1** Sound effects and distraction

Design that caused variability within the control condition included design that affected any gameplay outside of the interactions of the main task in a way that it could potentially invalidate the control condition required for the research. One such example was the sound design of the environment being too loud in some places which could stimulate the user too much and irregularly during the control condition.

- **RM3-1** Sound effects and distraction
- **RM3-2** Environment variability
- **RM4-1** Environments changing in between levels

Other feedback also included aspects outside of experimental validity, such as feedback on design regarding game motivation and accessibility (**RM2-3**,

RM4-3, **RM4-4**) or extending the the customization of the game (**RM2-7**, **RM4-2**).

The above context sets the stage for the main challenges of this project and, by extension, of academic games that aim to introduce ludic elements to validated experimental tasks rather than the other way around. While feedback regarding design that deviated from the research goal was typically incorporated quite easily by adding features to fit the required structure or removing features that conflicted with the research goal, feedback regarding minimizing distractions, biases, and variability presented a less binary problem. This resulted in a challenge of how ludic elements could be introduced while minimizing distractions and biases, as well as how variability in the task or control condition of the task could be minimized. This challenge was generally addressed with two strategies during development: using specific design and hidden manipulations.

5.1.2 Specific design

While global ludic elements as a whole have previously been validated or partially validated by researchers [7], [9], [14]–[18], the impact of how specific elements are designed are rarely discussed.

During the implementation of the environment, I believed that randomly procedurally generated terrain being different per participant could raise concerns that the control condition for each participant would have been different. This assumption led to a change where the environment was still generated procedurally, but in the same way for all players. Furthermore, the variability between the randomly generated environment chunks was minimized, with environments maintaining similar geometry and colors over time.

This decision to limit the impact of randomization was made preemptively to limit the concerns that were discussed at the start of this section. However, researchers may also assume that certain game elements may have no effects whatsoever. As exemplified by Tunnel Runner, completely altering the presentation of the task to an environment that includes colored graphical elements, 3D objects, and varying environments, can still result in (partially) valid data output [59]. This suggests that perhaps the design of specific elements itself can have an impact on the validity of the game as a task, rather than a ludic element as a whole.

This phenomenon was also found during the course of this development. The design and presentation of elements, such as the timer shown during the trials, the background environment during the trial, the volume of sound effects during the "on-rails" sections, and the framing and narrative of the game were all changed in their design throughout the development so that they would limit the distraction from the task and avoid potential biases.

Taking the trial mechanic as an example, the initial design originally consisted of a horizontal bar below the stimulus that would become less wide as it depleted, while the new design showed the timer as a circle behind the stimuli that would fill up radially. Both of these designs included a timer, but experts found the latter design less distracting than the former.

Another example can be found in the narrative framing of the game, which was changed to avoid inducing conceptual biases in the player. The player originally threw objects at enemies that prompted the tested properties. However, the "researcher" participants feared that objects that were meaningfully related to the action of throwing would bias the response of players during the experiment. In this case, the specific design of the narrative was in conflict with the task itself, though changing the narrative to reduce the relation between the framing and the experimental task resolved this concern.

These examples have shown how different ludic elements in itself did not produce perceivable issues in the validity of the experiment, but how the specific design of those elements did produce concerns regarding validity of academic games in general (distraction from the task), and validity of the specific research that the academic game served (biases in the task). Furthermore, by changing the specific execution of those elements, I was able to address those concerns. Future work may therefore benefit from a way to categorize and evaluate ludic elements in tasks, not just as a whole, but based on their specific design and their impact on the research goals.

5.1.3 Hidden Manipulations

Limiting potential variability between players was mainly done through hidden manipulations of the game state. Fail states were removed in the health mechanic and the vial encounters to eliminate the possibility of players becoming demotivated due to failure. The health-bar was manipulated so that it would never deplete entirely. Furthermore, the encounters with the vials would prompt the user to mash the controls of the game and would play sound effects and animations with the inputs of the player to give the illusion that they were building up strength to open the vial. However, the vials would always open after a certain amount of time. These design choices helped standardize the gameplay experience, reducing variability in experiment duration and preventing differences in player motivation from affecting results, while also providing an illusion of stakes that would not affect experimental validity.

However, while these hidden manipulations preserved the illusion of gameplay for less experienced players, more experienced players may have recognized the lack of genuine challenge. Additionally, the above approach raises concerns about re-usability in repeated experiments. Participants who play multiple times may realize that certain actions have no real consequences, leading to behavioral changes that could be difficult to predict and control. Wiley et al. previously recommended that games should align with participants' expectations [16]. However, their study found that participant engagement remained higher in the game condition that only added narrative elements compared to a traditional presentation of the task, even when the game's mechanics did not fully meet the expectations set by the narrative. Moreover, performance between the two groups were comparable. This finding was supported by a recent study by Held et al. [60] and suggests that even games with minimal interactive elements, such as those relying primarily on narrative and context, can still help enhance engagement and immersion. Therefore, elements that give the appearance of ludic games, such as hidden manipulations, may not necessarily be detrimental when considering studies where experiments only need to be performed once. However, this approach may come at the cost of reusability, as repeated exposure to the same game could reduce its effectiveness. More importantly, if the widespread use of minimally engaging academic games becomes common practice, it may contribute to negative perceptions of such games, potentially leading to biases that could preemptively impact participant performance in future studies.

Modulating difficulty in an academic game presented further challenges. Any effective difficulty-adjustment system would require rigorous testing across multiple iterations to counteract learning effects over repeated playthroughs. Fine-tuning difficulty parameters while maintaining experimental validity would significantly extend the development timeline and require substantial resources. Given these constraints, dynamic difficulty modulation was ultimately not pursued, reinforcing the notion that balancing ludic elements with experimental control remains a challenge in academic game development. Even so, modulated difficulty could also raise issues regarding the experience variability.

Every task, experiment, and study is different and may require different priorities when designing an academic game. In this particular study stakeholders wanted to focus on ecological validity rather than engagement. However, focusing on engagement over ecological validity may be as justified when considering the purpose of the research. As such, different game elements that may not have been appropriate for this project, such as point systems, could be a valid approach in different academic games.

5.2 Challenges in the Academic Development Process

The development process itself was shaped by various constraints, including practicality, stakeholder involvement, and the difficulties of aligning game design with experimental requirements.

One of the early challenges encountered was communication with stakeholders and participants, particularly those with limited experience in game development. Differences in familiarity with game mechanics and terminology led to discrepancies in envisioning the final product. "Researcher" participants, who had little prior exposure to games, sometimes struggled to fully grasp gameplay concepts when presented with low-fidelity representations such as storyboards and were confused on polysemous terms such as "concept", which is used in both game and psychology fields, but have different meanings in their respective fields. This resulted in misunderstandings, with stakeholders making false assumptions that later confused discussions and resulted in mid-development corrections on earlier established design decisions.

For example, during the first two milestones, no specific feedback was provided regarding the order of stimuli presentation. It was only in the third milestone, after significant development had already taken place, that this issue was raised, necessitating adjustments that could have been avoided with clearer early-stage communication. This illustrates a broader challenge: neither stakeholders nor developers could easily anticipate how certain ludic elements might impact experimental control, underscoring the importance of an iterative development process with low-fidelity gameplay prototypes to facilitate clearer feedback. It also emphasized the importance of the pro-active preparation of questions during meetings to raise and discuss potential issues that stakeholders had overlooked, as the problem with the presentation order was initially not identified by the participants and was only discussed during the milestone meetings after I brought it up in a discussion.

On the other hand, this project also highlighted the challenges of academic game developers without any prior experience in the target field. For this project, understanding of cognitive psychology practices was enhanced by the "researcher" participants providing relevant readings which included examples of conventional presentations of the task. This also helped establish the format of the output of the game, and how researchers set up the trials. By providing these examples, the format of the input and output of the game could be designed in a way that was close to how the researchers would format the input and output of the game in previous studies. However, this may have been a lot harder without the inclusion of these previous works and examples.

As the project progressed, significant changes were made to the game's core design due to the these challenges. The original vision aimed to evoke a sense of adventure and the thrill of defeating enemies, but later adjustments altered this fundamental experience. The original combat mechanic, throwing objects at enemies, was replaced with a more abstract interaction. Additionally, the visual and audio feedback was toned down to ensure experimental validity. These changes substantially altered the game's feel, moving it away from its initial vision. A more structured approach, where the final design was established earlier in the process, might have allowed for a more integrated design of mechanics within the game's context, rather than implementing changes retrospectively at the cost of already developed content. This may help prevent losing the intentionality of the design throughout implementation. One alternative approach could involve redistributing meeting schedules: rather than evenly spaced meetings across the project timeline, the initial phase could feature frequent, intensive meetings to thoroughly define the game's mechanics and visuals, reducing the need for later course corrections. The challenge, however, lies in achieving sufficient design detail within limited time and balancing that out with the ability to iteratively test mechanics through early implementation.

Furthermore, the strategy of presenting multiple concepts with increasing levels of gamification was helpful in determining not only what level of gamification the participants were comfortable with, but also which specific problems can be found in the design of each concept. This information could then be applied in future changes of the chosen design. One example of this could be found in the discussion of 4.1.4 during Milestone meeting 1, where participants noted that the reward of player's gaining money in the game could affect conceptual processing, which they specifically wanted to study using the game. This informed the refinement of the final game such as omitting overt reward mechanics.

Ultimately, academic game development places a strong emphasis on experimental validity over ludic engagement, presenting unique constraints compared to commercial game development. To bridge this gap, new design tools and methodologies may be required to accelerate high-fidelity prototyping, enabling stakeholders to better visualize gameplay experiences before full-scale development begins. Developing such tools could help academic game developers balance experimental rigor with engaging gameplay more effectively in future projects.

5.3 Limitations

This work is to some extent limited by the constraints of the collaborative development. The "researcher" participants of this research ultimately set the validity requirements of this work. This may have steered the development in a specific way which more closely resembles that of previous studies. Furthermore, this work does not contain a user study that could have quantitatively proved the efficacy of the design. While this is omitted, I believe that the evaluation of the game by the participants was a sufficient measure for the feasibility of using the game as a valid approach to the property verification task. Furthermore, this is a more realistic measure of success, as development of academic games is futile without willingness from researchers to apply these games for their work.

Chapter6

Conclusion

For this thesis I have performed an extensive development and documentation process with the aim of exploring how game design can be incorporated into the collaborative development of an academic validated task game. To this end, a property verification game has been developed in collaboration with a number of "expert" and "researcher" participants that satisfied the requirements set by the participants. The result of the study, the documentation of the participant's discussion regarding the game and reflection on the development process, has revealed several key takeaways.

First, based on the feedback from the "researcher" participants, I believe that the addition of narrative elements and a game-like context such as 3D environments and game-UI to a cognitive task may help generate ecological validity without impeding the validity of the results. Furthermore, the placement, visibility, and specific design of visual elements should be carefully designed to be able to provide this context while minimizing distractions from the main task. Narrative framing should also be checked for potential conflicts with the research goal, such as biases towards concepts, or interfering with the researched area of cognition.

Second, designs must carefully consider that variability between the experience of participants may cause concerns regarding comparability between participants. To this end, designing the game to be linear or using hidden manipulations to force the game state to be independent of the player's actions could help to limit this variability. Developers and researchers should also consider the use case for the game when developing using the previous strategies, as motivation of the participants in repeat experiments may suffer from negative biases.

Regarding the collaborative development process, I offer the following recommendations.

To gauge how comfortable researchers may be with game elements, offer several concepts with differing levels of "gamification". This process may help understand the stakeholder's sentiments towards specific mechanics, informing further design.

Developers may benefit from the use of low or high fidelity visual mock-ups, such as storyboards or video renders to help explain the mechanics of the game to stakeholders. Higher fidelity mock-ups may also help the stakeholder visualize the design of specific visual elements which helps them identify potential issues regarding the validity of the task.

Pro-actively preparing discussion points, such as discussing specific elements of the design that the developer suspects may influence the research validity, may help researchers identify validity issues that would otherwise go unnoticed.

Two key areas were identified where future works could benefit similar projects in the future. First, by investigating ways to more efficiently communicate design to stakeholders at the start of the development process, developers may be able to spend less time on design and more time on implementation, allowing for higher quality games. Second, academic task games currently lack understanding of specific designs. That is, certain game mechanics, such as points, are investigated on a macro-scale and researchers draw inconsistent conclusions when examining the effect of these mechanics. This invalidates choosing a mechanic for its specific effects as a sure-fire strategy. Therefore researchers would benefit from studies that examine the specific execution of a design on a micro-scale to investigate how the differences in designs of the same mechanic may affect results.

Chapter7

Tools and AI Acknowledgments

To ensure transparency with regards to the use of tools like AI and other software, this section exists to declare all software tools used for the duration of the Thesis and Research Topics.

During the preparation of this work, I used ChatGPT to rephrase sections of my thesis and aid the readability and structure of these sections. After using this tool/service, I thoroughly reviewed and edited the content as needed, taking full responsibility for the final outcome. I also used ChatGPT to summarize and explain unclear concepts within existing research papers. I also reviewed and verified the information I gained through this method, taking responsibility for any inaccuracies that may have ended up in the report. Furthermore, I used the transcription functionality of Microsoft Teams to transcribe the contents of the meetings that contribute to the results, After using this tool/service, I thoroughly reviewed and edited the content as needed, taking full responsibility for the final outcome. Finally, I used DeepL to help translate parts of the game text for localization purposes. After using this tool/service, I thoroughly reviewed and edited the content as needed, taking full responsibility for the final outcome.

- ChatGPT (Accessed 9/2024 until 3/2025)
- Microsoft Teams 2024/2025
- DeepL Translator
- ATLAS.ti 2024
- Unity 2022.3.52f1
- Jetbrains Rider 2024.2.7
- Audacity 3.2.5

- Blender 4.2
- Paint.NET 5.1.4

Bibliography

[1] M. Ma, A. Oikonomou, and L. C. Jain, "Serious games and edutainment applications," vol. 504, 2011. DOI: 10.1007/978-1-4471-2161-9.

[2] E. Donchin, "Video games as research tools: The space fortress game," *Behavior Research Methods Instruments & Computers*, 1995. DOI: 10.3758/BF03204735.

[3] M. A. Gómez-Maureira, M. v. Duijn, C. Rieffe, and A. Plaat, "Academic games - mapping the use of video games in research contexts," *International Conference on Foundations of Digital Games*, 2022. DOI: 10.1145/3555858.3555926.

[4] J. Huizinga, *Homo ludens: A study of the play-element in culture*. Martino Publishing, Sep. 2014.

[5] M. M. Morgan, M. J. Christie, L. De Lecea, *et al.*, "Stop-signal task," *Encyclopedia of Psychopharmacology*, I. P. Stolerman, Ed., pp. 1285–1285, 2010. DOI: 10.1007/978-3-540-68706-1_1120. [Online]. Available: https://doi.org/10.1007/978-3-540-68706-1_1120.

[6] A. Lieberoth, "Shallow gamification," 2015. DOI: 10.1177/1555412014559978.

[7] M. A. Friehs, M. Dechant, S. Vedress, C. Frings, and R. L. Mandryk, "Effective gamification of the stop-signal task: Two controlled laboratory experiments," *JMIR serious games*, 2020. DOI: 10.2196/17810.

[8] P. Schroeder, J. Lohmann, and M. Ninaus, "Preserved inhibitory control deficits of overweight participants in a gamified stop-signal task: Experimental study of validity," *JMIR serious games*, 2021. DOI: 10.2196/25063.

[9] H. Kirsten, M. Dechant, H. Gibbons, and M. A. Friehs, "Tasting inhibition: A proof-of-concept study of the food stop-signal game," 2023. DOI: 10.1016/BS.PBR.2022.12.002.

[10] K. Yoder and M. Belmonte, "Combining computer game-based behavioural experiments with high-density eeg and infrared gaze tracking.," *Journal of visualized experiments : JoVE*, 2010. DOI: 10.3791/2320.

[11] I. Kniestedt, I. Lefter, S. Lukosch, and F. M. Brazier, "Re-framing engagement for applied games: A conceptual framework," *Entertainment Computing*, vol. 41, p. 100 475, 2022.

[12] M. A. Gómez-Maureira, T. Yeung, and M. A. Friehs, "Mapping academic game development: An exploratory investigation," *Foundations of Digital Games*, 2025. DOI: 10.1145/3723498.3723799.

[13] K. Allen, F. Brändle, M. Botvinick, *et al.*, "Using games to understand the mind," *Nature Human Behaviour*, vol. 8, no. 6, pp. 1035–1043, Jun. 2024. DOI: 10.1038/s41562-024-01878-9.

[14] C. Groening and C. Binnewies, ""achievement unlocked!" - the impact of digital achievements as a gamification element on motivation and performance," *Computers in Human Behavior*, 2019. DOI: 10.1016/j.chb.2019.02.026.

[15] K. Wiley, P. Berger, M. A. Friehs, and R. Mandryk, "Measuring the reliability of a gamified stroop task: Quantitative experiment.," *JMIR Serious Games*, 2024. DOI: 10.2196/50315.

[16] K. Wiley, S. Vedress, and R. Mandryk, "How points and theme affect performance and experience in a gamified cognitive task," *International Conference on Human Factors in Computing Systems*, 2020. DOI: 10.1145/3313831.3376697.

[17] A. Khaleghi, Z. Aghaei, and M. A. Mahdavi, "A gamification framework for cognitive assessment and cognitive training: Qualitative study.," *JMIR serious games*, 2021. DOI: 10.2196/21900.

[18] N. Markiewicz, M. Russa, A. Fokkens, M. Dechant, and M. A. Friehs, "You got it in your hands: Stop-signal modality influences on reactive response inhibition with gaming controls," *International Journal of Human–Computer Interaction*, pp. 1–10, 2023.

[19] C. Johanson, C. Gutwin, and R. Mandryk, "Trails, rails, and over-reliance: How navigation assistance affects route-finding and spatial learning in virtual environments," *International journal of human-computer studies*, 2023. DOI: 10.1016/J.IJHCS.2023.103097.

[20] N. Markiewicz, M. Russa, A. Fokkens, M. Dechant, and M. A. Friehs, "You got it in your hands: Stop-signal modality influences on reactive response inhibition with gaming controls," *International Journal of Human-Computer Interaction*, 2023. DOI: 10.1080/10447318.2023.2285624.

[21] M. J. Apter, "Reversal theory and personality: A review," *Journal of Research in Personality*, 1984. DOI: 10.1016/0092-6566(84)90013-8.

[22] S. Putze, D. Alexandrovsky, F. Putze, S. Höffner, J. D. Smeddinck, and R. Malaka, "Breaking the experience: Effects of questionnaires in vr user studies," *International Conference on Human Factors in Computing Systems*, 2020. DOI: 10.1145/3313831.3376144.

[23] V. Schwind, P. Knierim, N. Haas, and N. Henze, "Using presence questionnaires in virtual reality," *International Conference on Human Factors in Computing Systems*, 2019. DOI: 10.1145/3290605.3300590.

[24] J. Lumsden, A. Skinner, A. Woods, N. Lawrence, and M. Munafo, "The effects of gamelike features and test location on cognitive test performance and participant enjoyment," *PeerJ*, 2016. DOI: 10.7717/peerj.2184.

[25] P. Kuhnke, M. Kiefer, and G. Hartwigsen, "Task-dependent recruitment of modality-specific and multimodal regions during conceptual processing.," *Cerebral Cortex*, 2020. DOI: 10.1093/cercor/bhaa010.

[26] P. Kuhnke, M. Kiefer, and G. Hartwigsen, "Task-dependent functional and effective connectivity during conceptual processing.," *Cerebral Cortex*, 2021. DOI: 10.1093/cercor/bhab026.

[27] R. Bruffaerts, P. Dupont, S. D. Grauwe, *et al.*, "Right fusiform response patterns reflect visual object identity rather than semantic similarity.," *NeuroImage*, 2013. DOI: 10.1016/J.NEUROIMAGE.2013.05.128.

[28] A. G. Liuzzi, R. Bruffaerts, P. Dupont, *et al.*, "Left perirhinal cortex codes for similarity in meaning between written words: Comparison with auditory word input," *Neuropsychologia*, 2015. DOI: 10.1016/J.NEUROPSYCHOLOGIA.2015.03.016.

[29] Nintendo EPD, *Big brain academy: Brain vs brain*, Nintendo, 2021.

[30] Nintendo EPD, *Dr kawashima's brain training for nintendo switch*, Nintendo, 2020.

[31] J. Allen, *"big brain academy: Brain vs. brain" review: Satisfyingly cerebral*, Dec. 2021. [Online]. Available: https://www.nme.com/reviews/big-brain-academy-brain-vs-brain-review-3114020.

[32] Namco, *Time crisis*, Namco, 1995.

[33] Walt Disney Imagineering, *Buzz lightyear laser blast*, Disneyland Park Paris, "https://www.disneylandparis.com/en-int/attractions/disneyland-park/buzz-lightyear-laser-blast/", 2006.

[34] Namco and Flying Tiger Development, *Time crisis*, Namco, 1995.

[35] F. T. Development, *Time crisis - project titan for sony playstation*. [Online]. Available: https://www.video-games-museum.com/en/game/Time-Crisis-Project-Titan/31/2/19817.

[36] Game Freak, *Pokémon red/blue*, Nintendo, 1999.

[37] Nintendo RD4, *The legend of zelda*, Nintendo, 1987.

[38] King, *Candy crush saga*, King, 2012.

[39] Gamelab, *Diner dash*, Playfirst, 2005.

[40] Ryu Ga Gotoku Studio, *Yakuza 0*, Sega, 2015.

[41] S. Spencer, *Diner dash*, Sep. 2008. [Online]. Available: https://www.pocketgamer.com/diner-dash-iphone/diner-dash/.

[42] W. Quick, *Yakuza 0: Hostess guide*, Sep. 2021. [Online]. Available: https://www.thegamer.com/yakuza-0-hostess-cabaret-club-minigame-guide/.

[43] K. Hoenig, E.-J. Sim, V. Bochev, B. Herrnberger, and M. Kiefer, "Conceptual flexibility in the human brain: Dynamic recruitment of semantic maps from visual, motor, and motion-related areas," *Journal of Cognitive Neuroscience*, 2008. DOI: 10.1162/JOCN.2008.20123.

[44] Hudson Soft, *Mario party*, Nintendo, 1999.

[45] Midway, *Mortal kombat*, Midway, 1992.

[46] Nintendo EAD, *Wii sports*, Nintendo, 2006.

[47] Nintendo EAD, *Wii sports resort*, Nintendo, 2009.

[48] C. t. W. S. Wiki, *Swordplay*. [Online]. Available: https://wiisports.fandom.com/wiki/Swordplay.

[49] C. t. Z. Wiki, *Windvane meadow*. [Online]. Available: https://zelda.fandom.com/wiki/Windvane_Meadow.

[50] Nintendo EPD, *The legend of zelda: Breath of the wild*, Nintendo, 2017.

[51] Nintendo EPD, *The legend of zelda: Tears of the kingdom*, Nintendo, 2023.

[52] Konami Computer Entertainment Japan, *Metal gear solid*, Konami, 1998.

[53] Bemani, *Dance dance revolution*, Konami, 1998.

[54] PlatinumGames, *Metal gear rising: Revengeance*, Konami, 2013.

[55] id Software, *Doom*, Bethesda Softworks, 2016.

[56] FromSoftware, *Dark souls*, Namco Bandai Games, 2011.

[57] J. T. Bowey, M. Birk, and R. Mandryk, "Manipulating leaderboards to induce player experience," *CHI PLAY*, 2015. DOI: 10.1145/2793107.2793138.

[58] D. M. Johnson, P. Wyeth, P. Sweetser, and J. Gardner, "Personality, genre and videogame play experience," *FnG '12*, 2012. DOI: 10.1145/2367616.2367633.

[59] B. Markovitch, P. Markopoulos, and M. V. Birk, "Tunnel runner: A proof-of-principle for the feasibility and benefits of facilitating players' sense of control in cognitive assessment games," *International Conference on Human Factors in Computing Systems*, 2024. DOI: 10.1145/3613904.3642418.

[60] L. Held, J. Pannermayr, A. Kaufmann, *et al.*, "No impact of story context and avatar power on performance in a stop-signal game," *Heliyon*, 2024. DOI: 10.1016/j.heliyon.2024.e41039.

n.b. *The games referenced use the earliest release date for the year of release. Game series in general are referenced by the first game of the series*

Appendix A:

Online materials including meeting plannings, screen shots, gameplay footage, and instructions can be found in the following link:

> https://drive.google.com/file/d/1PwUiKzig$_{R}BuxXqrbLSd2J8OF1$
> $Ma8bxT/view?usp = sharing$

Appendix B: Research Topics, "Mapping Academic Game Development: An Exploratory Investigation"

Mapping Practices of Academic Game Development

Anonymous Author(s)

Abstract

This study explores the development and use of games as research tools through eight interviews with researchers and developers. Interviewees employed academic games for motivation, immersion, or when the gaming context was seen as a necessity. The interview content was analyzed and grouped into nine key themes that illustrate the challenges and opportunities in using games for research. The study highlights collaboration challenges between developers and researchers, as well as communication barriers and disparities in gaming experience. Recommendations include enhancing design process documentation, establishing formal collaboration strategies, and careful consideration of how game elements might affect research validity. These insights contribute to a deeper understanding of academic game development and support more effective collaborations. Finally, the findings of this study underscore the potential value of a new 'academic game developer' specialization; a role that bridges the gap between research and development.

CCS Concepts

• **Software and its engineering** → **Interactive games**; *Collaboration in software development*; • **Human-centered computing** → Interactive systems and tools.

Keywords

Academic games, Research games, Games for research, Game development, Researcher-developer collaboration, Research tools, Ludo-utilitarian dissonance

ACM Reference Format:

Anonymous Author(s). 2018. Mapping Practices of Academic Game Development. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation emai (Conference acronym 'XX)*. ACM, New York, NY, USA, 10 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

Since their inception, video games have evolved into a widely popular medium, sparking growing interest in their potential uses beyond entertainment. Researchers, educators, and professionals have recognized the value of video games for various purposes, exploring their applications in fields beyond entertainment. This has paved the way for the development of serious games as effective tools for various purposes [17]. Apart from serious games for teaching specific skills, games as a research tool to study basic cognitive and motor functioning have been around since at least the 90s [5].

Games can be used in research by having participants perform experimental tasks in a game-like setting, which helps sustain the suspension of disbelief often required in such studies [8, 9]. For instance, cognitive psychology researchers have developed video games as alternatives to conventional tests like the Stop Signal Task, where participants must inhibit a response upon receiving a stop signal [20]. Framing tasks within a game setting can enhance engagement with minimal game elements [6, 12, 15, 23]. More complex games that incorporate narratives and immersive environments can enhance ecological validity, enable detailed event logging, and facilitate scaling through modular game designs [28]

However, video games made for research, and applied games at large, can present a tension between game mechanics that are designed for ludic engagement and those that exist to fulfill their applied purpose [13].

The development of academic games is often poorly documented. Most papers describing the use of games as tests or tasks include descriptions of the game design to varying degrees, but they rarely provide an in-depth elaboration of the design choices [6, 10, 12, 15]. More crucially, this means academics lack insight into the *rationale* behind these decisions and the overall development process.

This paper aims to address the knowledge gap surrounding academic game development by interviewing researchers and developers with relevant experience. Our goal is to uncover the various approaches to game development in this context, identify common challenges and problems, and explore potential solutions.

To investigate the current challenges, opportunities, and motivations for using games in task and test design within academic research, we conducted an interview study with eight experts. These interviewees have backgrounds in either developing or using games as experimental tests or tasks in academic research.

2 Related Work

This section reviews key literature and concepts underpinning our study. We explore 'ludo-utilitarian dissonance', examining tensions between games' entertainment and research purposes. We then discuss video game ontology, considering how game definitions affect academic game development. Finally, we summarize current knowledge on academic game development, identifying research gaps our study addresses.

2.1 Ludo-Utilitarian Dissonance

For this study we refer to 'ludo-utilitarian dissonance' as the tension that arises when attempting to balance the entertaining aspects of a game (ludic elements) with its intended practical purpose (utilitarian properties). In the context of games designed for research or educational purposes, the dissonance manifests as a challenge to create engaging experiences that simultaneously fulfill specific experimental objectives.

Conference acronym 'XX, June 03–05, 2018, Woodstock, NY

ACM ISBN 978-1-4503-XXXX-X/18/06
https://doi.org/XXXXXXX.XXXXXXX

Friehs and colleagues adapted a cognitive psychology Stop Signal Task into a digital game, allowing players to use gaming controllers [6, 18]. The original task presented white arrows on a black background, while the adapted version incorporated game elements. In the traditional version, participants press a key when an arrow appears on screen, but must refrain from pressing when they hear a beep. This task, used since the 1980s, is considered reliable [20]. The gamified version maintains these mechanics but replaces the arrow with a cartoon fairy and sets the task in a virtual forest, where the player's character must choose the correct path. This narrative framing transforms the test into a game-like experience.

From a formal game design perspective, these research games often lack key elements like meaningful choices, clear goals, and win/loss conditions [19, 22]. Participants may even point out these shortcomings [6, 7]. However, incorporating traditional game elements could potentially compromise the validity of the research. For instance, even adding a simple point system to a dot probe task has been shown to affect participant performance [27].

Any additional element to an established test might influence participants' performance and experience by reducing the task's faithfulness to the original, thereby lowering control of external variables and potentially decreasing the measurement accuracy of the cognitive process. Consequently, implementing more conventional forms of game design within these tasks falls victim to the clash between the purpose of games and the goal of research. A task may be *framed* as a game but avoid deeper game design elements in favor of heightened experimental control [6]. This new 'task game' can then be validated, to some degree, against the base task. Interestingly, when participants engage with such task games, their engagement and attention may increase. Previous studies even suggests that game elements could be beneficial for performance consistency [6, 26].

2.2 On the Ontology of Video Games

Defining games has challenged scholars for decades. Huizinga and Caillois laid foundational concepts, emphasizing play's separation from reality and non-productive nature [3, 9]. Modern definitions vary, focusing on aspects like meaningful choices, resource management, or learning systems [4, 14, 19]. Rather than relying on subjective definitions, a more effective approach is to identify specific markers of video game mechanics, aesthetics, and terminology. Initiatives like *OntoJogo* aim to standardize game classifications based on observable attributes [24]. This shift prioritizes players' experiences over designers' intentions, acknowledging that even activities like slot machines, which may not fit strict game definitions, are often perceived as game-like by users.

In the academic context, games are primarily designed for data collection rather than entertainment, mirroring how casinos create engaging environments to encourage spending. This approach allows researchers to leverage game-like elements for research purposes while acknowledging the broader spectrum of what constitutes a game in various contexts.

This exploration of game definitions underscores a crucial point for our study: the concept of what constitutes a game can differ significantly among various stakeholders in academic game development. An experiment crafted as a game might not be perceived as such by participants, while conversely, experimental tasks might inadvertently be experienced as games, even without the researchers intending this framing.

We argue that the effectiveness of academic games depends primarily on participants' engagement and experience, rather than researchers' or developers' intentions. This is similar to how a casino's success is determined by player engagement, not how operators classify their offerings. When studying academic game development, it is crucial to explore how different stakeholders perceive the 'gameness' of a system, as this perception may vary.

For this study, we broadly define games as digital creations perceived as games due to their visual style, presentation, or societal associations. We specifically use 'task games' to refer to games designed primarily to gather insights from participants.

2.3 Insights Into Academic Game Development

Researchers have attempted to conceptualize and guide the development process of serious games [2, 17, 28]. However, studies using games as research tools typically provide minimal insight into the design process, focusing instead on the final design and research methodology [6, 10, 12, 15, 18]. This gap in understanding the rationale behind specific design choices may lead to inadvertent omission of crucial elements in subsequent research.

Developing games for research requires a broader skill set than typically possessed by specialized researchers. Multidisciplinary collaboration is essential to bridge technical and design implementation gaps [2, 11]. However, such collaboration is seldom documented in academic game papers, obscuring valuable lessons and potential pitfalls.

Further investigation into the design process of academic games could yield important insights, particularly regarding interdisciplinary collaboration challenges and their impact on design decisions. This research could provide crucial knowledge about the complexities of game development for academic purposes, enabling more informed future project choices.

3 Method

This section outlines the methodology employed in our study, which involved conducting 8 interviews with stakeholders in research projects that incorporated game development. As a component of our interview process, we developed the Designer Implementer Researcher Developer Involvement Canvas (DIRDI), a conceptual tool designed to help participants visualize and articulate their roles within the development process of games for research purposes.

In the following subsections, we provide a detailed description of the interview design, procedure, and subsequent analysis.

3.1 DIRDI Canvas

Our interview methodology includes a canvas for participants to categorize their roles along two key dimensions. The first axis distinguishes between game development and integrating games for research, while the second axis contrasts technical implementation with design influence. This framework, termed the DIRDI canvas (Designer, Implementer, Researcher, Developer Involvement), also includes a third axis to represent the participant's overall involvement in the project relative to other stakeholders.

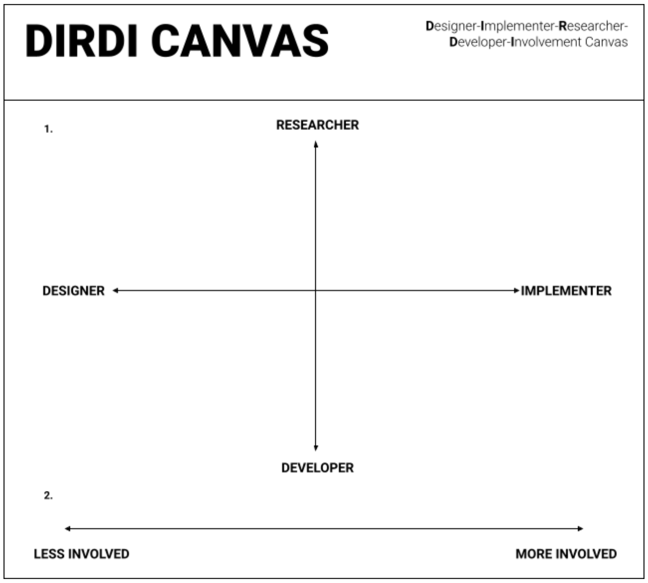


Figure 1: The Designer Implementer Researcher Developer Involvement Canvas used to indicate a participant's role within academic development. The canvas is meant to be augmented by participants by drawing on it.

3.2 Interview Design

The interview is designed to gather qualitative data on the usage of video games as a form of experimental test or task within any research field. The following questions are what we seek to address through the individual interview sessions:

(1) How is the development of games as tests or tasks within academia structured?
(2) What tensions exist between developers and researchers?
(3) What are reasons for using games as tests/tasks in academia?
(4) What causes games as tests/tasks within academia to fail?
(5) What resources do academic game development lack?

These questions assume that academic game development typically requires an interdisciplinary team for effective collaboration, particularly because many researchers may lack game development expertise.

The interview was structured into five sections: Introduction, Participant Background, Conception, Design, and Teams and Resources.

Introduction. This section serves as a guide to the starting procedure of the interview including the explanation of the scope and handling of informed consent.

Participant Background. Participants introduce themselves, share their background, and specify their role in developing games or tools for their academic projects. We collect this information to contextualize participant responses. As part of this, participants map their position on the DIRDI canvas.

Conception. This section explores the early project stages of using games for research experiments. Participants discuss the reasons for developing their games or game-like tasks, how their most recent project began, whether this is a common approach across their projects, and what considerations are made when starting a new project. Additionally, participants are asked whether their game-tasks are based on existing methods or projects, or if they represent novel experiments. Finally, we ask participants to reflect on any failures or instances where the final result was not utilized when employing games and elaborate on the reasons.

Design. This section explores the interplay between research and game development, focusing on design considerations throughout the development process. The questions probe the collaboration's impact on experimental validity while indirectly revealing insights into design decisions. Participants are asked about the level of creative freedom given to developers and the specificity of researchers' instructions. The section also examines researchers' awareness of developers' design choices and reasoning, as well as any potential 'hidden' decisions made by either party that might affect the other. Lastly, participants consider whether these undisclosed choices could influence the experimental outcomes.

Teams and Resources. The final part aims to contextualize participants' projects in terms of resources and scope. Participants are asked about available resources, including team size, project timeframes, and any material or resources provided through academic institutions. The section also explores whether resource limitations affect game design, execution, and what additional resources could benefit future works.

3.3 Participants

We invite participants through convenience sampling through our professional network. In the end, 8 participants completed the interview in full. Participants are invited based on whether their previous work contained either:

(1) A digital game for an experimental test or task
(2) A digital tool with game elements for an experimental test or task
(3) A virtual reality tool for an experimental test or task

We include Virtual Reality (VR) in our inclusion criteria due to its close association with games in development processes and consumer markets [21]. Additionally, the framing effect observed in games may extend to VR, potentially impacting engagement through VR interactions, similar to how framing an activity as a game can increase engagement [16].

3.4 Procedure

This section outlines the procedure for conducting the interviews. We recruited participants via email, inviting them to an online MS Teams interview about the study. Eight out of nine initial participants completed the interviews, all holding at least a master's degree in relevant fields. Semi-structured interviews were conducted, recording audio and optional video, with durations ranging from 30 to 87 minutes. Participants drew on canvases to visualize certain answers.

Interviews were designed with broad, open-ended questions addressing our research questions, allowing participants to share their experiences freely. Audio was automatically transcribed and manually anonymized, with personal names replaced by [REDACTED] and unintelligible speech tagged accordingly. We obtained informed consent for all data collected, including drawings, recordings, and transcriptions, with no compensation offered. The study was approved by the he ethics board at the [redacted].

3.5 Analysis

To conceptualize games as tests or tasks within academic development, we conducted an abductive thematic analysis [25]. We used the qualitative data analysis program *ATLAS.ti* for coding. The transcriptions underwent two rounds of coding. In the first round, we aimed to capture every noteworthy instance within the research context to gain a wide range of insights. We then grouped and combined these codes to create a new codebook, including only codes that appeared in multiple interviews. Using this new codebook, we performed a second round of coding. These final codes were used to generate the themes. Given the field's unconsolidated nature and our relatively small sample size, we separately evaluated codes that appeared in only one interview but were deemed significant to the overall knowledge. We considered these single-instance codes in our later discussion.

4 Results

For the results, we assigned each participant a letter code corresponding to their canvas results. Some discretion was used in marking participants' indicated positions, as the canvases required manual alignment.

Our DIRDI canvas mapping revealed that most participants viewed themselves as leaning slightly more towards "Designers" and "Researchers" than "Implementers" and "Developers". We observed a wide range of involvement levels in the development of games as tests or tasks. Of the eight participants, three identified more as designers (A, D, G), while five saw themselves as both implementers and designers (B, C, E, F, H). Four participants leaned more towards researchers (A, B, C, D), two towards developers (G, H), and two were balanced between researchers and developers (E, F). Five participants reported working collaboratively with other researchers or developers, excluding instances of student supervision (A, B, C, D, F). Six participants considered at least one of their research tests or tasks to be a game (A, C, E, F, G, H), while three did not (B, D, G).

Participants came from diverse academic backgrounds: two in computer science (D, H), three in human-computer interaction (C, F, G), one in psychology (B), one in game design (A), and one in industrial design (E). Of these, two participants held PhD degrees: one in psychology (B) and one in human-computer interaction (C).

5 Key Themes

This section presents the key themes that emerged from our analysis of the interviews. These themes cover key aspects of using games for research, including: collaborative challenges, engaging participants, adapting experimental tasks into task games, VR applications, design considerations, researchers' gaming experience, understanding design choices, resource limitations, and quality concerns in research games. Each theme is illustrated through brief excerpts from the interviews.

5.1 Challenging Collaboration

Among a number of participants, the collaboration between developers and the researchers ('stakeholders') was discussed. A majority of the developers who worked on collaborative projects with a stakeholder saw this collaboration as challenging. Their concerns could be categorized in roughly three categories:

(1) Problematic communication
(2) Imbalance of the stakes
(3) Insufficient knowledge of games and conceptualization capabilities

Problematic communication stemmed from stakeholders not being able to articulate their needs or expectations for the project, as well as differing levels of involvement. Both researchers and developers reported instances in which the end result was differed from what was intended.

> [A]: *There's ... the nightmare that was for years to get psychologists to properly spec what they ask for.*
> [B]: *I didn't make [it] very explicit that I need [the] commanding officer to disappear after that interaction... those are things that sometimes if you forget they have quite big implications... So that's one of the things that there was lately like "ohh yeah, I didn't say that explicitly"*

These instances were attributed to oversight and not knowing beforehand that the lack of specification would influence the end result. On the lack of involvement from stakeholders, participants commented that not all researchers would be as involved in the design of the final result, which resulted in developers having to make decisions on their own that would be different from what the researcher expected.

> [B]: *I must say that that not every researcher does that. Some really leave more of the designing to the implementer.*
> [G]: *[B]ecause the professor has not gone into the drawing. [If] it's not really similar to the interactions that he had physically, it just doesn't work*

Participants also discussed an imbalance of stakes, noting that the researcher who commissioned the game was often considered the final stakeholder due to the project's goal. Developers were limited in their executive capabilities, acting mostly as advisory experts and implementers who provide feasible options for the researcher to choose from.

> [B]: *I find it important that I keep the last say over the the design.*
> [C]: *I have a key role in that research. It [needs to] resonate with me, and it definitely needs to be agreed upon how the game looks like and what are the features and how it's being used.*
> [G]: *I can't take a design perspective and say well the easiest for me it's that we develop [this] and the best interaction that I have is to have the controller and the easiest one [is to] just grab the object. But that doesn't mean that for my stakeholder that actually would be the best one ...*

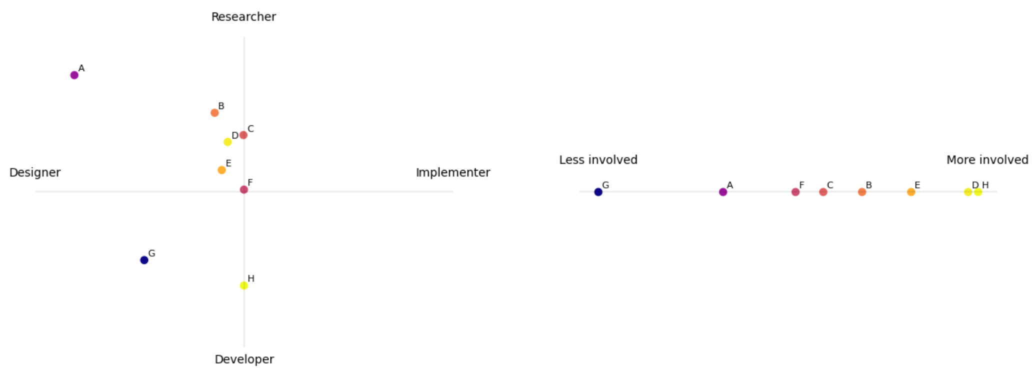


Figure 2: Participant positionality and involvement within their research games as indicated through the DIRDI canvas.

Participants with more development-focused roles noted researchers' limited knowledge of games and difficulties in conceptualization. They observed that researchers often lacked a comprehensive understanding of games due to their different academic backgrounds.

This knowledge gap was evident in the specification of the final game, where researchers sometimes struggled to articulate the precise requirements needed for their experiments.

> [A]: *They don't have any training on it. They have a lot of trouble conceptualizing and describing what they actually want or what they actually think they need for an experiment.*
> [A]: *[P]sychologists will very, very easily ask you for a fully realized, completely seamless virtual open world with a hyper detailed physics engine, without bothering to consider: in what way this is surplus to requirements for what they actually need for a given experiment.*
> [G]: *And the problem becomes when the professor himself cannot come up with the concept of a game.*

Participants elaborated on challenges in the development process, highlighting the importance of meetings and iterative development. Meetings were emphasized as a crucial tool to address communication issues and ensure researcher involvement, often by mandating attendance. Meanwhile, an iterative approach was mentioned primarily by participants who had a development background.

> [A]: *And it advises the proponent that there will be meetings within this period and you have to attend. Otherwise, this doesn't go forward.*
> [D]: *You have meetings, you discuss them and then slightly bring forward the problems and then you share what you all thought or how to solve these problems... because otherwise you implement something and then it's like, yeah, actually that's wrong because we needed it that way.*
> [F]: *... for the most part, it's like I'm just try the thing [game prototype] and see how it works... and then I'll get feedback from my supervisor.*
> [D]: *...people would become aware in the testing and see, "oh, this is wrong."*

In short, problematic communication and lack of stakeholder involvement were considered to lead to projects where the the end result was incongruent with the goals of the researcher. Some stakeholders struggled with communicating their needs, discussed differing levels of involvement throughout the development process, and indicated insufficient knowledge of game development. These problems led to misaligned outcomes, and complications in the development process. Strategies such as meetings and iterative development were cited as a way to address these issues.

5.2 Motivating and Engaging Experimental Participants

Participants discussed the usage of a game-like task to attract attention, increase participant engagement, or maintain attention throughout the duration of an experiment.

> [H]: *When you frame it as a game and people wanna do well, I guess they sometimes get competitive... They wanna do well and it's easier to get people to sign up for the the task.*
> [D]: *People get bored and therefore their concentration focus diminishes. So how can you keep them entertained in a sense that they stay focused on the actual task and not lose interest, therefore... affect the results.*
> [F]: *[I]f they're getting better, it's probably satisfying in some way, right? They're feeling competent, so maybe... that might be motivating to them to keep playing.*

5.3 Using Conventional Games as Tasks

Beyond motivating and engaging participants, several interviewees indicated that their experiments utilized games to examine phenomena inherently present in gaming environments.

> [C]: *[P]eople had to... finish [an] escape room game where they had to interact with the NPC using the speech. And for this game, we wanted to come up with this new system of speech recognition where you would use contextual information and so on.*
> [H]: *In games, you often have to move the mouse quickly to click on targets. For PC games, pointing tasks? Yep, are absolutely used in games.*

Additionally, one participant gave an example where the experimental test and subject came from a different source, but that they applied a gaming test to understand the previous study in the context of games.

[F]: *I was reading a textbook on human performance and it was giving all these different things you could do to modify practice... to improve performance... I'm like, well, damn, we care about performance a lot in games. Let's see if this works for games.*

Among participants who worked on projects involving these inherent game tasks, some also mentioned using existing games as sources of inspiration or using some of their mechanics inside of their task games.

[F]: *It was based off of a game called Speedrunners. [...] I was modeling the movement off Super Meat Boy [...] The first time I did it, I was having an undergrad create a clone of Super Hexagon.*
[H]: *Also later do like a... a pong game where you move the mouse up and down to move the pong paddle and then it would block the ball.*
[C]: *I think for 99% of the case that it was my own design that we came up with. And of course, it's inspired by many existing video games.*

Two interviewees reported that their games were guided or inspired by previous games made for research purposes.

[E]: *Similar like these type of augmented sandboxes, you may have seen those.*
[G]: *We had one professor who brought us [a] logistics game and he had already worked on quite some time on that project and he had already been working with a developer to understand how to transform things.*

5.4 VR: Practical but Not a Game

Participants that mentioned VR mainly described its primary functionality as being an alternative to real life experiments that would be less practical or ethical. They noted that VR allows researchers to put people in a wide range of scenarios and buildings as well as create an environment which would be otherwise difficult to recreate.

[G]: *My first like 5 different VR experiences was people telling me like "I wanna put my person in a church and then into prison".*
[B]: *We wanted to have a a spaceship in which... as a participant, you were in a sleeping quarter. You leave the sleeping quarters and a commanding officer is verbally harassing you.*
[B]: *I work with criminals and potential victims, and we know, of course, that doing a a crime in public is for me the best way to see what's going on. But it's ethically not on the table, so we use virtual environments [...] and then we see how they respond to manipulations and the environment.*
[G]: *We had this project with a student who wanted to do a project of simulation for... first respondents... So we have this guy standing at the edge of a building... And so your task is to approach the person.*

Notably, participants who utilized VR generally expressed concerns about participants not behaving realistically or acting differently compared to real-life situations. One participant specifically linked this issue to the association of games with VR.

[B]: *I try to really elicit the natural behavior that people have. So that's why it's important for me to look into presence, how people feel immersed in virtual environments because I don't want them to treat the virtual environment as a game, because then they would show behavior they might not normally show.*
[G]: *His analogy was: "but, you know, students are used to these video games where they kill everybody". And I was like, yeah, but because... you're trying to get the different type of... comprehension of the situation, that is a quite harsh measurement.*

5.5 Differences Between Tasks and Game Design

Interviewees discussed differences between the design of games and experimental tasks. One factor was that people might change their behavior because they engage in a game-like activity. One participant mentioned that this is also something that they believed to be intrinsic to games.

[A]: *It varies according to how the player is feeling, or basically a game is an environment that makes it possible for people to have conducts. Not just actions but conducts so the player can legitimately say OK here under these circumstances, given how this is going, I'm going to be this person or I'm going to be this other person. [...] But I'm gonna have a big impact on the game state or I'm going to play cautiously. A game, makes these conducts possible.*

Another interviewee also noted that something like playfulness could negatively affect their experiments as it would derail the experimental process

[C]: *When you need something in the game and people just become playful and throw it out of the window or something, and then we have to restart sometimes the game.*

Participants noted that experimental tasks are often designed to be very precise and that deviation from these tasks risks affecting the results.

[D]: *The problem what you have with these kinds of validated tasks, if you change too much of them, then it's not the task anymore. And then for therefore you're risking, oh, this is now something else.*
[F]: *Let's make this an experiment. That is as close to I can get to... what you could actually find in a game environment, except there's no actual like mechanics that are distracting.*
[B]: *There was a another intern for the programmer who said like yeah, but you can also put in... [that] you can also open doors and have trash et cetera and then the lead programmer said like: Yeah, but we're only focused on the person, so we need to keep everything constant.*

Interviewees noted that games used as tests or tasks are typically brief. One of them highlighted this as advantageous for project scope, reducing the need for extensive feature implementation. They further advised keeping games simple, explaining that participants require time to learn game rules. This learning period consumes part of the total experiment time, consequently limiting the duration available for actual gameplay.

[F]: *The other thing is like a research game. It's going to be played for probably about an hour at most. Most of the time, unless you're you're having people come back from multiple sessions.*

[D]: *[A gaming task] It's more short term and optimized towards "OK I want to have these variables."*
[F]: *You can get away with like only building a handful of levels or not implementing some features or things like that.*

Finally, it was implied that the process of developing a test or task game was different from that of conventional games. The setting of requirements and goals was mentioned by interviewees as a step before production or design of the game even begins. This step is then followed by the creation of the design of the game.

[G]: *Research will put you specific guidelines or... the specific requirements that you need to achieve in order to conduct the experiment.*
[G]: *So most professors will say... the model is A + B = C alright, so how do we turn that into a game?*
[E]: *But you can see it like if you know where to go, you know the endpoints, then you can then you know, then you're then, then all your decisions become guided.*

5.6 Differing Levels of Experience With Games

Interviewees reported that participants' varying levels of gaming experience impacted their interactions with game tasks. This diversity in experience manifested in several ways. Interviewees observed that players with different experience levels required varying amounts of time to learn the game mechanics and rules, and that gaming experience also influenced players' motivation to engage with and learn the game.

One interviewee highlighted that higher game difficulty could limit the sample size, as less experienced players might struggle to progress through the game. Additionally, gaming experience affected what players found interesting or challenging, with less experienced players being more easily engaged by simpler game mechanics.

[C]: *There are people that need more time to learn the game.*
[F]: *From an accessibility perspective... There's concerns there because... your participants are not necessarily as heavily invested in learning the game as... the typical player would be. [...] We had a bunch of people that couldn't make it through our tutorial levels... and then we had to basically filter out those people.*
[F]: *A really popular one is the infinite runner [...] and to me that game is really boring, but to someone who's never played a game, which some of our participants could have been, it's challenging and it's interesting.*

One interviewee noted that researchers sometimes mistakenly assume that people have sufficient gaming literacy.

[G]: *I press play and I was expecting that the participant knew that he had to run and walk. [...] They don't understand why, but people didn't look around in virtual reality. I said: "Did you tell them to look around?" "No!" I said, "well most of the people are not used to going into virtual reality and they have no clue."*

5.7 Lack of Understanding of Design Decisions

Interviewees mentioned occasions where they were unsure about the effects of design decisions within their games. They noted a lack of knowledge on how the results of their studies would be affected by design decisions, such as those implemented for 'fun' or to increase engagement.

[C]: *Small design decisions have an impact. Which you don't know exactly. You can anticipate to some extent, but you don't know exactly what is the reaction, right?*
[F]: *I feel like... finding and making a game fun is very confusing and nebulous.*
A: *[T]he problem is operationalizing it for game design and what I was trying to do with my experiments was operationalize these ideas of what games are... How does this translate to something... when designing a game resource?*
[F]: *It's kind of like we're just modeling our... things off of what other people have done and we don't necessarily know why they've made those decisions, just that they have made those decisions.*

A lack of design knowledge was noted with respect to specifically games as tests, which included how to scaffold complicated tasks and how different mediums for tests such as VR would influence the results of engagement.

[F]: *If you want them to do a task that's complicated, and if you need to be very careful with how your scaffolding them. And that's like almost a separate research.*
[G]: *We actually had to do a study to demonstrate that there was actually a higher engagement because... virtual reality actually changed just a little bit the way the person risked himself out of the fact that he felt way more engaged within virtual reality than in the 2D version.*

5.8 Lack of Time and Work Power

When discussing project resources, all but one of the interviewees indicated either a lack of time or a lack of developers. Lack of time resulted in a limitation of the final result through limited development time, which would sometimes be caused by limited academic time.

[C]: *You always have limited time and you have deadlines and you need to get that research done in this time frame that you plan for. And that means you have limited time for development.*
[B]: *[Discussing a project] I would love if the avatar [...] has even more facial features. More realistic, but that's not gonna happen because it takes too much time.*
[E]: *The final part of that project was focused on evaluating the game, getting data and analyzing data, writing a paper on [it], finishing the thesis. While, if that was not connected to this PhD project [you] could use the the final months to further develop the game in theory.*

Interviewees noted a lack of work power which would also worsen time limitations by slowing down the speed of development.

[C]: *Usually you don't have a big team for development [...] One person, maybe two people, maybe in best case scenario, three people.*
[F]: *I think that does limit me in some ways because I'm putting myself in a position where I have to be the developer of the game... and I'm a single person. I can't make a AAA game or anything that looks like it.*
[A]: *There's always a tremendous lack of developers.*

[H]: *Well, I, I mean, I guess having someone else do the, I don't know, 40 or 50 hours of coding would save me time.*

5.9 Bad and Buggy Games

Interviewees alluded to working with qualitatively lacking or otherwise shallow games. These were noted both the context of experimental games being made poorly, and games for research not being 'true' games. One of the interviewed participants noted the lack of expertise by a researcher when commissioning a game as a source of these issues.

[C]: *I've been [very] impressed with how well some of these games were designed and how amazing and fun the games [...] could be. And once you see that and [...] go over some really boring, horribly done games, it's hideous. And [it is] hideous for the participants. You're not respecting their time as well by designing something like this and and calling it like a research prototype.*
[D]: *[I]'s [a] chore and so like they are standing against one of the magic circle descriptions of games, which is like the freedom. Like the freedom to stop, which I don't have in a task.*
[A]: *If they make a game like that, what they're doing is they're treating the player like a circus animal. So circus animal does tricks. Receives treats. That's not my definition of a game, and I keep butting heads with my colleagues and psychology, who believe that something is a game as long as it's made in unity.*
[G]: *Unless [the researcher] really has the ability to understand that he requires a designer and a creative team to turn his concepts or ideas into something else, it becomes... something really clunky because they don't have the creativity to go out of the subject.*

Outside of the design of the game, interview participants also noted that execution wise, research games were prone to bugs which could affect participants.

[C]: *I think there are always cases that uh, you know functionality wise something is not working as people expect it to work and you could have maybe spent a couple of more weeks on the development to basically fix that and maybe the participants would have lower frustration rates or something like that or would be able to enjoy the game more.*
[B]: *I think it works to show as a proof of concept, but we can definitely improve for next time to make sure that there's no glitches or the least amount of glitches in the virtual environment.*

6 Discussion

This section interprets findings related to the design and collaboration challenges in academic game development, proposing practical strategies to address these issues.

Our interviews show two key aspects that are relevant for academic game development: (1) the purpose and design of academic games, and (2) the collaborative processes involved in their creation. The following discussion delves into these areas.

6.1 Purpose and Design of Academic Games

Researchers hold diverse views on the definition of games and struggle to operationalize these concepts. Nevertheless, they agree that there is value in digital tools that resemble games, whether through similar development processes, game-like visuals, or familiar mechanics, in academic contexts. This perspective supports moving away from the idea that purposeful games must share the same development goals as entertainment games, such as offering meaningful choices.

Our findings suggest that researchers assess academic games based on their added value to research, rather than categorizing them by purpose as in previous work [8]. Approaches like OntoJogo [24], which describe specific game elements, are less effective for conceptualizing games due to the required game design knowledge. We propose defining research task games by their specific contribution to the research, focusing on their added value rather than predetermined categories.

Our thematic analysis yields a framework categorizing academic task or test games into three types: engage, immerse, and necessitate. These categories reflect shifting design priorities based on game type. Games that engage aim to sustain participant engagement in otherwise monotonous tasks, such as the Stop Signal Game. Games that immerse, like VR-based museum environments, study realistic behaviors. Studies that necessitate games, exemplified by a modified replication of an existing racing platformer game, aim to examine specific behaviors within a gaming context while allowing for greater experimental control. Each category's distinct design needs are elaborated upon below.

Engage. Games that aim to engage primarily focus on maximizing participant engagement and must balance the utilization of game elements with the control and faithfulness to the test or task they adapt. These opposing values include aspects such as higher control, versus supplementary game mechanics, forcing a player to follow a specific experience versus giving them choices and freedom. As such, when developing games for these tasks, many researchers may employ the use of a narrative element and the framing of the task as a game through a visual language that mimics the visual language of games. Prior studies have shown valid results with increased engagement using these methods [6, 12, 15, 23].

Alternative strategies could further enhance this effect. Scoring systems have shown potential in increasing motivation [23], while fostering paratelic motivation by making tasks more enjoyable is another viable approach [1]. To avoid unintended consequences, developers should focus on enhancing task satisfaction rather than introducing mechanics that significantly alter the task itself [27]. Careful implementation and testing are crucial for each specific task, and evaluating different designs may require a large sample size to account for participants' increasing proficiency over time.

Immerse. While engaging games aim to capture attention, immersive games focus on creating realistic experiences that replicate real-world behaviors. Our interviewees particularly emphasized the use of virtual reality (VR) for this purpose. In these cases, researchers strive to minimize behavior that deviates from realistic responses. The association between VR and games in participants' minds presents a challenge, requiring researchers to carefully guide them and limit unwanted behavior. Participants' behavior validity can be compromised when individuals are too playful in the virtual environment. Consequently, interviewees avoided implementing additional game elements in these immersive experiences.

In our interviews, games that focused on immersion did so through literal replication of environments in VR. Examples included having people pretend they are in a museum, burgling a house, or dissuading someone from jumping off a building. Our findings also suggest that the design of the experimental process could be improved. Participants are observed to have varying levels of experience with games, and those unfamiliar with them require more detailed instructions. Furthermore, an important aspect discussed in an interview is the design of the entire experimental experience in VR research. The researcher noted that administering surveys between experimental conditions in real life often disrupted participants' immersion. This suggests that immersion extends beyond the game itself, necessitating careful design of the entire experimental process (from instructions to evaluations) to ensure seamless integration into the virtual environment.

Necessitate. Some research tasks necessitate the use of video games to study behavior within a specific gaming context. However, directly replicating commercial games presents challenges, such as the inability to log precise in-game events or modify game mechanics to investigate the effects of specific variations. Games inherently requiring a gaming context for testing can be considered the most authentic. Our findings suggest these games have fewer design restrictions since replicating the experience of a true game is part of the experimental environment. As a 'vertical slice' of a game, it's logical for researchers to draw inspiration from existing games, either to save time or to target a specific audience. One interview participant noted that as long as the experience remains consistent across all aspects except the manipulated variables, the results should remain valid, any effects from the overall experience ought to cancel each other out.

Beyond the specific observations for the three categories mentioned above, some general findings apply across all types: Test games for single experiments are typically short, enabling developers to concentrate on refining details. However, development often faces time constraints and a shortage of skilled developers due to budget limitations, insufficient workforce, or a lack of researchers with strong development skills. These constraints can affect the design's comprehensiveness and the code's quality.

Varying levels of gaming experience among participants can pose challenges, such as the need to account for skill disparities in the results. Borrowing designs from existing games or closely emulating familiar game elements can reduce learning time for participants and can save time in the design phase. Designing research games for a specific user group may further streamline the design process but may also limit the participant pool.

Inevitably, some players will have greater skill in certain games than others. To ensure validity, researchers can control for this variability by conducting surveys at the game's outset to assess familiarity and skill levels. Alternatively, implementing difficulty calibration trials before collecting study data can effectively level the playing field.

6.2 Collaboration and Logistics of Development

Collaborating with researchers who lack video game or design literacy presents a unique challenge and requires translating their goals effectively. This process demands effort from both parties to communicate efficiently and arrive at a game design that meets the researchers' needs. Although researchers may not excel in understanding game design, they are crucial in identifying validity issues during a game's development.

Developers should remain cautious of input from researchers that might inadvertently compromise research objectives unintentionally. However, researchers may also have very specific and precise needs in relation to the task that cannot be changed. Developers needs to be receptive of these boundary conditions.

Developers working on academic games would benefit from acquiring some research background and conducting preliminary research in their collaborators' field. The collaborator may also benefit from preparing a breakdown of the most-important concepts and methods that are relevant for the design of the game and the understanding of the larger context where it will be embedded. This approach enhances their understanding of the project's goals and stakes. To ensure the final product suits the research needs, it's crucial to establish clear research objectives and requirements before initiating the game design process.

Collaborators should agree on regular communication to avoid a lack of input from researchers during the development. 'Calibration' meetings should serve as moments to discuss design options, explain and declare relevant design decisions that have been made by the developers, and get feedback on the current design or implementation of the game. Highlighting the iterative nature of development can aid in managing expectations. Initially, rough prototypes should suffice where only core game mechanics are demonstrated without using sophisticated assets or graphics.

Finally, in studies employing games as task-performance tools, researchers should provide comprehensive documentation of the design process as supplementary materials. This enhanced transparency enables future researchers to better understand a game's design, allowing them to assess its capacity to yield valid results. Moreover, it would help anticipate the potential consequences of modifying specific design elements when adapting work.

Limitations and Future Work. This study has limitations, including potential selection bias from our convenience sample and limited generalizability due to small sample size. Future research could investigate the three game types through structured interviews or document a complete development process as a case study. Exploring participant experience levels, developer-researcher differences, and various task game categories could provide deeper insights into their impact on academic research.

7 Conclusion

In this study, we conducted interviews with researchers and developers about the creation and use of academic games. Our analysis revealed nine key themes that illuminate current practices in academic game development:

(1) Collaboration challenges: Differences in expectations, game literacy, and focus areas between stakeholders often complicate partnerships. Calibration meetings and iterative development can foster a more unified collaboration.

(2) Participant engagement: Using games for research is seen as advantageous for attracting participants and increasing engagement.
(3) Adaptation of conventional games: Researchers use existing games to study gaming-specific phenomena or apply gaming contexts to other fields. They often draw inspiration from popular games, adapting them either literally or conceptually for research purposes.
(4) Virtual Reality (VR) considerations: VR's overlap with games can be problematic when game connotations are unsuitable for research.
(5) Research-focused design: Academic games prioritize specific experimental goals over entertainment. Development concentrates on essential features supporting research objectives while avoiding potential distractions.
(6) Participant gaming experience: Varying levels of gaming experience and literacy among participants can unintentionally impact task performance and experimental results.
(7) Unintended design effects: Design decisions in academic games can unpredictably influence study results. Reducing game elements may improve experimental control but can also diminish intended properties.
(8) Resource constraints: Time and personnel shortages often limit the final product in academic game development. Developers mitigate these constraints by focusing on specific game elements crucial for research outcomes.
(9) Quality concerns: Poorly designed and buggy games frustrate participants and compromise research results, underscoring the importance of allocating sufficient time for development and testing.

Our research led to the classification of academic test games into three categories based on their research value. We offered guidelines for developing these games and enhancing collaboration in academic game projects. The development of academic test games faces challenges in translating research tasks into engaging gameplay and fostering effective teamwork. We advocate for a 'academic game developer' specialization that combines research acumen with game development skills to improve team collaboration and shared project understanding.

To advance this field, we recommend further research to standardize development processes and improve communication strategies. Thorough documentation and publication of academic game design methodologies are essential to bridge knowledge gaps for future researchers and developers. Addressing these challenges will not only enhance current research practices but also lay the foundation for a discipline that integrates game development with experimental research.

References

[1] Michael J. Apter. 1984. Reversal theory and personality: A review. *Journal of Research in Personality* (1984). https://doi.org/10.1016/0092-6566(84)90013-8

[2] F. Bellotti, Riccardo Berta, and A. D. Gloria. 2010. Designing Effective Serious Games: Opportunities and Challenges for Research. *Int. J. Emerg. Technol. Learn.* (2010). https://doi.org/10.3991/IJET.V5S3.1500

[3] Roger Caillois. 1977. *Jeux et les hommes.* GALLIMARD.

[4] Greg Costikyan. 2002. I Have No Words & I Must Design: Toward a Critical Vocabulary for Games. In *CGDC Conf.* https://api.semanticscholar.org/CorpusID:39630945

[5] Emanuel Donchin. 1995. Video games as research tools: The Space Fortress game. *Behavior Research Methods Instruments & Computers* (1995). https://doi.org/10.3758/BF03204735

[6] Maximilian A. Friehs, Martin Dechant, Sarah Vedress, Christian Frings, and Regan L. Mandryk. 2020. Effective Gamification of the Stop-Signal Task: Two Controlled Laboratory Experiments. *JMIR serious games* (2020). https://doi.org/10.2196/17810

[7] R. Gallagher, K. Kessler, J. Bramham, M. Dechant, and M. A. Friehs. 2023. A proof-of-concept study exploring the effects of impulsivity on a gamified version of the stop-signal task in children. *Frontiers in Psychology* (2023). https://doi.org/10.3389/FPSYG.2023.1068229

[8] Marcello A. Gómez-Maureira, Max van Duijn, C. Rieffe, and A. Plaat. 2022. Academic Games - Mapping the Use of Video Games in Research Contexts. *International Conference on Foundations of Digital Games* (2022). https://doi.org/10.1145/3555858.3555926

[9] Johan Huizinga. 2014. *Homo ludens: A study of the play-element in culture.* Martino Publishing.

[10] Colby Johanson, Carl Gutwin, and Regan Mandryk. 2023. Trails, rails, and over-reliance: How navigation assistance affects route-finding and spatial learning in virtual environments. *International journal of human-computer studies* (2023). https://doi.org/10.1016/J.IJHCS.2023.103097

[11] Jaana Kemppainen, Tanja Korhonen, and Teija Ravelin. 2014. Developing Health Games requires multidisciplinary expertise. *Finnish Journal of eHealth and eWelfare* 6 (Nov. 2014), 200–205. https://journal.fi/finjehew/article/view/48213

[12] Hannah Kirsten, Martin Dechant, Henning Gibbons, and Maximilian Achim Friehs. 2023. Tasting inhibition: A proof-of-concept study of the food stop-signal game. (2023). https://doi.org/10.1016/BS.PBR.2022.12.002

[13] Isabelle Kniestedt, Iulia Lefter, Stephan Lukosch, and Frances M Brazier. 2022. Reframing engagement for applied games: A conceptual framework. *Entertainment Computing* 41 (2022), 100475.

[14] Raph Koster. 2004. *Theory of fun for game design.* Paraglyph Press.

[15] Andreas Lieberoth. 2015. Shallow Gamification. (2015). https://doi.org/10.1177/1555412014559978

[16] Andreas Lieberoth, M. Møller, and Andreea Catalina Marin. 2015. Deep and Shallow Gamification in Marketing: Thin Evidence and the Forgotten Powers of Really Good Games. (2015). https://doi.org/10.4018/978-1-4666-8342-6.CH006

[17] Minhua Ma, Andreas Oikonomou, and Lakhmi C. Jain. 2011. Serious Games and Edutainment Applications. 504 (2011). https://doi.org/10.1007/978-1-4471-2161-9

[18] Nikola Markiewicz, Miruna Russa, Annemiek Fokkens, Martin Dechant, and Maximilian A. Friehs. 2023. You Got It in Your Hands: Stop-Signal Modality Influences on Reactive Response Inhibition with Gaming Controls. *International Journal of Human-Computer Interaction* (2023). https://doi.org/10.1080/10447318.2023.2285624

[19] Sid Meier. 2018. Sid Meier's Interesting Decisions. https://www.youtube.com/watch?v=WggIdtrqgKg.

[20] Michael M. Morgan, MacDonald J. Christie, Luis De Lecea, Jason C. Halford, Josee E. Leysen, Warren H. Meck, Catalin V. Buhusi, Marcy J. Bubar, Kathryn A. Cunningham, Marcy J. Bubar, and et al. 2010. Stop-signal task. *Encyclopedia of Psychopharmacology* (2010), 1285–1285. https://doi.org/10.1007/978-3-540-68706-1_1120

[21] Joseph O'Hagan, M. Khamis, and J. Williamson. 2021. Surveying Consumer Understanding & Sentiment Of VR. *MMVE@MMSys* (2021). https://doi.org/10.1145/3458307.3460965

[22] Jesse Schell. 2020. *The Art of Game Design: A book of lenses.* CRC Press, Taylor & Francis Group.

[23] P. Schroeder, J. Lohmann, and M. Ninaus. 2021. Preserved Inhibitory Control Deficits of Overweight Participants in a Gamified Stop-Signal Task: Experimental Study of Validity. *JMIR serious games* (2021). https://doi.org/10.2196/25063

[24] Salete Teixeira, Raul V. Boas, Francisco Oliveira, Cristiana Araújo, and Pedro R. Henriques. 2020. OntoJogo: An Ontology for Game Classification. In *2020 IEEE 8th International Conference on Serious Games and Applications for Health (SeGAH).* 1–8. https://doi.org/10.1109/SeGAH49190.2020.9201876

[25] Jamie Thompson. 2022. A Guide to Abductive Thematic Analysis. *The Qualitative Report* (2022). https://doi.org/10.46743/2160-3715/2022.5340

[26] Katelyn Wiley, Phaedra Berger, Maximilian Achim Friehs, and Regan Lee Mandryk. 2024. Measuring the Reliability of a Gamified Stroop Task: Quantitative Experiment. *JMIR serious games* (2024). https://doi.org/10.2196/50315

[27] Katelyn Wiley, Sarah Vedress, and R. Mandryk. 2020. How Points and Theme Affect Performance and Experience in a Gamified Cognitive Task. *International Conference on Human Factors in Computing Systems* (2020). https://doi.org/10.1145/3313831.3376697

[28] K. Yoder and M. Belmonte. 2010. Combining computer game-based behavioural experiments with high-density EEG and infrared gaze tracking. *Journal of visualized experiments : JoVE* (2010). https://doi.org/10.3791/2320