Human-Centred ASL Learning: Effects of Interface Modality, Game Mechanics, and AI-Based Diagnostic Guidance

JINDI WANG

Human-Centred ASL Learning: Effects of Interface Modality, Game Mechanics, and AI-Based Diagnostic Guidance

Jindi Wang

A Thesis presented for the degree of
Doctor of Philosophy

Department of Computer Science
Durham University
United Kingdom
November 2025

Abstract

This dissertation presents an exploration of the integration of cutting-edge technologies to enhance American Sign Language (ASL) learning and to improve the efficacy of mediated human-human textual conversations. Emphasizing user experience (UX) design, it comprises a series of empirical studies that collectively demonstrate how user-defined interactions, gamification in camera-based environment can significantly augment traditional methods of learning.

This research investigates the development and evaluation of a user-defined hand gesture interface for learning static ASL. The findings indicate that custom gestures substantially increase user motivation and engagement, contributing valuable insights into the role of personalised interfaces in language acquisition. This is followed by a series of experiments focusing on the implementation of an innovative VR camera-based, gamified learning environment. These studies offer a critical comparison between 2D and 3D VR systems in the context of ASL education, revealing that while 3D environments enhance user engagement and satisfaction, they do not exhibit a significant advantage over 2D environments in terms of learning outcomes. This suggests that engagement and effectiveness in learning are not always linearly related.

Extending the scope of immersive virtual environments, the dissertation introduces a novel approach to ASL learning through immersive VR, specifically targeting numeric ASL. The approach is evidenced to elicit higher user engagement in comparison to conventional web-based methods, offering evidence that immersive learning environments can provide more satisfying and efficacious educational experiences.

Collectively, this dissertation contributes to the fields of educational technology and human-computer interaction. It offers empirical evidence supporting the integration of user-defined interactions and gamification in enhancing learning processes. This work provides interesting insights into the design of user-centric educational tools, highlighting the transformative potential of these technologies in revamping traditional methodologies and bolstering user engagement and educational experiences.

Declaration

The work in this thesis is based on research carried out at the Department of Computer Science, Durham University, United Kingdom. No part of this thesis has been submitted elsewhere for any other degree or qualification and it is all my own work unless referenced to the contrary in the text.

Acknowledgements

I would like to express my deepest gratitude to my thesis advisor, Dr Ioannis Ivrissimtzis, whose expertise, understanding, and patience, added considerably to my graduate experience. He provided me with the tools that I needed to choose the right direction and successfully complete my dissertation.

My time at Durham University was made enjoyable in large part due to the friendship and support of my colleagues and friends. I am also grateful for having a wonderful family who provided me with endless support and encouragement during my years of study. This accomplishment would not have been possible without them. Thank you, Dr Lei Shi, Dr Zhaoxing Li, Dr Yunzhan Zhou, Beining Zhang (RA) and Qilin Wang (RA), for cheering me on to the finish line and providing a constant source of comfort and encouragement. I thank my examiners, Prof. Effie Law (Internal Examiner) and Dr Fernando Loizides (External Examiner), for their careful reading of the thesis and for constructive feedback that strengthened the final manuscript.

Finally, I would like to acknowledge the financial support I received from the Department of Computer Science at Durham University, which allowed me to dedicate my time fully to my research.

This journey has been a transformative phase of my life, and I appreciate all who have been part of it and helped me along the way.

Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

1.1 Background and Motivation

The integration of technology in education has transformed traditional teaching methodologies, particularly in specialized domains like language learning. Innovations in Virtual Reality (VR), gamification, and Artificial Intelligence (AI) have redefined educational experiences, providing new avenues for engagement and interaction [1–3]. American Sign Language (ASL) learning, which relies heavily on visual comprehension and spatial awareness, presents unique challenges that these technologies aim to address [4].

Historically, traditional ASL teaching methods often lacked immersive and interactive elements crucial for effective language acquisition [5] as shown in Figure 1.1. The advent of VR has introduced realistic interaction platforms essential for effective language learning, providing environments that mimic real-life interactions [6,7]. Concurrently, gamification has proven to enhance learner engagement through elements of play, increasing both motivation and participation [8].

The application of VR in ASL education has evolved from simple virtual sign language dictionaries to complex interactive modules that allow users to engage

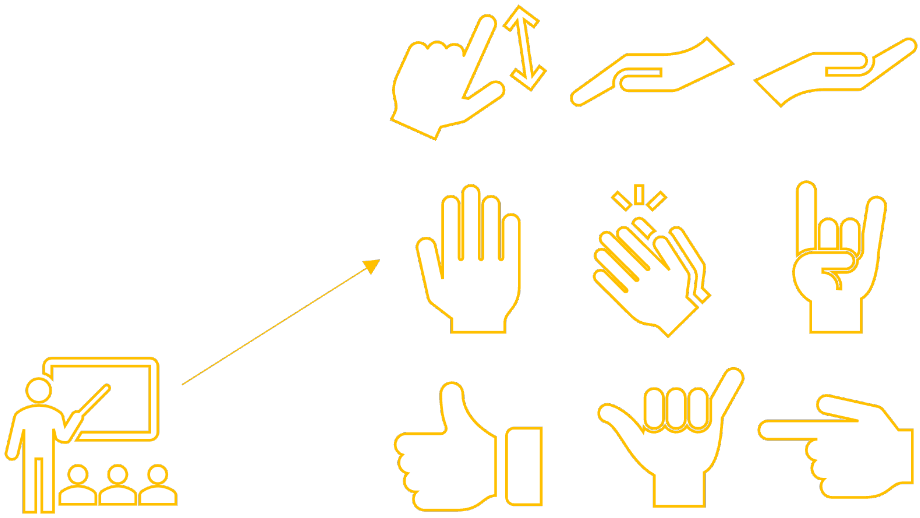

Figure 1.1: Traditional method to learn sign language.

in virtual conversations, improving both receptive and expressive language skills. These developments show significant promise in providing a deeper, more intuitive understanding of ASL syntax and grammar, essential components often difficult to master through traditional methods [9]. Moreover, gamification in ASL learning not only increases engagement but also introduces a competitive element that motivates learners to practice more diligently. By embedding ASL instruction within the framework of games, learners experience a less formal, more enjoyable approach to language acquisition, which can lead to increased long-term retention and more spontaneous language use [10].

This broadened approach not only addresses the technological enhancements in education but also emphasizes the evolving role of AI in facilitating communication, thus setting the stage for a detailed exploration of technology-enhanced learning and communication processes. Despite the potential benefits, integrating these technologies into a cohesive educational strategy remains a significant challenge. The effectiveness of VR, gamification, and AI in education and textual communication requires comprehensive research to explore their synergistic potential and practical implications. This thesis aims to bridge this gap by examining how these technologies can be integrated effectively to enhance both ASL education and human-human textual communication.

Despite recent advances, ASL education remains underserved by adaptive and immersive technologies. Many existing tools lack interactivity, personalisation, or accessibility for non-traditional learners, and empirical studies comparing interface modalities or gamified learning experiences are rare. As noted by Dalgarno and Lee [11], the affordances of immersive VR—particularly for spatially complex learning tasks—warrant deeper investigation in language learning contexts.

Moreover, gamification and AI hold promise not just for engagement but for scaffolding learner autonomy and real-time feedback—critical needs in ASL acquisition, where expressive feedback is non-verbal and gesture-based [12]. This research hypothesizes that the integration of immersive environments, game mechanics, and AI assistance can significantly enhance learning outcomes in ASL by reducing cognitive load, increasing motivation, and providing more natural interaction channels. The novelty lies not only in the combination of these tools, but in evaluating their comparative and synergistic impact through multiple focused studies.

1.2 Research Questions and Hypotheses

Based on the outlined objectives, the research questions (RQs) and corresponding hypotheses (H) guiding this dissertation are:

- **RQ1 (Chapter 3):** Does immersive VR enhance user engagement and learning outcomes in ASL education compared to traditional learning methods?

  H1: Participants using immersive VR will report significantly higher levels of engagement and usability (as measured by the UEQ and SUS) than those using non-immersive methods.

- **RQ2 (Chapter 4):** How does the integration of gamified interaction (e.g., a Whack-a-Mole game) affect user motivation and perceived enjoyment in ASL learning?

  H2: The addition of gamified modules will result in improved scores in the stimulation and novelty dimensions of the UEQ, compared to non-gamified VR.

- **RQ3 (Chapter 5):** How does allowing user-defined gestures in VR ASL platforms influence user experience and satisfaction?

  H3: Learners using self-defined gestures will show higher satisfaction and novelty ratings but may face higher cognitive load due to customisation efforts.

- **RQ4 (Chapter 6):** Does a 3D interface provide better usability and engagement than a 2D interface in ASL VR learning environments?

  H4: The 3D version will lead to higher scores in attractiveness and stimulation but lower dependability due to increased complexity.

- **RQ5 (Chapter 7):** To what extent does AI-assisted learning—comprising real-time sign recognition and adaptive, model-generated feedback—compared with a non-AI mode, influence learner accuracy and response time across repeated and final ASL assessment sessions?

  H5: Learners using AI-assisted ASL learning (with real-time sign recognition and adaptive, model-generated feedback) will demonstrate significantly higher accuracy and faster response times across repeated practice sessions and in the final assessment compared to learners using the non-AI mode.

These research questions are designed to be distinct yet complementary, collectively exploring how emerging technologies impact ASL learning from different perspectives: immersion, gamification, personalisation, and spatial design.

1.3 Research Methodology

1.3.1 Research Design

This dissertation comprises a sequence of interconnected empirical investigations designed to evaluate the efficacy of innovative technological interventions—namely Virtual Reality (VR), gamification, and Artificial Intelligence (AI)—in enhancing the learning of American Sign Language (ASL) and improving mediated communication. Each chapter of this dissertation describes a separate study, each with distinct research objectives:

Study 1: Immersive VR Environments for ASL Learning (it has been published in ECTEL 2023)

Chapter 3: This study explores the potential of immersive VR to enhance the learning experience of ASL learners, with a focus on the impact of immersive sensory environments on learner engagement and effectiveness [13].

Study 2: Gamification in VR Learning Environments (it has been published in INTERACT 2023)

Chapter 4: Investigates the effects of gamification techniques within VR environments on learning motivation and effectiveness, analyzing how game-like elements influence educational outcomes in ASL education [14].

Study 3: Customization of Gesture Interfaces in ASL Learning (it has been published in ITS 2023)

Chapter 5: Explores user-defined hand gesture interfaces within VR settings, aiming to understand how personalization of gesture controls affects learning engagement and retention in ASL [15].

Study 4: 2D vs. 3D VR Learning Environments (it has been published in IEEE VR 2024)

Chapter 6: Compares the effectiveness of 2D and 3D virtual environments on user experience and learning outcomes, providing insights into optimal environmental configurations for ASL instruction [16].

Study 5: AI vs non-AI Learning Comparison

Chapter 7: Conducted a controlled, repeated-measures study comparing an AI-assisted ASL learning interface (real-time recognition with targeted feedback) against a non-AI, predefined-materials condition across three rounds, evaluating effects on assessment accuracy, response time, and game performance.

These studies employ a combination of quantitative and qualitative methodologies to provide robust and comprehensive insights into the applied technologies.

1.3.2 Participants

The participant recruitment strategy was designed to ensure a comprehensive analysis across various demographics, educational backgrounds, and technological proficiencies, corresponding to the nuanced requirements of each study in this dissertation.

General Recruitment Strategy

Participants were recruited from a wide array of sources, including university databases, online forums dedicated to language learning and technology, social media platforms, and local community centres known for educational outreach. Efforts were made to ensure a diverse participant pool in terms of age, gender, and educational background to enhance the generalizability of the study results.

Study-Specific Recruitment Details

Studies Involving VR (Studies 1, 3, and 4) Recruitment targeted individuals from the general population with varying degrees of familiarity with ASL—from beginners to intermediate learners—who had limited prior exposure to VR technologies to mitigate any pre-existing biases towards the technology.

Study Involving Gamification (Study 2) This study specifically sought individuals with prior gaming experience or exposure to gamified educational tools, aiming to assess the differential impacts of gamification on individuals accustomed to such environments.

Demographic Information Collection

Detailed demographic information was collected through pre-study surveys, which included questions on age, gender, educational attainment, proficiency in ASL, and familiarity with VR and AI technologies.

1.3.3 Data Collection Methods

Data were systematically collected across all studies to ensure comprehensive coverage of both qualitative and quantitative aspects. The following subsections detail the data collection methods employed in each study:

Quantitative Data Collection

- **Surveys and Questionnaires:** Standardized surveys including pre- and post-study questionnaires were administered to capture data on user experience, learning outcomes, and participant satisfaction with the technological interventions.

Qualitative Data Collection

- **Interviews:** Semi-structured interviews were conducted post-interaction to gather in-depth insights into participant perceptions, experiences, and any challenges faced during the studies.
- **Focus Groups:** For studies involving communication tools as mentioned in Chapter 8, focus group discussions were utilized to explore collective views and debate the implications of AI features on communication efficacy.

1.3.4 Instruments and Tools

The following instruments and technological tools were utilised to facilitate data collection and ensure the accuracy of data analysis:

Data Collection Instruments

- **System Usability Scale (SUS):** Employed to quantitatively measure the usability of the VR and AI systems used in the studies.
- **Custom Questionnaires:** Developed specifically for each study to assess variables pertinent to the research questions, such as user engagement, learning effectiveness, and AI interaction satisfaction.

1.3.5 Data Analysis

Data collected from the various studies were analysed using both qualitative and quantitative methods to provide a holistic understanding of the research questions.

Quantitative Analysis

- **Statistical Testing:** Data were subjected to statistical tests such as ANOVA, chi-square tests, and regression analysis to determine significant differences and relationships among the study variables.
- **Learning Analytics:** In VR and gamification studies, user data analytics techniques were used to analyse interaction data and assess learning progression and outcomes.

Qualitative Analysis

- **Thematic Analysis:** Used to identify, analyse and report patterns within qualitative data. This method was instrumental in interpreting diverse participant responses during interviews and focus groups.
- **Content Analysis:** Employed to systematically categorise text data from interviews and open-ended survey responses to quantify and analyse the presence, meanings, and relationships of such words and concepts.

1.3.6 Ethical Considerations

All study procedures adhered to the ethical guidelines provided by the Institutional Review Board (IRB) at Durham University. Participants were informed about the study's purpose, their right to withdraw at any time, measures taken to ensure confidentiality, and the handling of their data post-study. Informed consent was obtained from all participants prior to their involvement in the research activities.

1.3.7 Authorship and implementation

Chapters 3–6 were designed by the author with targeted code contributions from two research assistants under the author's direction; the author integrated the system, ran all sessions, and performed all analyses. Chapter 7 (AI vs. non-AI) was entirely by the author—design, implementation, data collection, and analysis.

1.4 Significance of the Study

1.4.1 Results

This dissertation presents a comprehensive evaluation of the impact of integrating Virtual Reality (VR), gamification, and Artificial Intelligence (AI) into American Sign Language (ASL) education and mediated human-human textual communication. The results demonstrate significant advancements, as detailed below:

- **Educational Impact:** The application of VR and gamification in ASL education has significantly enhanced learner engagement and retention. Empirical evidence confirms that these technologies elevate the learning experience, making it more immersive and effective, particularly in enhancing communication skills within the deaf community.

- **Technological Advancements:** This research provides substantial insights into the operational effectiveness and utility of VR and AI within educational settings. It outlines both the challenges and benefits of these technologies, offering a roadmap for their future application.

- **Development of AI-Driven Communication Tools:** The study's findings on AI chat assistants illustrate their capability to significantly improve the quality and efficiency of mediated textual communication, making these interactions more adaptable and user-centric.

1.4.2 Relevance

The broader implications of this research are profound, impacting various societal, policy, and economic areas:

- **Social Implications:** By improving ASL educational tools, this work significantly contributes to educational inclusivity and accessibility for individuals with hearing impairments, promoting greater societal integration.
- **Policy and Educational Reforms:** The empirical evidence provided may inform policy-making and lead to reforms that advocate for the integration of advanced technologies in educational frameworks, thereby promoting equality in educational opportunities.
- **Economic Impact:** Integrating VR and AI into educational models has the potential to reduce dependence on physical resources and increase access through digital platforms, leading to more cost-effective educational approaches.

1.5 Thesis Structure

The thesis is meticulously organized into chapters that delve into the application of cutting-edge technologies such as Virtual Reality (VR), gamification, and Artificial Intelligence (AI) within the context of American Sign Language (ASL) education and mediated textual communication. This structured approach provides a coherent narrative from the exploration of technological implementations to their practical outcomes and theoretical implications. Below is an outline of the thesis structure with corrections:

1. **Chapter 2: Literature Review**
   The literature review presents a synthesis of existing research related to the technologies being investigated. It evaluates previous studies on VR and gamification in education, particularly in language learning, and discusses the evolution of AI-driven communication tools in educational settings.

2. **Chapters 3 to 5: VR and Gamification in ASL Learning**
   These chapters focus on the application of VR and gamification in ASL education:

   - **Chapter 3:** Examines the use of immersive VR environments and their impact on the engagement and comprehension of ASL among learners.
   - **Chapter 4:** Analyzes the role of gamification elements in motivating learners and enhancing their participation and retention rates.
   - **Chapter 5:** Discusses the integration of VR and gamification, presenting a combined analysis of their synergistic effects on learning outcomes.

3. **Chapter 6: Comparative Analysis of VR Implementations**
   This chapter explores different VR learning environments, comparing their effectiveness in enhancing ASL education. It specifically focuses on assessing the comparative advantages of various VR technologies and their configurations in supporting ASL acquisition.

4. **Chapter 7: AI-Assisted vs Non-AI (Human-Centred Comparative Study)**
   A controlled repeated-measures experiment (three rounds: pre/learn/post + timed game) compares AI-assisted feedback with a non-AI mode using identical assessments. We analyse accuracy, response time, improvement, retention, learning time, adherence, and game scores (linear mixed-effects), plus interviews via TF–IDF+NMF with SBERT validation. Both groups improve; AI generally achieves higher post-session accuracy and lower response times without extra time-on-task. Qualitatively, "clarity" diverges (perception- vs. correction-based learning), with calls for transparent, controllable replay artefacts.

5. **Chapter 8: Discussion**
   An integrated synthesis in Chapters 3-7 answers the RQs: immersion / gamification increases engagement but requires diagnostic specificity to produce durable gains; 2D can match 3D when cues and feedback are dependable; AI's

advantage reflects guidance quality rather than study time. We situate results in motor-learning and feedback theory, note mixed/null findings (e.g., learning-time parity, condition-dependent variance).

6. **Chapter 9: Conclusion and Future Work**
   The final chapter synthesizes the findings from the individual studies, discusses the implications for educational technology and communication, and proposes directions for future research in the field.

1.6 Publication to Date

Note on publications included in this thesis: At the time of submission, six chapters of this thesis are heavily based on papers submitted for publication or published in conferences and journals:

- **Chapter 3: J Wang**, I Ivrissimtzis, Z Li, Y Zhou, L Shi. ***Exploring the Potential of Immersive Virtual Environments for Learning American Sign Language***. European Conference on Technology Enhanced Learning, pp. 459–474. [13]

- **Chapter 4: J Wang**, I Ivrissimtzis, Z Li, Y Zhou, L Shi. ***Developing and Evaluating a Novel Gamified Virtual Learning Environment for ASL***. IFIP Conference on Human-Computer Interaction, pp. 459–468. [14]

- **Chapter 5: J Wang**, I Ivrissimtzis, Z Li, Y Zhou, L Shi. ***User-defined hand gesture interface to improve user experience of learning American sign language***. International Conference on Intelligent Tutoring Systems, pp. 479-490. [15]

- **Chapter 6: J Wang**, I Ivrissimtzis, Z Li, L Shi. ***Comparative Efficacy of 2D and 3D Virtual Reality Games in American Sign Language Learning***. The 31st IEEE Conference on Virtual Reality and 3D User Interfaces. [16]

Note on publications not included in this thesis: In addition to the publications mentioned earlier, several other works were published during the research period for this thesis that contributed to my understanding of the subject matter. However, these publications do not directly align with the main narrative of this thesis and therefore are not incorporated into the main body of the text.

- Zhaoxing Li, Jujie Yang, **Jindi Wang**, Lei Shi, Jiayi Feng, Sebastian Stein. ***LBKT: A LSTM BERT-based Knowledge Tracing Model for Long-Sequence Data***. The 20th International Conference on Intelligent Tutoring Systems.

- **J Wang**, I Ivrissimtzis, Z Li, L Shi. ***Hand Gesture Recognition for User-defined Textual Inputs and Gestures***. Universal Access in the Information Society.

- Z Li, L Shi, **J Wang**, AI Cristea, Y Zhou. ***Sim-GAIL: A generative adversarial imitation learning approach of student modelling for intelligent tutoring systems***. Neural Computing and Applications 35 (34), 24369-24388.

- Z Li, M Jacobsen, L Shi, Y Zhou, **J Wang**. ***Broader and Deeper: A Multi-Features with Latent Relations BERT Knowledge Tracing Model***. European Conference on Technology Enhanced Learning, 183-197.

- Y Zhou, L Shi, Z He, Z Li, **J Wang**. ***Design Paradigms of 3D User Interfaces for VR Exhibitions***. IFIP Conference on Human-Computer Interaction, 618-627.

- Z Li, L Shi, Y Zhou, **J Wang**. ***Towards student behaviour simulation: a decision transformer based approach***. International Conference on Intelligent Tutoring Systems, 553-562.

- C Xiao, B Sun, **J Wang**, M Liu, J Feng. ***Breaking through Inequality of Information Acquisition among Social Classes: A Modest Effort on Measuring "Fun"***. Proceedings of the Second Workshop on NLP for Positive Impact (NLP4PI), 101-112.

CHAPTER 2

Literature Review

2.1 Introduction

This chapter presents a literature review that maps the main strands of research underpinning the empirical work in this thesis. The review is organised around five themes that align directly with the experimental chapters: immersive VR for sign language learning (Chapter 3), gamification in education (Chapter 4), gesture-based learning interfaces (Chapter 5), 2D vs. 3D interfaces in education (Chapter 6), and AI/LLM-assisted feedback (Chapter 7). Each section synthesises representative work, highlights gaps, and explains how these motivate the subsequent experiments.

The review is best characterised as a *scoping, narrative review* rather than a formal Systematic Literature Review (SLR). The goal is to identify and synthesise influential and illustrative studies that inform the design decisions and research questions in this thesis, rather than to exhaustively catalogue every published article on each topic according to a pre-registered protocol. Accordingly, the search and selection process is described transparently in Section 2.2, but this chapter does not claim adherence to the PRISMA framework or to SLR conventions such as multi-stage screening and structured coding by multiple reviewers.

2.2 Approach to the Literature Review

2.2.1 Nature and purpose of the review

The review adopts a *scoping, narrative* approach that is appropriate for a design-oriented, multi-study thesis spanning several related domains: sign language learning technologies, immersive and non-immersive environments, gesture-based interfaces, gamification, and AI-assisted feedback and assessment. Rather than evaluating a single narrowly defined intervention, the chapter integrates diverse strands of work to:

- identify key application areas and design patterns relevant to technology-supported sign language learning;
- synthesise how different types of interfaces (2D/3D, VR, gesture-based, gamified, AI-assisted) have been used in educational contexts;
- highlight conceptual and practical gaps that motivate the research questions and system designs in later chapters.

The emphasis is therefore on conceptual coverage and design relevance, not on statistical aggregation or meta-analysis.

2.2.2 Search strategy

The literature search was conducted iteratively using Google Scholar and the ACM Digital Library as primary sources, with IEEE Xplore and SpringerLink consulted where appropriate. The main search period covered publications from 2010 onwards, with earlier foundational work included when identified through citation chaining. Searches were carried out between 2020 and 2023.

To identify studies relevant to the five themes of the thesis, combinations of the following queries were used:

"Virtual Reality AND sign language learning"

"Gamification AND education"

“Gesture-based learning interfaces”

“2D vs 3D VR in education”

“AI OR LLM-assisted feedback in language learning”

Search terms were adjusted as new terminology emerged (for example, specific platform names or model types) and as relevant references were discovered in the bibliographies of key papers. Backward citation chaining (examining reference lists) and forward citation searches (examining later works that cite key papers) were used to broaden coverage.

2.2.3 Inclusion and exclusion criteria

The following criteria guided the inclusion of studies in the review:

- **Topical relevance:** The study addressed at least one of the core themes of the thesis: sign language learning and practice; VR/AR or 3D environments for learning; gesture-based or vision-based interfaces in education; gamification or serious games in learning contexts; or AI/LLM-supported feedback and assessment.
- **Publication type and quality:** Priority was given to peer-reviewed journal articles, conference papers, and scholarly book chapters written in English.
- **Educational focus:** Studies were required to have a clear educational or training component, rather than focusing exclusively on technical or engineering performance.

Studies were generally excluded if they:

- focused primarily on clinical rehabilitation or therapeutic applications without a transferable educational dimension;
- reported only technical benchmarks for recognition models or hardware without involving learners or learning tasks;
- dealt with sign languages solely from a linguistic theory perspective, without technology-mediated learning or practice.

These criteria were applied pragmatically. Because the review is scoping and narrative rather than strictly systematic, some borderline cases that offered particularly relevant design insights were retained even if they did not match every criterion perfectly (for example, technical papers that included illustrative educational scenarios).

2.2.4 Screening and selection

The selection process proceeded in three main stages:

1. **Initial screening of titles and abstracts.** Search results were first screened at the level of title and abstract to remove clearly irrelevant material (e.g. work on unrelated medical imaging or non-educational natural language processing).

2. **Full-text inspection and thematic grouping.** For the remaining items, full texts were consulted to confirm relevance to one or more of the five themes. At this stage, studies were grouped thematically according to their primary contribution (e.g. immersive sign language systems, gamified learning environments, gesture-based interaction, 2D vs. 3D comparisons, AI/LLM-supported feedback).

3. **Refinement via citation chaining.** Within each theme, key studies served as anchors for backward and forward citation searches, which added further relevant work and, in some cases, led to the substitution of earlier or less complete studies with more recent or comprehensive ones.

The search initially yielded several hundred records across the different query combinations. After de-duplication, screening and full-text inspection, a subset of approximately one hundred studies was retained as most relevant to the aims of this thesis. In this chapter, the focus is on representative studies that together provide an overview of the field and highlight major trends and research gaps; the intention is not to exhaustively review every included work.

2.2.5 Data extraction and synthesis

For the studies that were retained, information was extracted on:

- the learner population and learning goals;
- the type of technology or interface used (e.g. desktop 2D, VR/AR, gesture-based, mobile game);
- the role and form of feedback and/or assessment;
- study design and key findings related to performance, engagement, and user experience;
- design implications, limitations, and open questions reported by the authors.

The findings are synthesised narratively within thematic sections. Rather than computing aggregate effect sizes, the synthesis compares design choices, highlights converging and diverging findings, and traces how these patterns motivate the research questions, hypotheses, and system design decisions developed in Chapters 3–7.

2.3 Immersive VR for Sign Language Learning (Chapter 3)

In line with the scoping, narrative approach described above, this section highlights representative work on immersive VR for sign language learning and related educational contexts that informs the design and motivation of Chapter 3.

Virtual reality (VR) has been widely applied in education for its ability to foster immersion, presence, and embodied interaction. Within language learning, [17] developed an immersive tutor for American Sign Language (ASL), showing that learners reported higher engagement when practising in VR. Similarly, [18] evaluated a VR-based sign practice system and found positive usability results, particularly for repetitive rehearsal of gestures. Both studies demonstrate the promise of VR in ASL education, but do not compare against alternative platforms. Broader VR research confirms these benefits. [19] reviewed immersive learning technologies and concluded that VR enhances motivation through embodied presence. [20] conducted controlled studies in VR science education, finding higher spatial learning gains than 2D materials. Likewise, [21] reported that immersive VR produces stronger engagement than 2D, though not in ASL contexts. Interaction modalities have also been investigated. [22] highlighted how naturalistic gesture interaction in VR improves engagement, while [23] showed immersive gesture-based learning increases satisfaction compared with traditional interfaces. In the ASL domain, [24] designed a gamified VR platform for sign practice, but evaluation was limited to usability reports.

Across these works, VR demonstrates educational benefits, but there is no controlled study comparing immersive VR with non-immersive, web-based platforms in ASL learning. Chapter 3 addresses this by experimentally contrasting ASL numeral learning in a VR environment versus a web environment.

2.4 Gamification in Education (Chapter 4)

Following the same scoping logic, this section surveys representative literature on gamification in education with a focus on implications for ASL learning and the design of Chapter 4.

Gamification has emerged as a strategy for increasing motivation in learning contexts. [25] showed that gamified VR environments improve collaboration and motivation. In language education, [26] demonstrated that gamification fosters better retention. [27] similarly found positive impacts of gamification on student engagement across digital platforms. Specific to sign language, [12] created a gamified ASL environment that encouraged practice through interactive challenges. However, evaluation was limited in scope and lacked a controlled study of learning outcomes. Other reviews, such as [28], confirm that gamification boosts motivation in education broadly, but stress the need for domain-specific evaluation. [25] and [21] also note that gamification often amplifies the benefits of immersive environments, though the empirical evidence for ASL contexts is absent. [29] examined mobile-based language apps and found that game mechanics improve retention, yet these findings have not been extended to VR-based ASL practice.

While gamification is proven to motivate learners, there is little rigorous evaluation in ASL-specific VR settings. **Motivation:** Chapter 4 addresses this by developing a gamified "Whack-a-Mole" ASL game and evaluating its effects on motivation and learning.

2.5 Gesture Interfaces and Personalisation (Chapter 5)

This section focuses on gesture-based interfaces and personalisation as they relate to ASL and to the study design in Chapter 5.

Gesture interfaces provide natural interaction in immersive learning. [30] reviewed gesture technologies in VR, confirming they enhance immersion. [19] studied gesture use in immersive contexts and found learners valued the embodied experience. In sign language research, [31] advanced neural recognition models for ASL, demonstrating technical accuracy but with limited pedagogical focus. [32] and [33] both used gesture recognition in VR for ASL, yet the systems relied on predefined gestures, limiting learner personalisation. Beyond ASL, [34] evaluated gesture interfaces for general education tasks, showing usability benefits but again without customisation. [35] and [36] highlight that gesture input contributes to engagement in VR language learning, but neither investigated the effects of user-defined gestures.

Most gesture-based ASL systems emphasise recognition accuracy but not learner personalisation. Chapter 5 investigates predefined versus user-defined gestures, examining their impact on usability and satisfaction.

2.6 2D vs 3D VR Interfaces (Chapter 6)

Here, the focus shifts to comparisons between 2D and 3D/VR interfaces, and how this literature motivates the study in Chapter 6.

Across STEM and skills training, 3D/VR often yields higher *presence* and *engagement* than 2D media, with several studies reporting stronger subjective immersion and, in some cases, improved learning [17, 37–41]. However, the evidence on *learning outcomes* is mixed: advantages emerge when 3D affordances (e.g., spatial manipulation, multi-view inspection) are instructionally leveraged, but diminish when content and practice do not require those affordances [42]. Notably, there are no controlled comparisons of *2D vs. 3D* interfaces for ASL learning; prior work using VR for ASL does not isolate dimensionality effects [17], and mobile 3D gains have been shown outside the ASL context [43].

Positioning of Chapter 6. We therefore test whether the well-documented presence advantage of 3D translates into measurable ASL vocabulary gains relative to a matched 2D interface, holding pedagogy and practice dosage constant.

2.7 AI and LLM-Assisted Feedback (Chapter 7)

Finally, this section reviews work on AI- and LLM-assisted feedback and tutoring, with a focus on implications for ASL learning and the experiment in Chapter 7.

Artificial intelligence (AI) tutors are increasingly used in education. [44] demonstrated that intelligent tutoring systems improve learning outcomes through adaptive feedback. Recent reviews, such as [45], highlight ChatGPT's potential in providing contextualised feedback for students. [46] also identified AI feedback as a critical trend in immersive learning. In sign language, AI has mostly been applied to recognition. [47] proposed transformer-based ASL recognition, while [48] used gesture recognition for ASL tutoring. These works show technical progress but do not provide adaptive feedback for learners. [49] and [50] explored AI-enhanced language apps, reporting higher personalisation, but not applied to ASL. [51] noted the integration of AI into VR environments, but without empirical evaluation of tutoring effects.

While AI/LLMs are emerging as powerful tutors, no studies apply them to ASL learning. Chapter 7 pioneers this by comparing an LLM-assisted adaptive learning path with a predefined one for ASL.

CHAPTER 3

Exploring the Potential of Immersive Virtual Environments for Learning American Sign Language

In this chapter, we elaborate on the study comparing immersive virtual environments with websites for learning American Sign Language (ASL). The research presented in this chapter was published at the 2023 European Conference on Technology Enhanced Learning.

3.1 Introduction

The World Health Organization (WHO)[1] predicts that, by 2050, about 2.5 billion people will have some degree of hearing loss, and at least 700 million of them will need some sort of hearing rehabilitation. The rehabilitation training procedures for people with hearing loss include the use of sign language and various alternative sensory techniques such as voice reading, writing with fingers on the palm of the hearing-impaired, and vibration sensing. While the most common form of communication for those who are deaf is sign language, most people without hearing loss have never

[1] https://www.who.int/zh/news-room/fact-sheets/detail/deafness-and-hearing-loss

studied sign language, making communication between these two groups challenging. Thus, the learning of sign language has become a key topic in educational research to break down communication barriers between diverse groups.

In the past years, face-to-face teaching of sign languages has been severely affected by the COVID-19 restrictions, with alternative online approaches filling in the gap. The majority of the latest approaches to teaching sign languages [52–54] employ website-based tools, while approaches based on immersive virtual reality (VR) technology [14] are more sparse in the literature. The following are some of the few studies on the VR-based approach. Adamo *et al.* [55] proposed the use of an immersive 3D learning environment to increase the mathematical skills of deaf children by teaching mathematical concepts and ASL's math terminology through user interaction with fantasy 3D virtual signers and environments. Schioppo *et al.* [56] proposed a sign language recognition method using features extracted from data acquired by a Leap Motion controller from an egocentric view. The method was tested on the 26 letters of the ASL alphabet. In a related development, Phan *et al.* [57] used motion tracking to trial a number of different methods for providing user feedback in a sign language learning system.

For the purposes of our study, we created a learning environment employing widely used VR technology to provide users with an immersive environment for learning numbers 0-9 in American Sign Language (ASL). Several issues with existing website-based and immersive approaches were identified and addressed in the design of our system, including small datasets for training the gesture recogniser, a lack of real-world settings, and most importantly, a lack of user satisfaction for sustained engagement with the system. To improve user experience, inspired by the ASL Sea Battle [12], a sign language game that was created by Bragg *et al.* to facilitate the gathering of user data, we developed and introduced into the system a Whack-a-Mole type of game.

To the best of our knowledge, no previous user studies are focusing on comparing website-based with VR-based systems, especially for learning 0-9 in ASL. Hence, we conducted a user study based on the survey scheme proposed by Schrepp *et al.* [58], aiming at comparing differences in user experience between an immersive and a website-based environment. The website-based environment was developed by Quizlet [2]. To make fair comparisons, the design aimed at maximising the consistency between the two environments.

Summarising, the main research question motivating our work: ***"Does an immersive ASL learning environment provide a better user experience compared to a web-based learning environment?"***, was looked into within the context of an immersive environment for learning numbers between 0 and 9 in ASL. Our main contributions are as follows:

1. We implemented a novel immersive virtual environment for ASL learning with a Whack-a-Mole type of game.

2. We provide initial evidence that immersive virtual environments can enhance users' learning experience and engagement when compared to website-based learning environments for ASL. This suggests that incorporating immersive elements into ASL education may be a promising direction for improving learning outcomes and user satisfaction.

3.2 Related Work

First, we briefly review sign language recognition, which is a crucial technology for computer-assisted sign language learning and an important technical part of our system. Then, we go over some current research on web-based sign language learning, concentrating on the issues of effectiveness and usability of the website's features. Finally, we review research on sign language as a communication tool in general, going beyond the learning of the language.

Sign language recognition: Bheda *et al.* [59] proposed a method based on deep convolutional neural networks (CNNs) to recognize images of gestures of the letters and digits in ASL. Kim *et al.* [60] proposed a novel sign language recognition method, which employs an object detection network for a region of interest (ROI) segmentation to preprocess the input data. Battistoni *et al.* [61] described a

[2]https://quizlet.com/560702085/asl-numbers-0-9-flash-cards/

method for ASL alphabet recognition based on CNNs, which allows for monitoring the users' learning progress. Jiang *et al.* [62] proposed a novel fingerspelling identification method for Chinese Sign Language via AlexNet-based transfer learning and evaluated four different methods of transfer learning. Camgoz *et al.* [63] introduced a novel transformer-based architecture that jointly learns Continuous Sign Language Recognition and Translation while being trainable in an end-to-end manner. Zhang *et al.* [64] proposed MediaPipe Hands, a real-time on-device hand tracking pipeline to compute hand landmark positions from a single RGB camera frame for AR/VR applications. Goswami *et al.* [65] created a new dataset for ASL recognition and used it to train a CNN-based model for hand gesture recognition and classification. Finally, Pallavi *et al.* [66] presented a deep learning model based on the YOLOv3 architecture, reporting high recognition rates on the ASL alphabet.

Having reviewed the existing work on sign language recognition, we concluded that Mediapipe is the most suitable tool for the purposes of this paper, and thus, we used it for sign language recognition, benefiting from its highly accurate, real-time detection of hand landmark points. Moreover, as an open-source hand gesture detection framework from Google, it is well-documented and supported.

Website-based sign language learning: Kumar *et al.* [67] proposed a sign language translation system based solely on visual input, employing deep learning for accurate translation. Joy *et al.* [68] proposed SignQuiz, a finger-spelt sign learning application for Indian sign language (ISL), utilizing automatic sign language recognition techniques. Vaitkevičius *et al.* [69] presented a system capable of learning gestures using the data from the Leap Motion device and classifying them with Hidden Markov Classification (HMC). Bird *et al.* [70] used a late fusion approach for sign language recognition from multi-modal data. They significantly improved the overall accuracy compared to single-modality image classification (88.14%) and Leap Motion data classification (72.73%). John *et al.* [71] proposed a system aiming at handling regional variations in vocabulary and grammar through a common vision-based platform. Empe *et al.* [52] developed a smartphone app called SimboWika to help deaf primary school kids learn Filipino Sign Language. They were motivated by the observation that communication between those who have hearing loss and those who do not can be challenging, especially when considering the unique perspective of persons with hearing loss language, and the lack of proficiency in sign language by those without hearing loss. Estrada *et al.* [53] proposed a web tool aimed at kids with and without hearing difficulties. They used a game to help children learn sign language through play. Patricks *et al.* [54] developed the interactive website application Sign2Sign, integrating real-time sign language detection AI, dynamic 3D avatars, and conversation-focused sign language instruction to support sign language education.

We note that, despite the significant size of some of the research projects behind these proposed methods and systems, there is no systematic evaluation of how users are affected by their interactions with these systems and, ultimately, how well the sign language students are progressing. Thus, we argue that user research is necessary, and the questionnaire interview format seems to be appropriate for an initial investigation.

Sign language applications: Bantupalli *et al.* [72] presented a vision-based application for translating sign language to text, thus aiding communication between signers and non-signers. Their recognition model takes video sequences and extracts temporal and spatial features from them. Schnepp *et al.* [73] argued that an animated sign language dictionary is a valuable resource for caregivers learning to communicate with residents who use sign language. They developed such a tool using a human-centred design methodology. Samonte [74] developed an e-tutor system, assisting instructors with course delivery and assessment. Economou *et al.* [21] developed a Serious Game (SG) aiming at closing the communication gap between the able hearing people and those with hearing impairment. Their tool facilitates sign language learning, specifically targeting the adult population. Wang *et al.* [15] developed a gamified sign language environment with characteristics that users could personalise. They found that gamification improved user experience. The research we reviewed on sign language applications shows that dictionary searches and gamification can improve users' motivation, which informed our choice to include such features in our system design.

Our survey of the literature verified that, as we have already noted, the majority of the existing work utilizes web-based solutions for the creation of sign-language learning interfaces. This is often the most convenient solution, making it easy for the user to access all the learning materials. However, on the other hand, such an approach can lead to users developing a sense of repetitiveness and feelings of boredom. This is an observation supported by user feedback on such systems, and in fact, some users would drop out of the learning process altogether due to a lack of motivation [75].

Thus, the main objective of our work is to investigate whether virtual environments can provide a better user experience in sign language learning. We built a VR-based system that provides users with an immersive experience and added a quiz and a small game to stimulate users' interest and improve their experience. We note that recent advances in VR technology mean that there is a more general trend of migration of online tasks from web-based systems on the 2D planar screen to the immersive 3D space [76]; our system does something similar for sign language learning. Moreover, noting the absence of user research on the subject, in order to determine if an immersive learning environment can indeed improve the user experience when learning a sign language, we invited two groups of users to undertake an interview survey. Our aims were to evaluate our system and compare the user experience of learning 0-9 in ASL with a website-based learning environment.

3.3 User Interface of Immersive Environment and Website

This section provides an overview of the key components of the proposed immersive environment and the main features and user interfaces to compare with the web-based learning environment, as shown in Figure 3.2 and Figure 3.3.

3.3.1 The Immersive Learning Environment

As shown in Figure 3.2, the immersive environment was designed with clear spatial separation of instructional and interactive elements. The dictionary wall (centre)

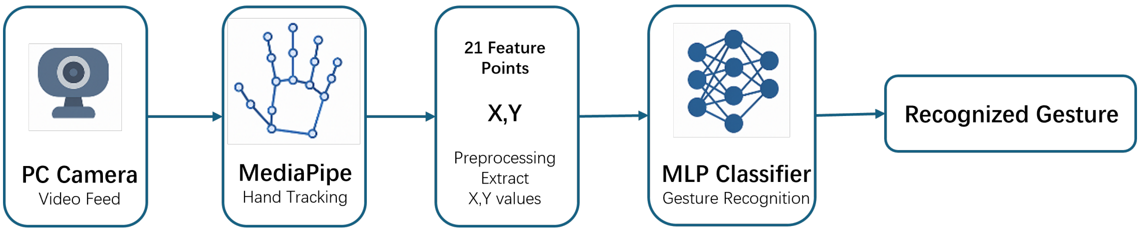

Figure 3.1: System architecture for hand gesture recognition using 21 hand landmark points. The PC camera captures real-time video input, which is processed by MediaPipe to extract 21 feature points (x, y coordinates) from the hand. These features are preprocessed and then fed into a Multi-Layer Perceptron (MLP) classifier to recognize hand gestures, producing the final output.

helps users learn ASL numbers, while the right instruction panel guides their task sequence. This layout was chosen to minimise confusion and promote flow in the learning experience. The whole scene was created in Unity (2020.3.32f1). The system supports basic object interaction through gaze-based selection (e.g., clicking or picking an object after 3 seconds of fixed attention).

The image acquisition was done by an integrated camera, linked to a PC using openCV (3.4.2) [77]. Regarding gesture recognition, Mediapipe was used to detect the user's hand and extract a sequence of 21 feature points (p_0, p_1, p_2,...,p_{20}), corresponding to landmarks on the detected hand. We set p_0, the point at the bottom of the palm near the user's wrist, as the origin of the frame's coordinate system. Let (x_i, y_i) be the coordinates of the point p_i. They are normalized by

$$x_i = \frac{x_i - x_0}{x_{max}}, \quad y_i = \frac{y_i - y_0}{y_{max}}, \qquad i = 1, 2, \ldots, 20. \tag{3.1}$$

where

$$\begin{aligned} x_{max} &= max(\,|x_1 - x_0|, \quad |x_2 - x_0|, \quad \ldots, |x_{20} - x_0|\,) \\ y_{max} &= max(\,|y_1 - y_0|, \quad |y_2 - y_0|, \quad \ldots, |y_{20} - y_0|\,) \end{aligned} \tag{3.2}$$

and are then fed as a feature vector to the classifier as shown in the Figure 3.1. The classifier is a multilayer perceptron consisting of three fully connected layers, implemented in Python 3.6 and Tensorflow 2.6.0. We trained the classifier on a commodity PC with an RTX3080 GPU. The obtained recognition accuracy rates were above 90%, a result that was deemed sufficient for the purposes of this study as it is expected to support an overall smooth user experience.

The implemented user interfaces comprise four different modules: **Instructions**, **Sign Language Dictionary**, **Quiz**, and **Whack-a-Mole Game**, respectively. Each module is described in more detail below.

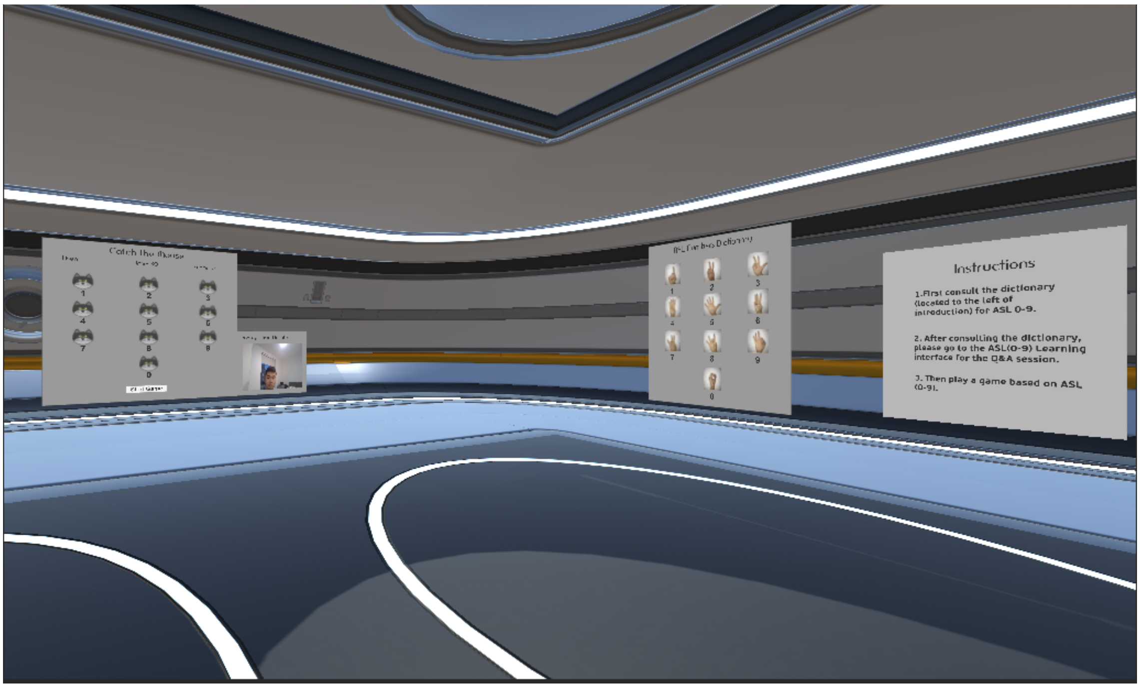

Figure 3.2: The user's perspective inside the VR learning environment, showing the dictionary wall in the centre, the instruction panel on the right, and game interface objects distributed on the left. Users interact with the system using hand gestures recognised by the front-mounted camera.

Instructions

In order to make the system easy to use, we created an **Instructions** interface, which also serves as the point of entry when the user logs into the system. In this interface, the user is introduced to the following three basic steps of the learning process, see also Fig. 3.4 (right).

1. Consulting the numbers 0–9 in an ASL dictionary (located to the left of the **Instructions**) to familiarise themselves with ASL 0–9 for about 5 minutes.

2. Users can self-assess their study of ASL 0–9 by accessing the **Quiz** module.

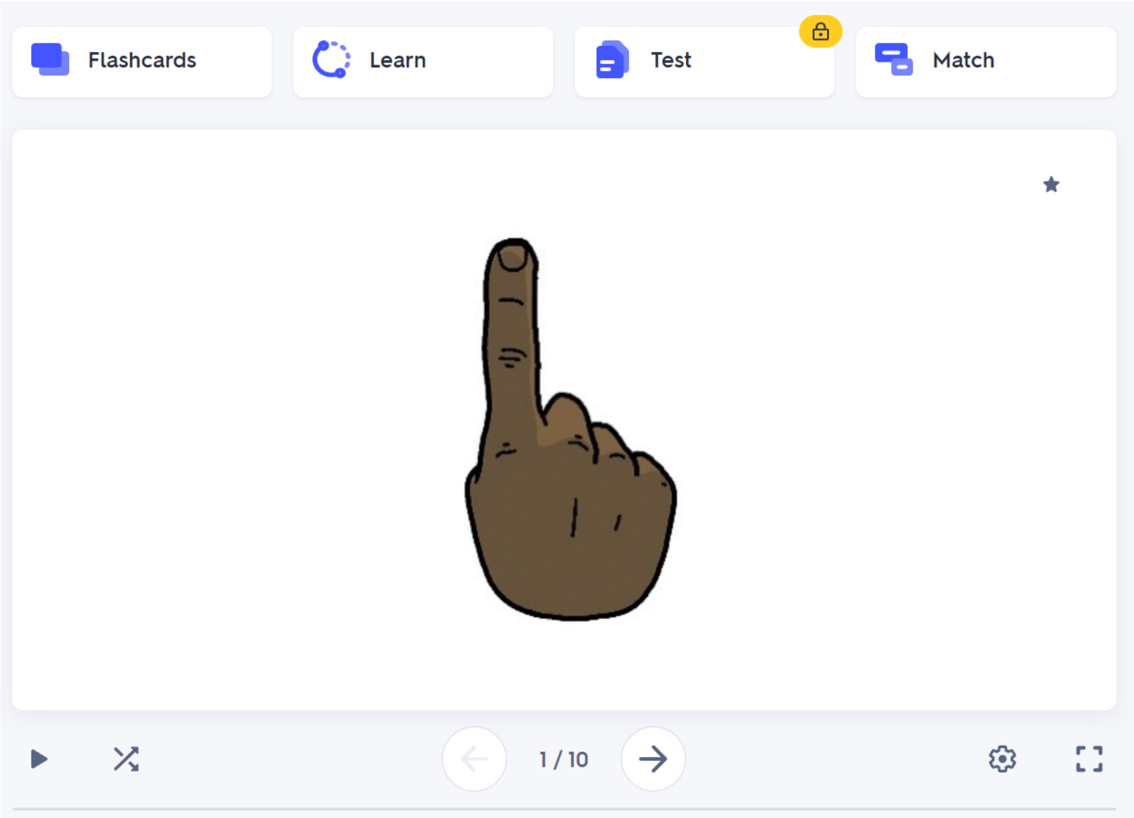

Figure 3.3: The website-based 0-9 ASL learning environment.

3. Users can engage in a game of Whack-a-Mole.

Sign language dictionary

We created an ASL dictionary for users to search. Fig. 3.4 shows illustrations of how the user can sign/express the numbers 0 to 9 (to the left of the Instructions).

Quiz

To improve the efficacy of the learning process, we integrated into the system a question-answer module that allows the users to assess their level of competence and, at the same time, exercise their signing skills by responding to a series of questions generated at random from a data bank. In the example shown in Figure. 3.5, the system pulled from the data bank the question "Can you sign for 9?". After perhaps consulting the dictionary, and if they have developed the appropriate skill, the user can sign the word "9". Alternatively, they can select the "I don't know" option, and the system will demonstrate to them the appropriate expression. In that case,

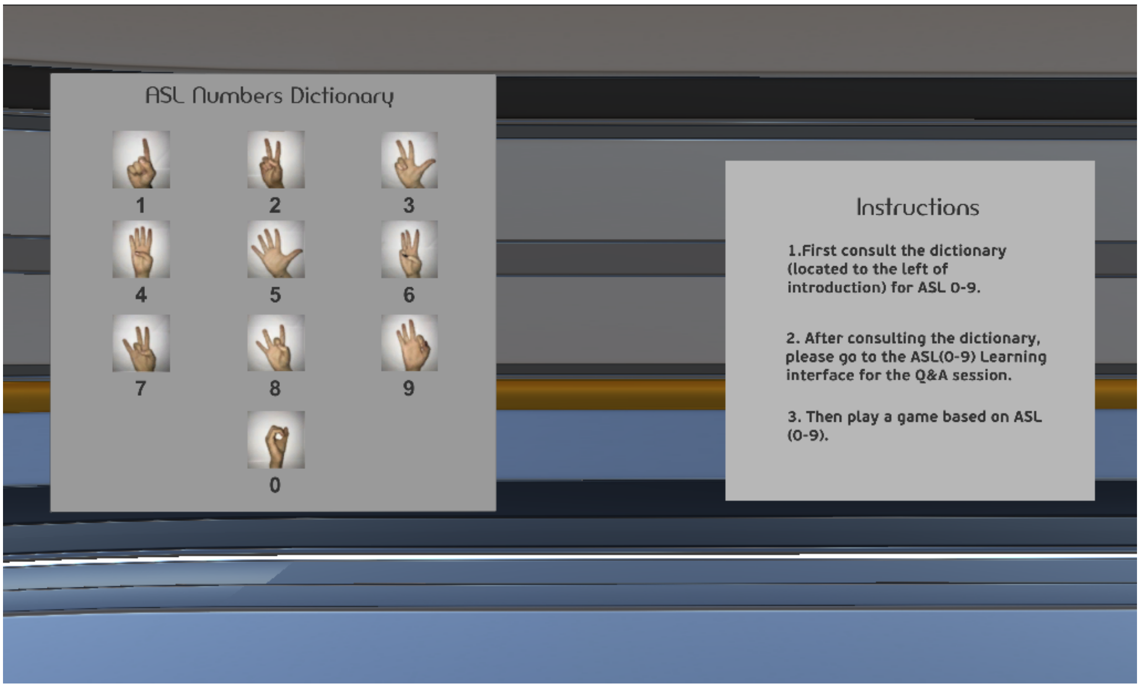

Figure 3.4: The user's initial perspective of 0-9 ASL **Dictionary** and **Instructions**.

the user will also be advised to continue exercising until they indicate that they feel comfortable signing that digit by pressing the relevant button.

Whack-a-Mole game

We adopted the Whack-a-Mole game and implemented a sign language-based version of it, aiming to engage users and improve their learning experience. In our game, as shown in Figure 3.6, each location is marked by a unique identifier. If the user signs correctly the current position of the gopher, one point is added; otherwise, no point is awarded. By default, if the user does not sign the gopher's location within 3 seconds, a new gopher will appear. The total duration of the game is 30 seconds.

3.3.2 Website Environment

Fig. 3.3 shows the user interface of learning 0-9 ASL in a website environment, which includes **Flashcards**, **Learn** and **Match**.

The **Flashcards** corresponds to the user interface of the sign language dictionary in the VR environment so that users can quickly become familiar with the representation of 0–9 numbers. Besides, the **Learn** module, like the Quiz interface, is

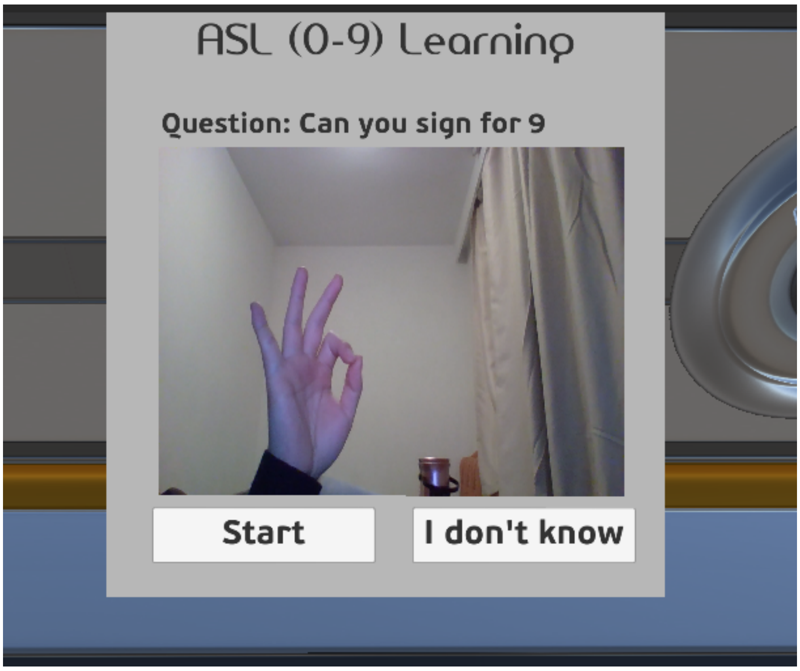

Figure 3.5: The 0-9 ASL learning quiz module in a VR environment.

used to further enhance the user's familiarity with the 0-9 ASL as shown in Fig. 3.7. Moreover, the **Match**, module as shown in Fig. 3.8, like the Whack-a-Mole game, is a match game used to enhance the entertainment of learning 0-9 in ASL.

3.4 Methodology

With the intention of collecting user feedback, which would serve as the study's data source, we invited 15 users (M = 8, F = 7) to test the immersive sign language learning environment we developed, and 15 different users (M = 8, F = 7) to test the website-based ASL learning environment. All participants had no formal training in ASL or any other sign language. A few participants reported casual exposure—for instance, having seen one or two signs in videos or during informal encounters—which defined as 'minimal' knowledge.

Users of the immersive environment first read the Instructions interface to become familiar with the learning procedure and then moved to the 0-9 ASL dictionary to spend 5 minutes studying the representations of the 0-9 digits. Users were encouraged to use the Quiz interface and take quizzes after they felt comfortable with them, and to use the Whack-a-Mole interface to play the game when they consid-

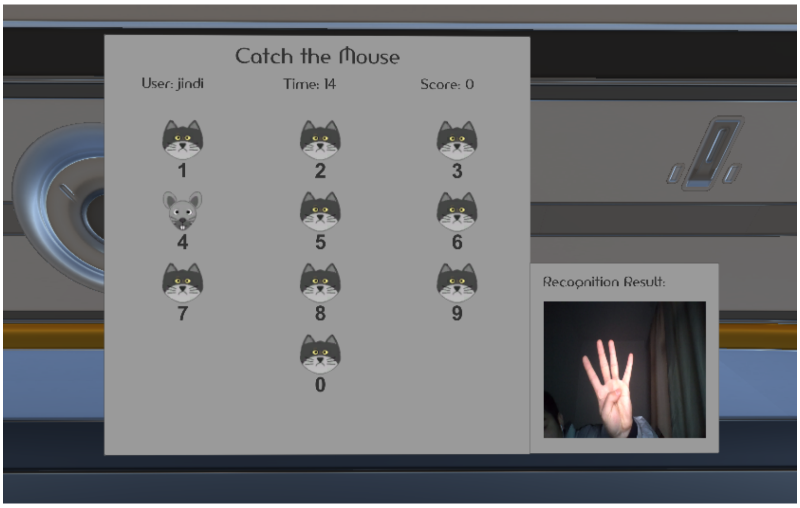

Figure 3.6: The 0-9 ASL learning game in a VR environment.

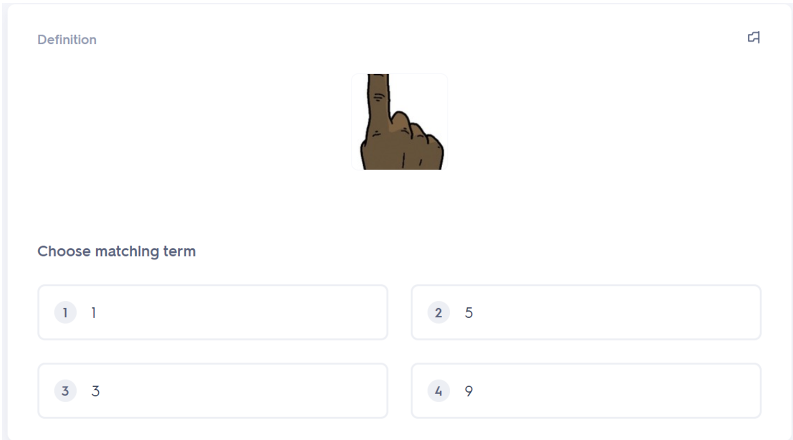

Figure 3.7: The 0-9 ASL learning quiz module in a website environment.

ered themselves familiar with them. In the website-based environment, users first queried the **Flashcards** for 5 minutes to get familiar with 0-9 in ASL, then they were encouraged to click the "Learn" button to take quizzes that contain multiple choice questions, and finally to click the "Match" button to enter the match game.

We adopted the user survey scheme proposed by Schrepp *et al.* [58], which is commonly used for evaluating the user experience in interactive systems, comprising six *scales*, each one representing a distinct user experience aspect: **Attractiveness**, **Efficiency**, **Perspicuity**, **Dependability**, **Stimulation**, **Novelty**. Each scale is divided into either six or four more specific items, as shown in Table 3.1. Following the recommended protocol, we evaluated the user experience, on each of the 26 items, on a 7-point Likert scale ranging from -3 (fully agree with a negative term)

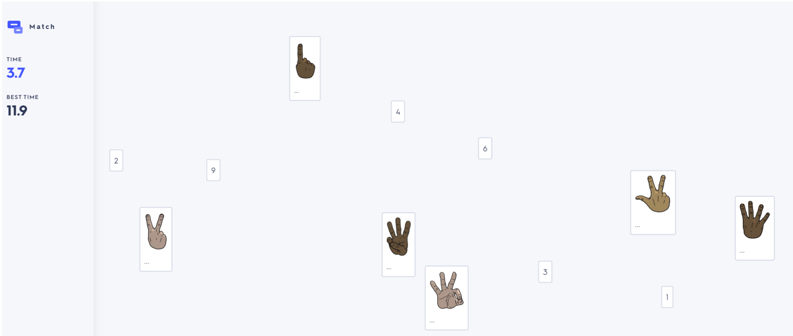

Figure 3.8: The 0-9 ASL learning game in a website environment.

to +3 (fully agree with a positive term). After the completion of the questionnaire, qualitative feedback was collected through follow-up interviews with the users.

Table 3.1: User experience questionnaire.

Attractiveness	**Perspicuity**
A1: annoying / enjoyable	**P1**: not understandable / understandable
A2: good / bad	**P2**: easy to learn / difficult to learn
A3: unlikable / pleasing	**P3**: complicated / easy
A4: unpleasant / pleasant	**P4**: clear / confusing
A5: attractive / unattractive	
A6: friendly / unfriendly	
Efficiency	**Dependability**
E1: fast / slow	**D1**: unpredictable / predictable
E2: inefficient / efficient	**D2**: obstructive / supportive
E3: impractical / practical	**D3**: secure / not secure
E4: organized / cluttered	**D4**: meets expectations / does not meet expectations
Stimulation	**Novelty**
S1: valuable / inferior	**N1**: creative / dull
S2: boring / exciting	**N2**: inventive / conventional
S3: not interesting / interesting	**N3**: usual / leading edge
S4: motivating / demotivating	**N4**: conservative / innovative

3.5 Result Analysis

In this section, we present the analysis of the data collected from the two groups, aiming at a comparison of user experience between the VR and the Web-based environments. Figure 3.9 and Figure 3.10 show the distribution of scores for each question in the two learning environments.

Figure 3.9 and Figure 3.10 show mean value scores for each item in the two learning environments. Clearly, the distribution of the mean scores of VR is more

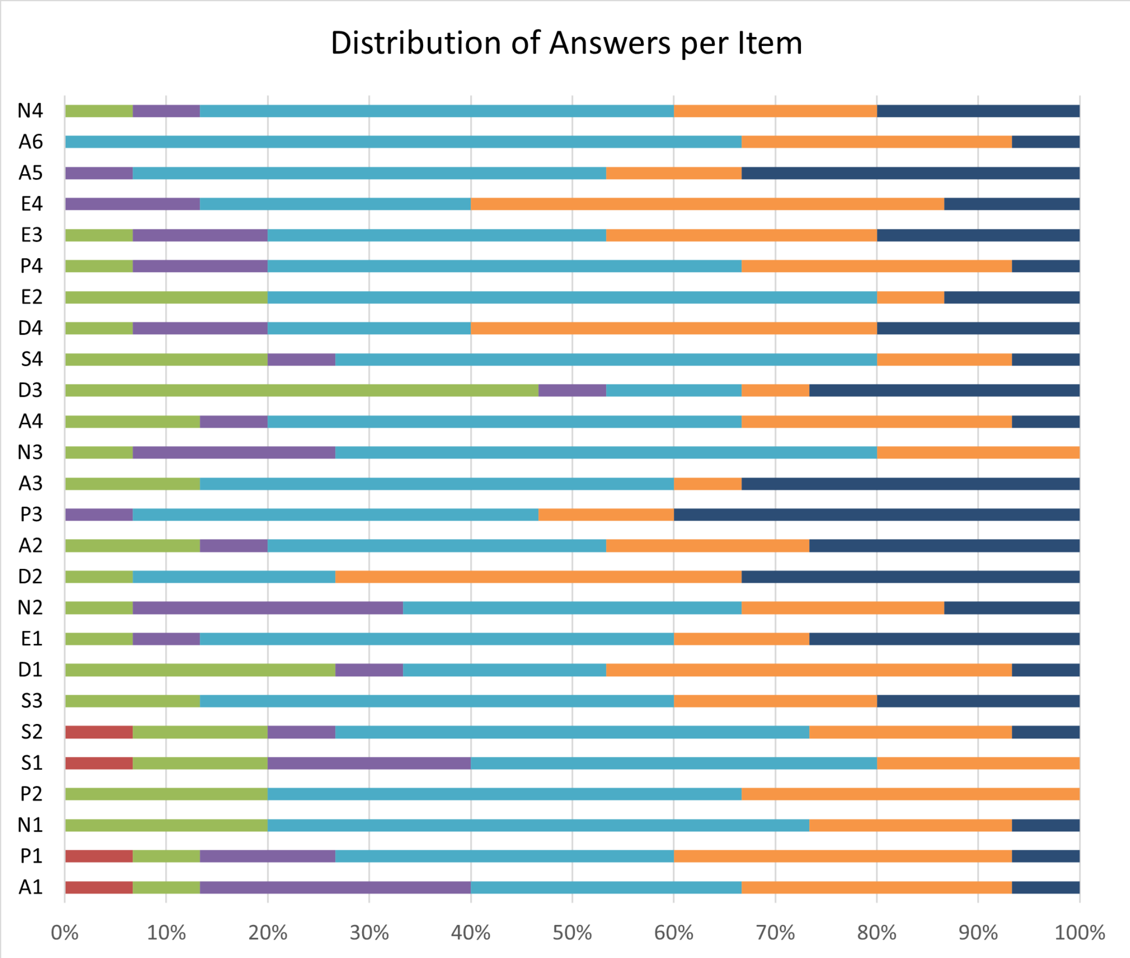

Figure 3.9: The score distributions for the VR environments.

favourable than that of the web-based environment, indicating higher user satisfaction when using the former. We also notice that all mean scores for the VR are positive, while those for the Web are mixed. From qualitative user feedback, we found that the items with negative mean scores are mostly because users felt bored during the Web learning sessions.

To further compare the user experience in the two learning environments, we studied user feedback against the benchmark proposed in [78]. In that paper, the authors analysed a large database of questionnaire responses and proposed the benchmark intervals in Table 3.2.

- **Excellent**: In the range of the 10% best results.
- **Good**: 10% of results better, 75% of results worse.

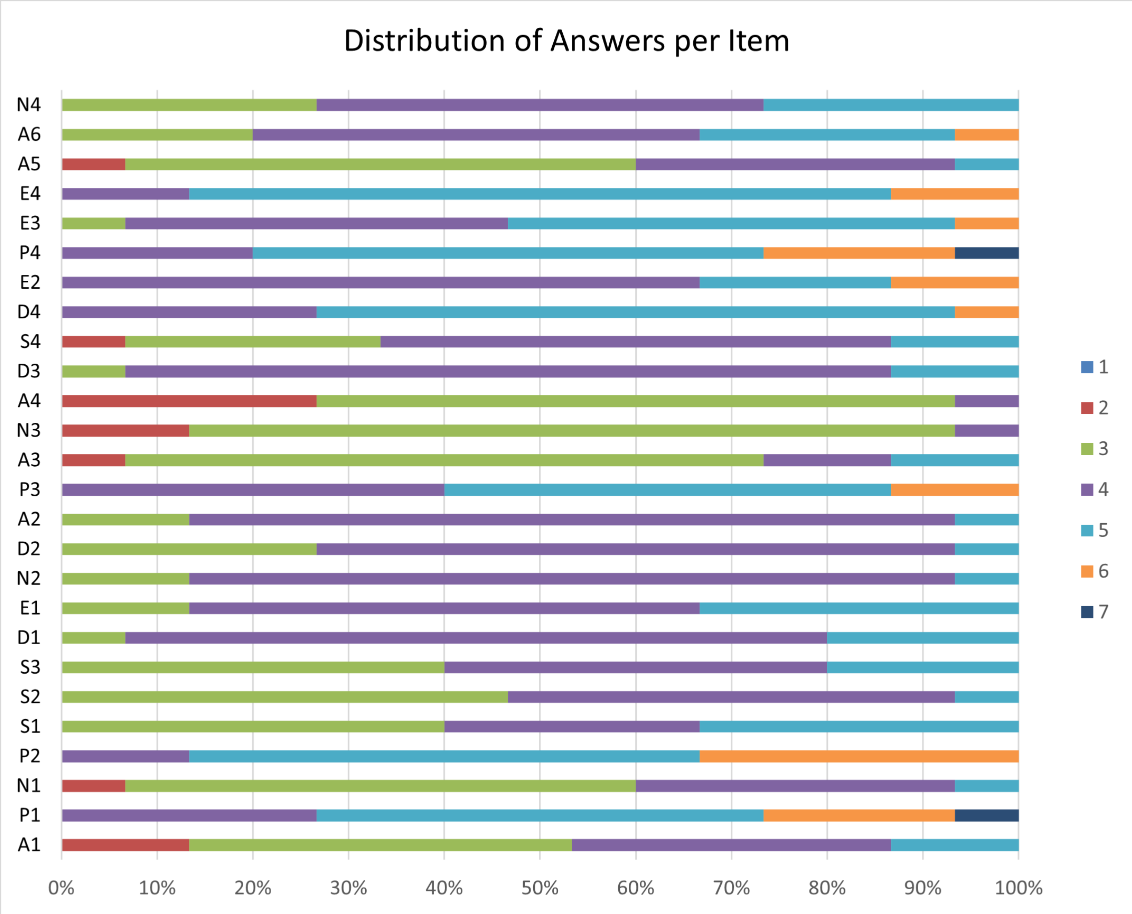

Figure 3.10: The score distributions for the Web environments.

- **Above average**: 25% of results better, 50% of results worse.
- **Below average**: 50% of results better, 25% of results worse.
- **Bad**: In the range of the 25% worst results.

Figure 3.13 and Figure 3.14 depict the distribution of six scales in Benchmark intervals, in which different colours represent different benchmark intervals. We can observe that the mean values of six scales except **Stimulation** in the VR environment are above average; this indicates that a large number of users can accept learning ASL 0-9 in VR, while in the Web environment, the mean values of six scales are almost below the average, indicating that users have a low acceptance of learning ASL 0-9 on website. The six scales in the two learning environments are analysed below.

Attractiveness: the average score in VR is 1.311 (SD = 0.791) in the "above

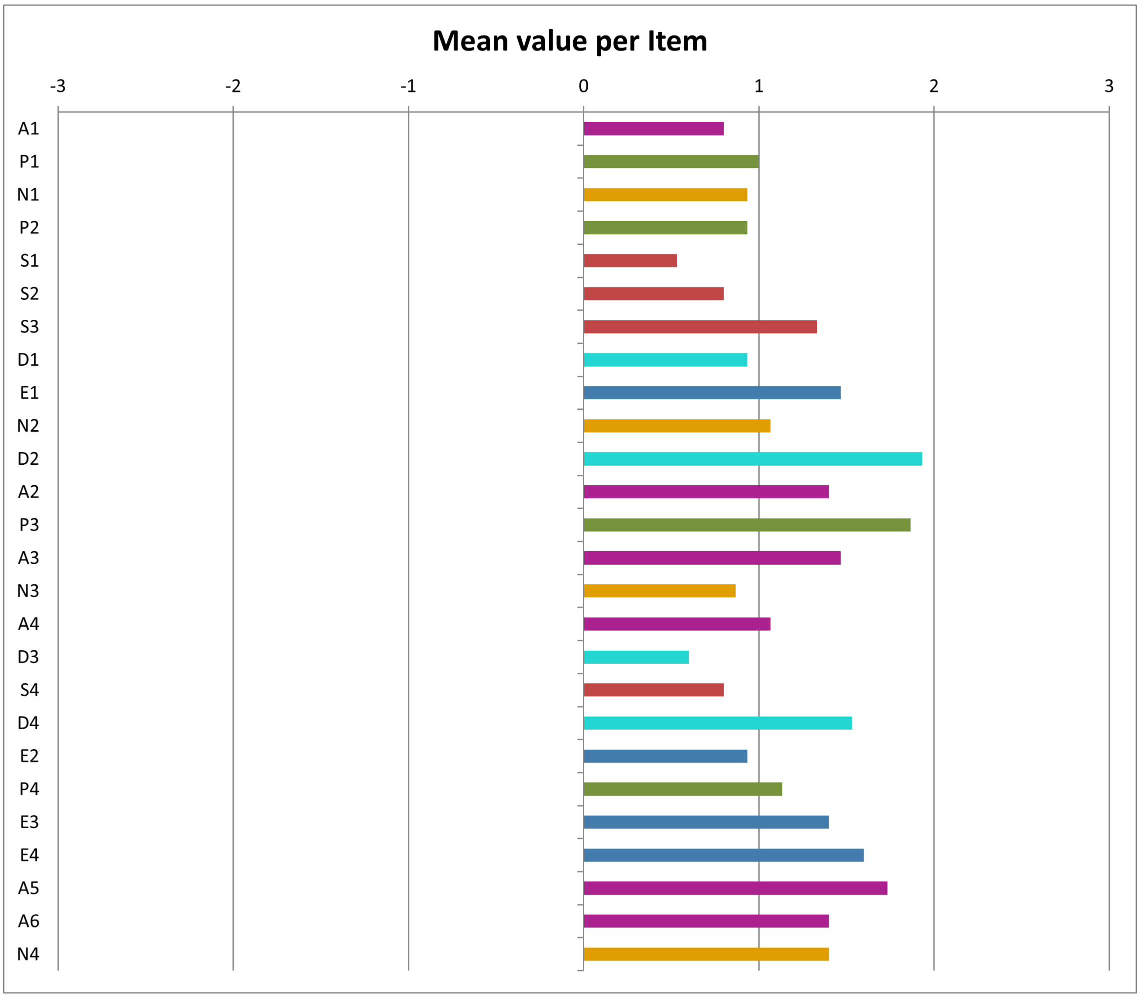

Figure 3.11: Mean value scores for each item in VR environments.

average" category, while for Web it is -0.478 (SD = 0.511) in the "Bad" category. This suggests that users find the VR environment more attractive for them to learn sign language. However, it received low scores from several users in terms of "annoying/enjoyable", as shown in Figure 3.9. This could be due to the fact that the user needs to wear a headset which increases their learning burden. Besides, for the web-based environment, it received low scores in terms of "unpleasant/pleasant", as shown in Fig. 3.10. This may be because the content they were learning was not interesting enough.

Perspicuity: the average score for VR is 1.233 (SD = 0.848) in the "above average" category, while for Web it is 1.033 (SD = 0.611) in the "Below average" category. The two platforms have similar scores on this scale and are both around the average, indicating that the perspicuity of both platforms is well-received by

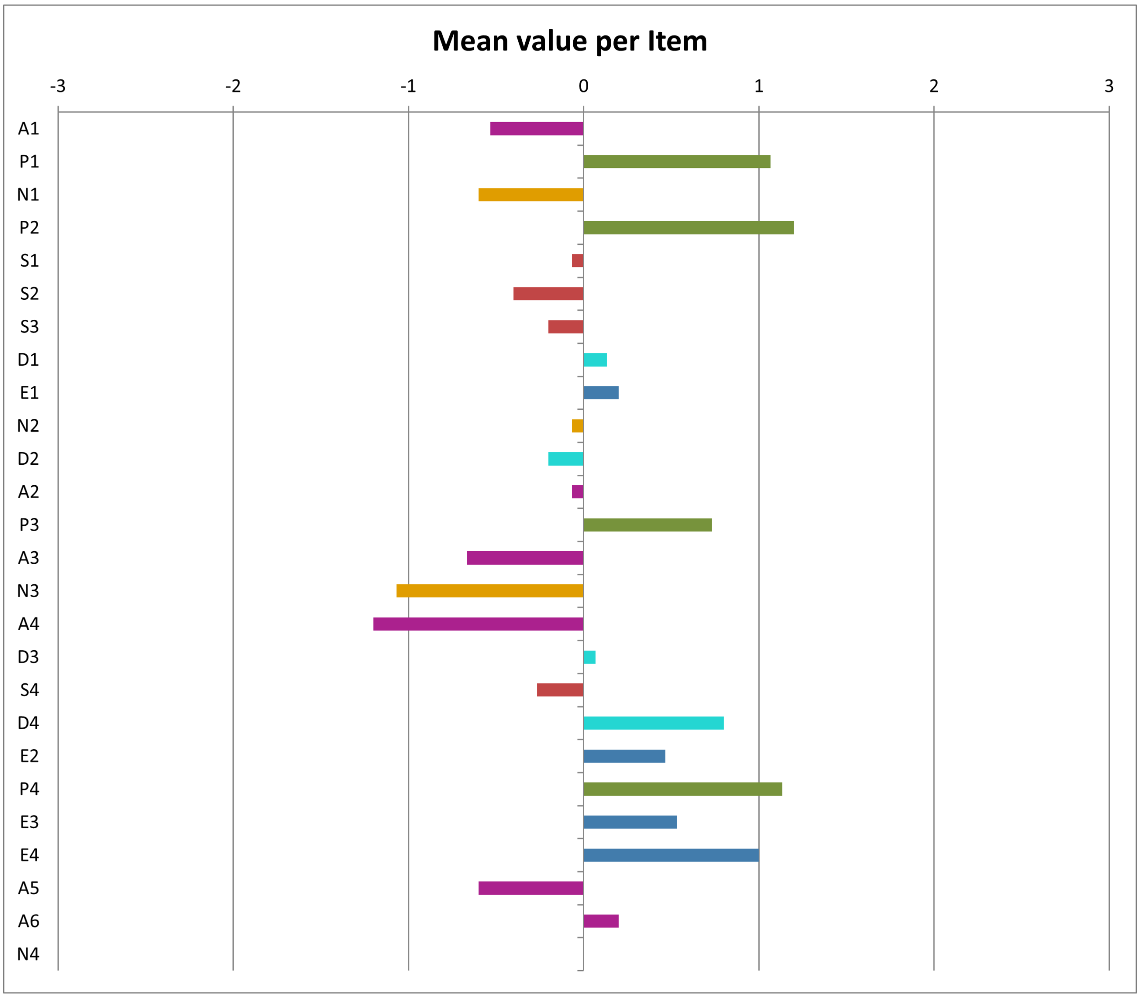

Figure 3.12: Mean value scores for each item in the Web environments.

users.

Efficiency: the average score for VR is 1.350 (SD = 0.915) in the "above average" category, while for Web it is 0.550 (SD = 0.536) in the "Bad" category. Although the score of VR is higher than that of the Web, the standard deviation of VR is also larger than that of the Web, indicating that VR fluctuates greatly under the efficiency scale. The possible reason is that some users have used VR equipment before and did not need to re-learn the use of VR equipment; while some users had no experience in using VR before and needed to learn the use of VR, which might increase their learning load.

Dependability: the average score in VR is 1.250 (SD = 0.829) in the "above average" category, while for Web it is 0.200 (SD = 0.316) in the "Bad" category. Although the standard deviation of VR on this scale is large, it means that some

Table 3.2: Benchmark intervals for the user experience scales.

	Attractiveness	Perspicuity	Efficiency	Dependability	Stimulation	Novelty
Excellent	≥ 1.75	≥ 1.78	≥ 1.90	≥ 1.65	≥ 1.55	≥ 1.40
Good	[1.52, 1.75)	[1.47, 1.78)	[1.56, 1.90)	[1.48, 1.65)	[1.31, 1.55)	[1.05, 1.40)
Above average	[1.17, 1.52)	[0.98, 1.47)	[1.08, 1.56)	[1.14, 1.48)	[0.99, 1.31)	[0.71, 1.05)
Below average	[0.70, 1.17)	[0.54, 0.98)	[0.64, 1.08)	[0.78, 1.14)	[0.50, 0.99)	[0.30, 0.71)
Bad	<0.70	<0.54	<0.64	<0.78	<0.50	<0.30

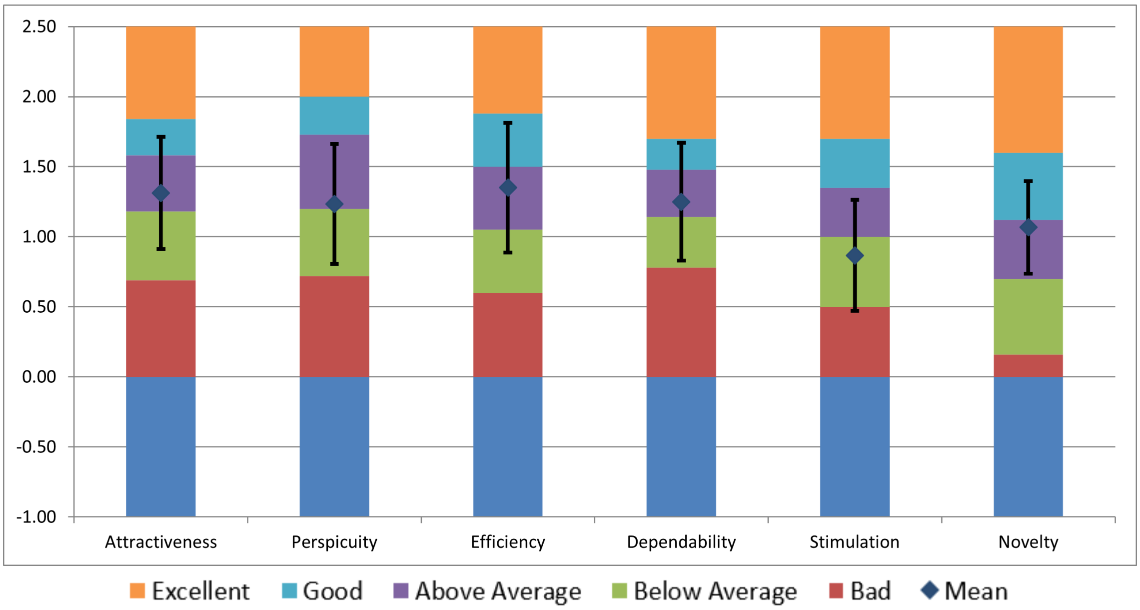

Figure 3.13: Benchmark intervals for the six scales in VR.

users feel that VR needs to be strengthened in terms of Dependability. Several users thought they executed the right gesture but were judged as being wrong. The back-end gesture recognition algorithm may be underperforming due to inadequate training data for particular gestures, resulting in low recognition accuracy rates. Future developments are planned to train recognition models to better generalize to user hand physiological variances.

Stimulation: the average score in VR is 0.867 (SD = 0.784) in the “Below average” category, while for Web it is -0.233 (SD = 0.563) in the “Bad” category. The score of VR on this scale is higher than that of the Web, because according to the feedback of users, learning sign language on the Web was only a selection with a mouse click, while in VR, sign language could be used to express numbers in an immersive way, which is more attractive to them.

Novelty: the average score in VR is 1.067 (SD = 0.651) in the “above average” category, while for Web it is -0.433 (SD = 0.306) in the “Bad” category. Today’s

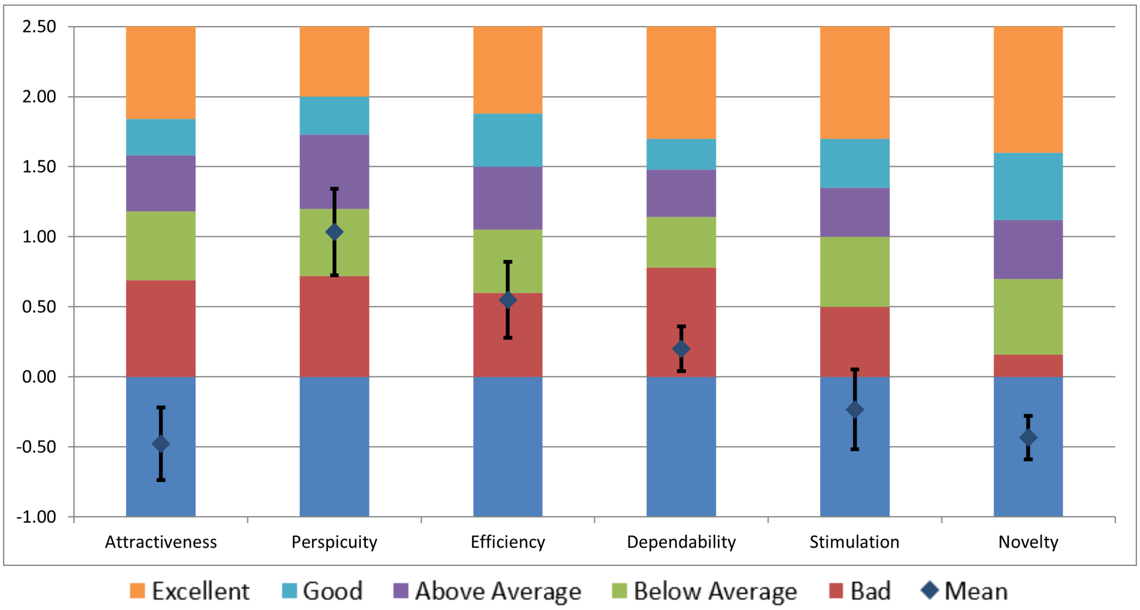

Figure 3.14: Benchmark intervals for the six scales in the Web.

Table 3.3: Means and p-values for the three groups of scales.

Pragmatic and Hedonic Quality	**VR**	**Website**	**P-Value**
Attractiveness	1.31	-0.48	5.73216E-20
Pragmatic Quality	1.28	0.59	1.86984E-09
Hedonic Quality	0.97	-0.33	1.23737E-18

users are acclimated to searching and learning content on the web, hence the Web's score on this scale is low. VR scores higher than the Web because many people have never learned in VR and find it novel.

In [78], the scales of the user experience questionnaire are grouped into *pragmatic quality* (Perspicuity, Efficiency, Dependability), and *hedonic quality* (Stimulation, Novelty), while Attractiveness is a pure valence dimension. Pragmatic quality describes task-related quality aspects, and hedonic quality describes the non-task-related quality aspects. Table 3.3 shows the mean scores for the two environments over these grouped scales, and the corresponding p-values. We note that, in all cases, we have statistically significant differences, especially regarding attractiveness and hedonic quality.

3.6 Discussion

We used VR technology to create an immersive environment for learning numbers in ASL and conducted a user study on it. We investigated whether immersive environments can provide a better user experience. Our hypothesis that users prefer to learn ASL 0-9 in an immersive, rather than a website-based environment, was tested with a questionnaire-based study on two groups of users.

We employed six assessment scales — **Attractiveness**, **Efficiency**, **Perspicuity**, **Dependability**, **Stimulation**, **Novelty** — on which users judged the system's design. The results show that our immersive ASL learning system performs effectively and can essentially meet user requirements for ASL learning. Thus, we were able to answer affirmatively the research question of whether an immersive ASL learning environment provides a better user experience to learn ASL. The analysis of environmental experience reveals that most users prefer to learn sign language in an immersive VR environment, probably because it is more experiential, while only very few users prefer to learn ASL via the web-based learning environment, primarily because they find it convenient and user-friendly to search for sign language information on web pages.

We note that, even though most users prefer immersive VR learning environments, the technology is still in its early stages of development. The system needs to be further streamlined and tweaked for specific functions to enhance the user experience. In particular, qualitative user feedback revealed the following limitations of the immersive VR learning environment.

3.6.1 Limitations

One important limitation of this study is the potential influence of the novelty effect associated with VR. Users were introduced to the immersive learning environment for the first time, and their high satisfaction ratings may partly reflect the novelty of the experience rather than the long-term usability or educational effectiveness of the system. Future longitudinal studies involving repeated use over multiple sessions are necessary to determine whether user engagement and satisfaction persist once the novelty wears off.

For each assessment scale, the main limitations of the system's design and implementation were as follows. **Attractiveness**: some users complained that there weren't any animated hints when sign language was correctly deciphered. The user's experience could be improved by including animated hints. **Efficiency**: some users have complained that our interface is not sufficiently automatic because playing the sign language game requires them to press the start button actively. **Perspicuity**: some users have reported that they did not know how to move around in the scene. **Dependability**: some users reported that they had made the correct sign language while playing the game, but the system judged that it was wrong, resulting in a lower score for the user. **Stimulation**: from the analysis of the stimulation scale results, a small number of users gave low scores, indicating that our system needs to be designed more creatively. **Novelty**: a small number of users thought that the system was not innovative enough, perhaps because they perceived the learning model as too easy.

Additionally, there were several methodological issues with our study that need to be considered. Firstly, with only 30 people taking part in the user survey, the sample size is relatively small and may not give an accurate representation of the intended audience. Secondly, as our study included only Chinese participants, the findings might not apply to people from different cultural backgrounds. To ensure broader validity of the findings, future studies should be based on larger and more varied samples.

3.7 Conclusion

We developed a virtual setting that gives users an immersive experience for learning ASL. We created four user interfaces corresponding to four distinct types of functionality, enabling users to easily comprehend the system's workflow and each phase of ASL learning: an instruction module, an ASL dictionary, a quiz module, and an ASL game based on Whack-a-Mole. To determine the acceptability of our UIs and evaluate user satisfaction with the virtual environment design, the findings of a user questionnaire were analysed (N = 30). The results indicated that users were generally satisfied with the virtual environment we built, and they preferred it against the website-based learning mode. Overall, the outcome supports our initial hypothesis that immersive virtual environments can improve users' experience of learning ASL.

In the future, we plan to include in our immersive environment more interactive elements such as backdrop movement, scene changes, and animation prompts. We also plan to implement various automatic settings to minimize the need for user interaction for the control of the system, and a follow-through user interface will be created so that the user can comprehend how to move the items in the scene. Furthermore, a more robust sign recognition model will be created, allowing us to include more sophisticated sign language learning material.

CHAPTER 4

Developing and Evaluating a Novel Gamified Virtual Learning Environment for ASL

In this chapter, we further the investigations begun in Chapter 3 by integrating additional educational content, specifically enabling users to learn the alphabet through an immersive virtual reality (VR) environment. This development represents an enhancement over Chapter 3, which was limited to the exploration of numeric content within similar VR settings. By broadening the scope to include alphabetic learning, we aim to provide a more comprehensive educational tool that caters to a wider range of learning objectives. The expanded research and its implications were presented at the 2023 IFIP Conference on Human-Computer Interaction.

4.1 Introduction

Sign language is a visual language that uses hand gestures and facial expressions to convey meaning. It is primarily used for communication with individuals who are deaf or hard of hearing or who experience difficulty speaking. Learning sign language is important for several reasons. Firstly, it enables better communication and social interaction with the hearing-loss community, thereby promoting inclusion and understanding. By learning sign language, one can break down communication barriers and establish meaningful connections with individuals who might otherwise feel excluded. Secondly, learning sign language has been shown to have numerous cognitive benefits, including enhancing cognitive development and language skills [79]. It is widely acknowledged that learning a second language has cognitive benefits, and the same is true for sign language. Finally, for individuals who experience hearing or speech impairments, sign language can serve as a crucial mode of communication, allowing them to participate more fully in society. Despite the importance of learning sign language, traditional web-based methods of learning have not been able to generate much interest among learners, partly because of a lack of novelty. Therefore, there is a need for more engaging and innovative approaches to learning sign language that can increase user engagement and promote effective learning.

While immersive VR systems have demonstrated promise in improving user experience in ASL learning (as explored in Chapter 3), the role of gamification—particularly in expanding content from numeric to alphabetic learning—remains underexplored. Prior studies in learning sciences and HCI suggest that integrating game mechanics can enhance motivation, increase engagement, and foster more sustained interaction with educational systems [12]. Drawing from this literature, we propose the following research hypothesis:

- **H4.1:** Incorporating game elements into an immersive ASL learning environment (covering both numeric and alphabetic content) will lead to higher user experience ratings in the areas of Stimulation, Novelty, and Engagement than traditional immersive learning environments without gamification.

Our main contributions are as follows:

1. We extend prior VR-based ASL systems by designing a comprehensive learning environment that includes both numeric and alphabetic content and incorporates a gamified module to enhance learner engagement.

2. We evaluate the system through a dedicated user study using the User Experience Questionnaire (UEQ), with results analysed against industry benchmarks to assess how gamification affects user perception and satisfaction.

3. We compare the outcomes to the findings of Chapter 3 to better understand the distinct effects of gamification, separate from the benefits of immersion alone.

4.2 User Interface of VR Environment

This section provides an overview of the main components of our user interface (UI) and highlights the main features of our VR environment. The UI is comprised of four different modules designed to facilitate effective ASL learning.

1. The **Instructions** module, which consists of six basic steps, provides users with an overview of the ASL learning process and guides them through the initial stages of the programme.

2. The **Sign Language Dictionaries** module, which enables users to consult and search for the signs of numbers or letters. This module serves as a reference tool for users as they progress through the learning process.

3. The **Quiz** module, which contains question-answer quizzes that allow users to test their signing skills and self-assess their level of competence. This module serves as a valuable feedback mechanism for users and encourages them to actively engage with the learning material.

4. The **Whack-a-Mole Game** module, which is to increase user motivation and engagement with the learning process. This module presents users with a fun and interactive way to practice their ASL skills, reinforcing their learning and providing a welcome break from more traditional learning methods.

Together, these four modules work in concert to provide users with a comprehensive and engaging VR-based ASL learning experience. By incorporating elements of gamification and interactivity into our VR environment, we hope to improve user satisfaction and facilitate more effective ASL learning outcomes.

We separated the scene of the immersive environment into two parts. Adopting the concept of a simple to complex learning process, the first part is for learning the numeric ASL, something that is considered a relatively easy task. The second part of the scene is for the more challenging task of learning the alphabetic ASL, excluding J and Z, which require dynamic gesturing.

Fig. 4.2 shows the initial view of the user when entering the VR environment, which includes the **Instructions** and **Sign Language Dictionary** interfaces. Fig. 4.2 shows the **Quiz** and **Whack-a-Mole Game** interfaces of numerical ASL learning, which are located to the left of the numerical ASL dictionary. Fig. 4.2 shows the **Quiz** and **Whack-a-Mole Game** interfaces of alphabetic ASL learning, which are located to the right of the alphabetic ASL dictionary.

The scene was developed in Unity 2020.3.32f1, and user interaction was facilitated through eye tracking using HTC Vive Pro. After 3 seconds of fixed attention, users can click or select objects in the scene. An integrated camera was used to acquire images; openCV (version 3.4.2) [77] was used for image processing on a PC. Hand gestures were detected using Mediapipe, which extracted a feature vector of 21 points corresponding to landmarks on the detected hand. An multilayer perceptron (MLP) consisting of 3 fully connected layers was implemented in Python 3.6 [80] and Tensorflow 2.6.0 [81] for gesture recognition. The classifier was trained on a standard PC with an RTX3080 GPU, achieving recognition accuracy rates above 90%, deemed sufficient to ensure a smooth user experience in our study.

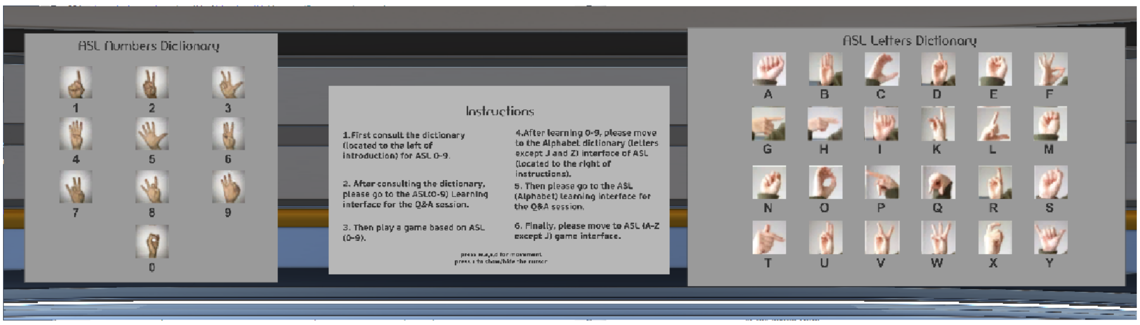

Figure 4.1: The implemented immersive virtual environment. (a) **Left:** the numeric ASL sign language dictionary. **Centre:** Instructions interface. **Right:** the A-Y except for J sign language dictionary.

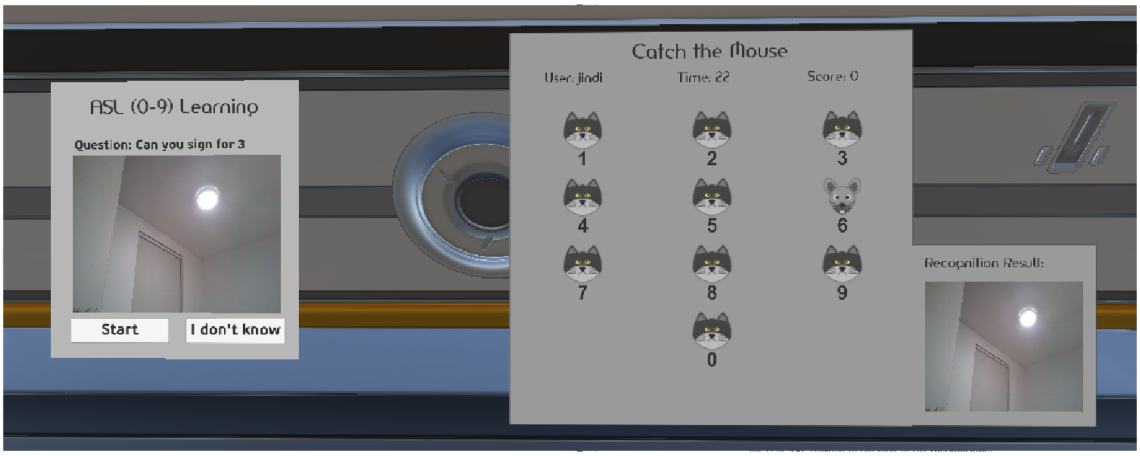

Figure 4.2: The implemented immersive virtual environment of the numeric ASL learning quiz (left) and game (right).

Figure 4.3: The implemented immersive virtual environment of the alphabetic ASL learning quiz (left) and game (right).

4.3 User Study Design

In order to evaluate the immersive environment design, we adopted the user survey scheme proposed by Schrepp *et al.* [58], which is commonly used to evaluate user experience in human-computer interaction systems. It consists of six evaluation factors, called *scales*: **Attractiveness**, **Efficiency**, **Perspicuity**, **Dependability**, **Stimulation**, **Novelty**. Each scale is further divided into four or six *items*, as shown in Table 3.1. We evaluated the proposed VR environment, on all scales and items, on a 7-point Likert scale ranging from -3 (fully agree with a negative term) to +3 (fully agree with a positive term), and studied the user feedback against the benchmark proposed in [78]. In that paper, the authors analysed a large database of questionnaire responses and derived the benchmark intervals shown in Table 3.2. These intervals correspond to the distribution:

- **Excellent**: In the range of the 10% best results.
- **Good**: 10% of results better, 75% of results worse.
- **Above average**: 25% of results better, 50% of results worse.
- **Below average**: 50% of results better, 25% of results worse.
- **Bad**: In the range of the 25% worst results.

We conducted the user study obtaining feedback from 15 participants, 8 males and 7 females, aged between 19 and 21 years old, who had little or no prior experience with ASL or any other sign language. None of the participants involved in this study had previously participated in the user study described in Chapter 3; the two groups were completely independent. At the start of the session, participants had the freedom to explore the system and consult the Instructions module. Then, each participant followed a six stages learning process:

1. Learn numeric ASL for 3 minutes from corresponding dictionary module.
2. Improve numeric ASL comprehension for 3 minutes in numeric quiz module.
3. 30 seconds on numeric ASL game module.
4. Learn alphabetic ASL from corresponding dictionary module for 3 minutes.
5. Improve alphabetic ASL literacy for 3 minutes in alphabetic quiz module.
6. 30 seconds on alphabetic ASL game module.

4.4 Result Analysis

Figure 4.4 shows the average scores for the six scales, denoted by 'x', plotted over a colour code of the corresponding benchmark interval. For each scale, the minimum and the maximum of the average scores on its individual items are also shown. In the figures of each item, the box plots show the minimum, first quartile, median, third quartile, and maximum, for each individual item of each scale.

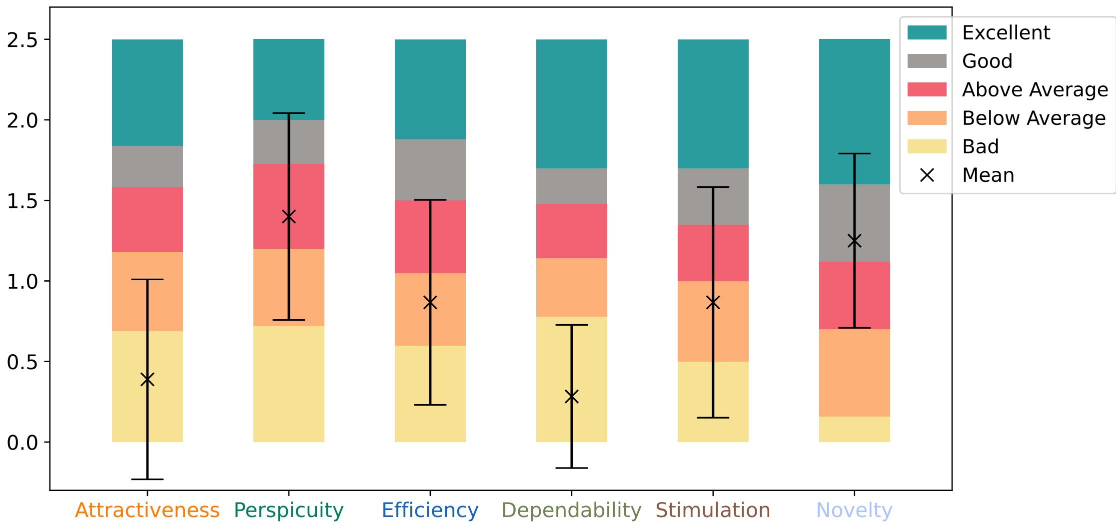

Figure 4.4: Benchmark intervals for the six scales

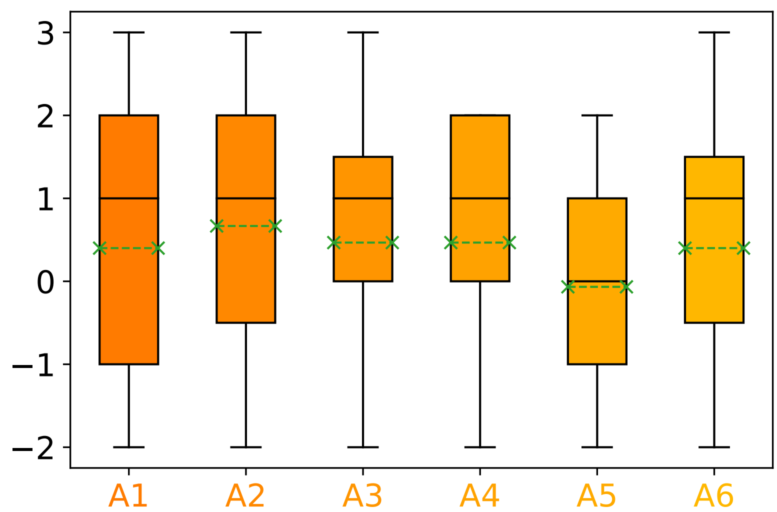

Figure 4.5: Box-plots of the scores for each item of the attractiveness.

Attractiveness: The mean value of the user scores is 0.39 (SD = 1.24), placing it in the "Bad" category, indicating that their overall impression of the VR environment was not favourable, and the system requires further improvements. Notably, the average score for item **A5**, shown in Fig. 4.4, is slightly below 0, which suggests that the users did not find the system particularly appealing. This may be because

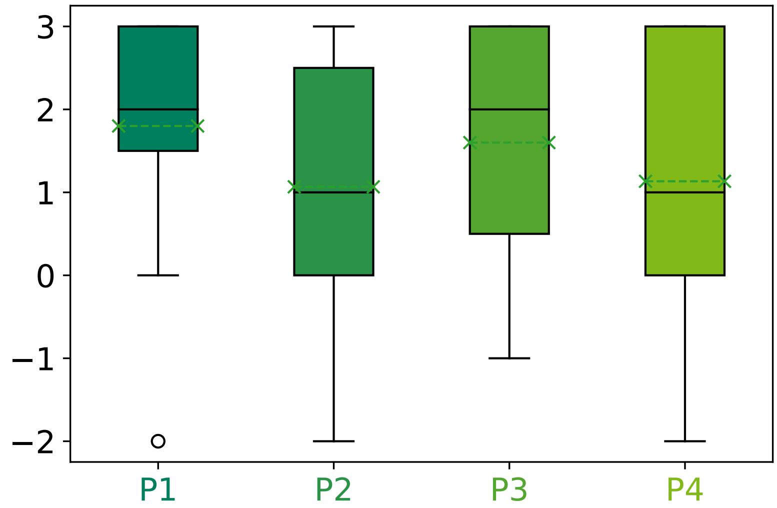

Figure 4.6: Box-plots of the scores for each item of the perspicuity.

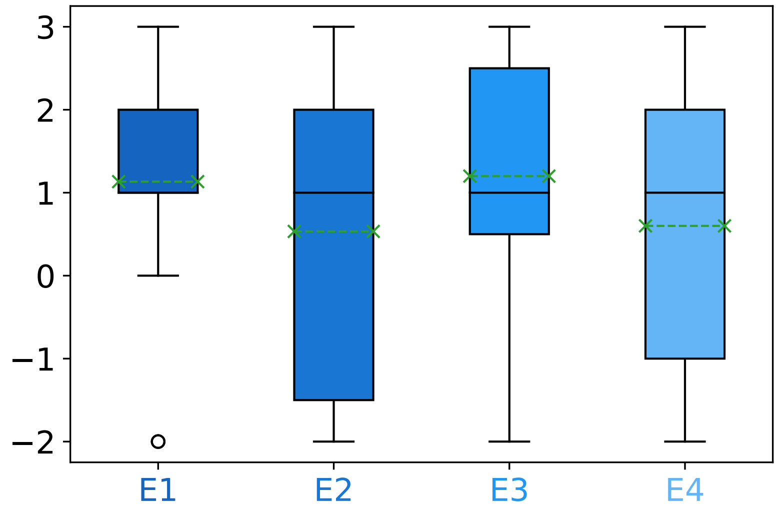

Figure 4.7: Box-plots of the scores for each item of the efficiency.

the learning environment relies on 2D user interfaces, whereas incorporating 3D elements may be more visually engaging for users. Therefore, we plan to integrate 3D user interfaces in future iterations of the ASL learning environment, aiming at enhancing its attractiveness.

Perspicuity: The average score is 1.40 (SD = 1.28), placing it in the "Above average" category, indicating that users perceive the VR environment as clear and understandable, facilitating their ASL learning experience. However, it seems that some of the users may have encountered some problems when using the environment, possibly due to their unfamiliarity with VR devices, and they may require some

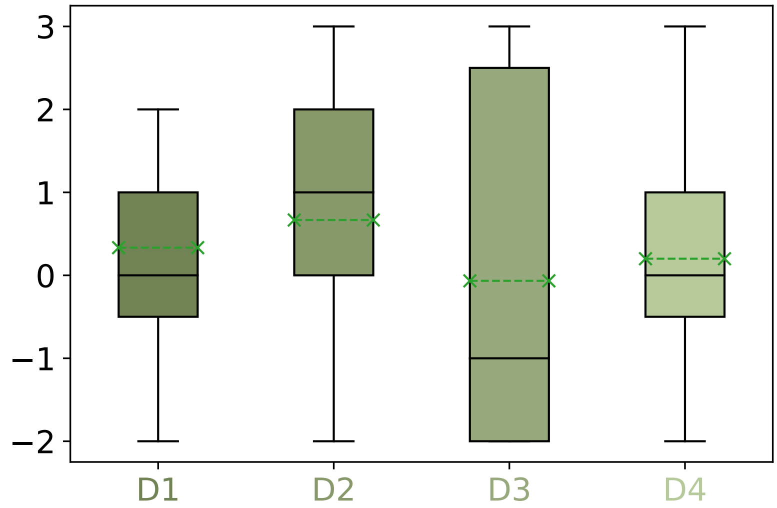

Figure 4.8: Box-plots of the scores for each item of the dependability.

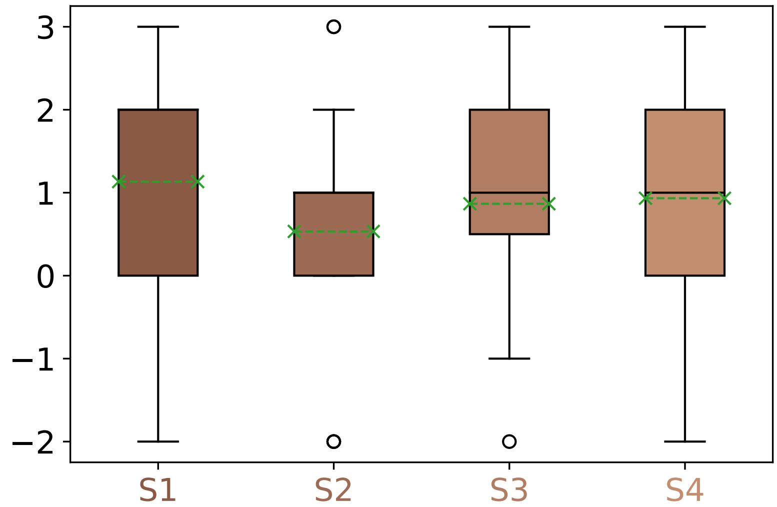

Figure 4.9: Box-plots of the scores for each item of the stimulation.

initial training.

Efficiency: In the "Below average" category, the average score is 0.87 (SD = 1.27). We note that, while the average score over the whole scale is slightly below average, analysis of individual item scores shows that our VR environment adequately fulfills some users' requirements. In particular, users found the system easy to use (as reflected by item **E1**) and believed that they could practice ASL effectively in the scenario (as reflected by item **E3**), see Fig. 4.4.

Dependability: In the "Bad" category, the average score is 0.28 (SD = 0.89). That means that the VR environment's dependability needs significant improve-

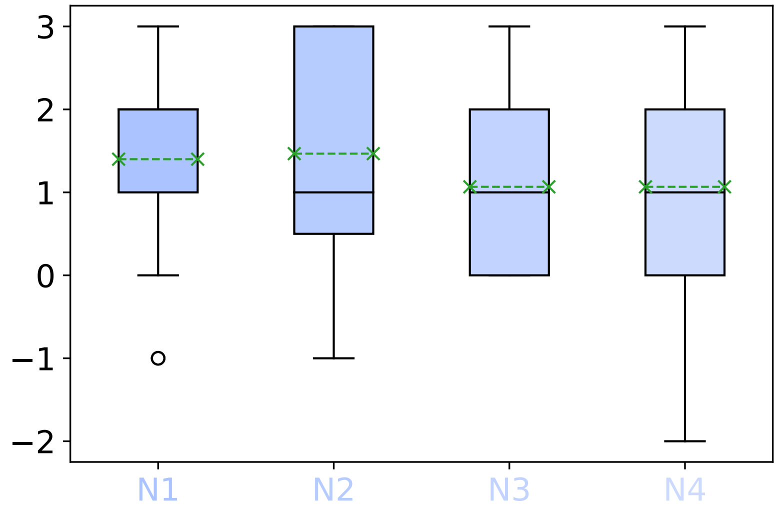

Figure 4.10: Box-plots of the scores for each item of the novelty.

ment. Despite the low overall average score, some users still believed that on individual items, particularly **D2** and **D4**, the system adequately fulfilled their requirements, see Fig. 4.4.

Stimulation: In the "Below average" category with an average score of 0.87 (SD = 1.43). Even though the score is slightly lower than average, the large variance indicates that some users find the learning environment stimulating. As shown in Fig. 4.4, the first quartile of all items is non-negative, indicating that a majority of users have a consistently favourable outlook regarding this scale.

Novelty: In the "Good" category with an average value of 1.25 (SD = 1.08). Again, the first quartile of all items is non-negative, see Fig. 4.4, indicating a consistently favourable view from a majority of users. They perceive the VR environment as a novel and innovative way of learning ASL.

4.5 Discussion

This chapter presented the design and evaluation of a gamified immersive VR learning environment for American Sign Language (ASL), expanding the scope of Chapter 3 from numeric-only to alphabetic ASL content. The study aimed to assess user experience through the UEQ framework across six dimensions: Attractiveness, Perspicuity, Efficiency, Dependability, Stimulation, and Novelty.

The findings revealed that while the VR environment was perceived as novel ("Good") and relatively clear to understand ("Above average" for Perspicuity), other aspects such as Attractiveness and Dependability received lower ratings ("Bad"). This indicates that although the gamified system successfully engaged users with its innovative presentation and learning format, several design and interaction elements still need refinement.

The high score in Novelty reflects users' appreciation for the fresh, immersive format of the learning experience—particularly the inclusion of the Whack-a-Mole-style game. Similarly, moderate ratings in Stimulation and Efficiency suggest that interactive elements positively affected motivation and learning flow, although not consistently across all users. Feedback pointed to areas such as system responsiveness, gesture recognition reliability, and visual engagement where improvements are necessary. The low score in Attractiveness highlights a potential disconnect between visual design and user expectations, potentially due to the reliance on flat 2D UI components within a 3D immersive space.

Compared with Chapter 3, which showed stronger scores across all UEQ dimensions when contrasting immersive and web-based environments, the results in this chapter suggest that simply adding gamification is not sufficient to guarantee high user satisfaction—particularly when the novelty wears off or when the gameplay does not integrate deeply with learning objectives. This chapter thus refines our understanding of what makes VR-based ASL learning effective, emphasizing the importance of cohesion between game mechanics and educational goals.

These findings confirm the importance of thoughtful gamification design in educational VR environments. They also support our original hypothesis that users would find the gamified VR approach engaging, though with the caveat that user experience depends on more than novelty and gameplay alone—it must also include usability, trust in system feedback, and aesthetic quality. Future work should investigate these issues in greater depth and through longitudinal studies to understand how engagement and learning outcomes evolve over time.

4.6 Conclusion

We have developed a VR system for learning numeric and alphabetic ASL and conducted a questionnaire-based user study to evaluate the user experience of learning ASL in the system. We found that to some extent it satisfied some user satisfaction factors, however, the system needs further development to enhance user experience, especially on the factors of attractiveness and dependability.

There are several limitations to our ASL learning system, which have been discussed for each scale of user experience separately. The identified shortcomings include a lack of animated hints; an interface that requires users to actively press a start button to commence an action; difficulty in moving around the VR scene; a relatively large number of incorrect judgments of correct signs, i.e., many false negatives; user expectations for a more creatively designed system; and an overall perception that the learning task was too easy. Additionally, the user study included 15 only participants, primarily between the ages of 19 and 21, and there was a complete lack of research on users in other age groups.

To address these limitations, we plan to revise the content, design, and implementation of the system as follows: add more interactive elements; implement automatic settings; create a follow-through user interface; develop a more robust sign recognition model; and include more sophisticated sign language learning material. We also plan to recruit a larger and more diverse group of participants for a follow-up user study.

CHAPTER 5

User-Defined Hand Gesture Interface to Improve User Experience of Learning American Sign Language

In this chapter, we introduce a novel user interface that allows users to create their own sign language gestures to engage with a game designed for learning American Sign Language (ASL). This innovative approach not only enhances interactivity but also personalizes the learning experience, catering to individual preferences and learning styles. The research findings from this chapter were showcased at the 2023 International Conference on Intelligent Tutoring Systems.

5.1 Introduction

According to the World Health Organization, around 2.5 billion people will have some degree of hearing loss by 2050[1], and at least 700 million of them will require some kind of hearing rehabilitation. The use of sign language, as well as several other alternative sensory approaches, such as voice reading, writing with the hands, or vibration sensing, are all part of the rehabilitation training courses for people with

[1]https://www.who.int/zh/news-room/fact-sheets/detail/deafness-and-hearing-loss

hearing loss. Although sign language is the most popular means of communication for the deaf, most persons who do not have hearing loss have never taken sign language classes, making communication between these two groups difficult. Thus, in an effort to remove communication barriers between various groups, learning sign language has emerged as a major research area in education.

The majority of the most recent approaches to the teaching of sign languages [52–54] employ predefined gestures, while approaches based on user-defined interaction technology are more sparse in the literature. For example, Adamo *et al.* [55] proposed the development of a new immersive 3D learning environment to increase the mathematical skills of deaf children by teaching mathematical concepts and American Sign Language (ASL) math terminology through user interaction with fantasy 3D virtual signers and traditional interaction environments. Schioppo *et al.* [56] proposed a sign language recognition method using features extracted from data acquired by a Leap Motion controller from an egocentric view. The method was tested on the 26 letters of the ASL alphabet. In a related development, Phan *et al.* [57] used motion tracking to trial a number of different methods for providing user feedback in a sign language learning system.

Regarding research on the processes by which users can define themselves a vocabulary of hand gestures, Piumsomboon *et al.* [82] conducted research on hand gesture guessability in an Augmented Reality (AR) environment. They invited users to make gestures corresponding to certain tasks, and created user-defined gesture sets to guide the designers in implementing user-centred hand gestures for AR. To the best of our knowledge, there are no studies on learning ASL through user-defined interaction techniques. Hence, the purpose of this paper is to investigate if user-defined interaction techniques can enhance users' learning of ASL. We believe that this is an important research topic since most such systems use hand gestures created by system designers, which do not always reflect user intention.

To accomplish our research goals, we developed a simple system with a user-defined gesture interface for learning static ASL. In the system design, we have identified and taken into account shortcomings of prior systems, including the small data sets used to train the gesture recognizer, the absence of a realistic environment, and most importantly, the user's difficulty in engaging with the system for an extended period of time. With inspiration from Bragg's ASL Sea Battle [12], a sign language game created to help gather user data, we created and integrated a Whack-a-Mole style game with a user-defined hand gesture interface into the system, aiming at boosting user motivation. Finally, we conducted a user study based on a survey designed according to Schrepp's [58] recommendations and concentrated on user experience analysis.

Summarising, the main research question motivating our work, ***"Can user-defined interaction techniques enhance user motivation to learn static ASL?"***, was looked into within the context of a gamified environment for learning static ASL. Our main contributions are as follows:

1. We implemented a user-defined hand gesture interface for ASL learning with a Whack-a-Mole type of game.

2. We conducted a user study to examine if user-defined interaction affected users' experience. The initial results indicate a positive user attitude towards gamified learning environments and a strong interest of the users in user-defined interactions.

5.2 Related Work

The back-end of the proposed system supporting ASL learning with user-defined interaction mainly consists of hand gesture detection and a recognition model. Hence, we review prior research on hand gesture detection and recognition in ASL and user interfaces for creating user-defined hand gestures.

5.2.1 Sign Language Detection and Recognition

Real-time detection of dynamic hand gestures from video streams is a challenging task since: (i) there is no indication when a hand gesture starts and ends in the video; (ii) a performed hand gesture should only be recognized once; and (iii) the entire system should be designed considering memory and computational power constraints. Bheda *et al.* [59] proposed a method based on deep convolutional neural networks (CNNs) to recognize images of the letters and digits in ASL. Kim *et al.* [60] proposed a novel sign language recognition method, which employs an object detection network for a region of interest segmentation to preprocess the input data. Battistoni *et al.* [61] described a method for ASL alphabet recognition based on CNNs, which allows for monitoring the users' learning progress. Jiang *et al.* [62] proposed a novel fingerspelling identification method for Chinese Sign Language via AlexNet-based transfer learning and evaluated four different methods of transfer learning. Camgoz *et al.* [63] introduced a novel transformer-based architecture that jointly learns Continuous Sign Language Recognition and Translation while being trainable in an end-to-end manner. Zhang *et al.* [64] proposed MediaPipe Hands, a real-time on-device hand tracking pipeline to compute hand landmark positions from a single RGB camera frame for AR/VR applications. Goswami *et al.* [65] created a new dataset for ASL recognition and used it to train a CNN-based model for hand gesture recognition and classification. Finally, Pallavi *et al.* [66] presented a deep learning model based on the YOLOv3 architecture, reporting high recognition rates on the ASL alphabet.

Having reviewed the existing work on sign language recognition, we concluded that Mediapipe is the most suitable tool for the purposes of this paper, and thus, we used it for sign language recognition, benefiting from its highly accurate, real-time detection of hand landmark points. Moreover, as an open-source hand gesture detection framework from Google, it is well-documented and supported.

5.2.2 User Interfaces for User-defined Hand Gesture

A lot of work has already been done on user-defined hand gesture user interfaces, but most of them support limited functionalities, such as letting the user select one out of two established hand gestures as the one they want to use. For example, Wu *et al.* [83] proposed an interface for users to customize hand gestures and apply them to VR shopping applications in 2019, while they [84] proposed a user-defined hand gesture interface that could be used on in-vehicle information systems in 2020. Besides, conventional means of accessing visual communication markers (VCM) rely on input entry methods that are not directly and intimately tied to expressive nonverbal cues. Koh *et al.* [85] addressed this issue, by facilitating the use of an alternative form of VCM entry: hand gestures. Moreover, to fill this gap Dai *et al.* [86] presented a training system, called CAPG-MYO, for user-defined hand gesture interaction. Takayama *et al.* [87] perform two user studies to derive a user-defined gesture set allowing 13 types of table manipulation.

To address the issue that pre-defined gestures do not completely reflect user intent, we evaluated earlier work on user-defined gestures. As a result of these studies, we were also motivated to consider whether the addition of user-defined gesture interaction will reduce sign language learners' weariness and boost their motivation for learning sign language.

Therefore, the primary goal of our research is to investigate if user-defined gesture interaction affects ASL learning. In order to give users an immersive experience, we developed a VR-based system. To stimulate users' curiosity and boost their motivation, we also included a simple game with a user-defined hand gesture function. In addition, because there is a lack of user research on the subject, we used a questionnaire to survey users to investigate whether customised gesture interactions can actually inspire more people to learn sign language. Our main objective was to critically assess our system and gather user feedback on how their interaction with our system affected their learning experience.

5.3 System Components

This section provides an overview of the key components of the proposed system. The system's recommended workflow is shown in Fig. 5.1. When the user enters their account information through the login interface, the system initialises their location to the Instruction interface. After users familiarize themselves with the user introduction information of the Instruction interface, they study the ASL dictionary for five minutes and then visit the sign language game interface to play the game and increase their understanding of sign language through it.

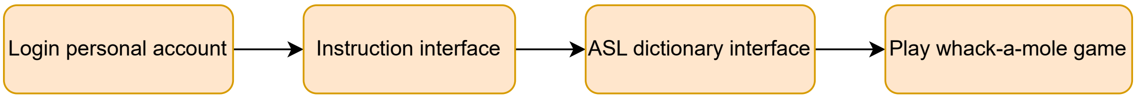

Figure 5.1: The workflow of ASL learning system.

5.3.1 Learning Environment

The learning tools and the game for learning the numbers 0 to 9 in ASL are displayed in Fig. 3.2, with the user's viewpoint tilted 45 degrees to the left. The entire scene was created in Unity (2020.3.32f1). Regarding the user's engagement with the system, we used the eye-tracking functionality of the HTC Vive Pro and enable clicking or picking an object after 3 seconds of the user's fixed attention. An inbuilt camera that was connected to the PC via openCV (version 3.4.2) [77] was used to acquire the images. Regarding gesture detection and recognition, Mediapipe is used to detect the user's hand and extract a series of 21 points matching corresponding landmarks on the detected hand. The feature vector from this sequence is then supplied as input to the classifier, which is an MLP with three fully connected layers, implemented in Tensorflow 2.6.0 [81] and Python 3.6 [80]. We used an RTX3080 GPU on a standard PC to train the classifier. The study's objectives were satisfied with an overall recognition accuracy rate of over 90%, which is expected to offer a generally positive user experience.

5.3.2 Whack-a-Mole Game and User-defined Interface

We adopted the Whack-a-Mole game and implemented a sign language-based version of it, aiming to make learners more interested in the material, increasing their motivation and, eventually, their engagement with the process. In our game, as shown in Fig. 5.2, each location is marked by a unique numeric identifier. If the user signs correctly the current position of the gopher, one point is added; otherwise, no point is awarded. The total duration of the game is 30 seconds.

The user-defined interface is a feature that is hidden from the user while they are playing the game. It will only appear when a mole is killed and trigger a hitherto hidden functionality, calling for the user to specify a wake-up gesture to be utilised later in the game. In the example shown in Figure 5.2 and Figure. 5.3.2, the user is

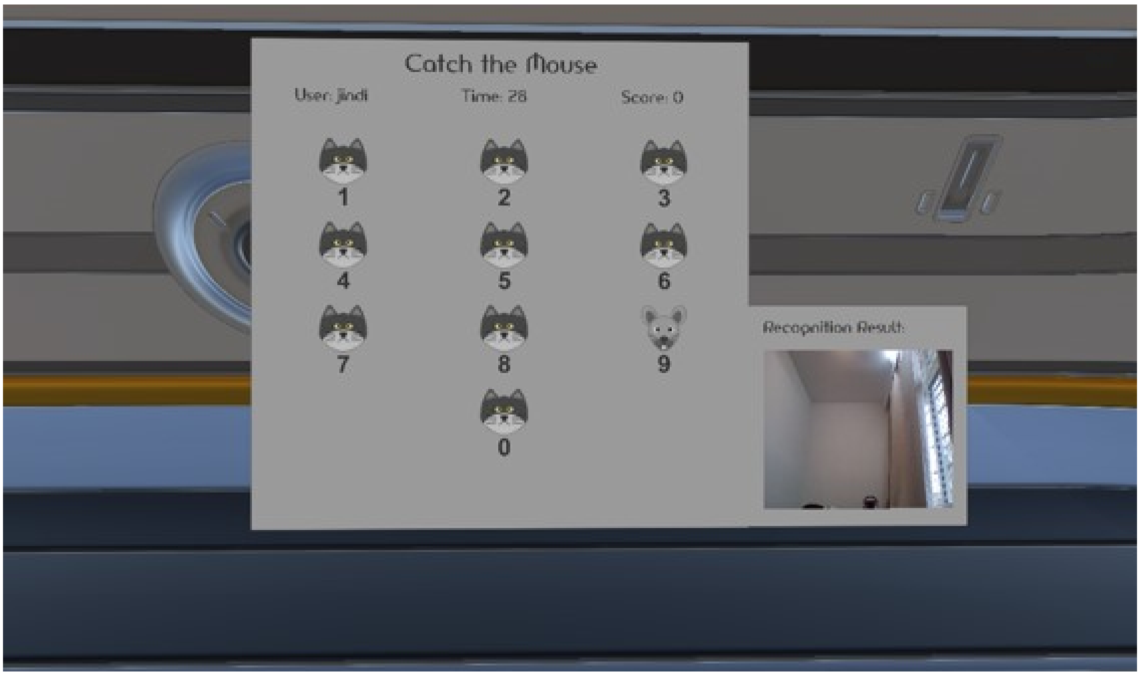

Figure 5.2: The Whack-a-Mole game for ASL learning.

given the special game skill to “Clear all Moles”. At the end of one iteration of the course, the system will collect user-defined gesture data for 5 seconds, retrain the recognition model, and the user-defined interface will be hidden again. Now that the user has picked their special skills, they may start playing the game again by clicking “Start” on the game screen, and they can use their newly acquired special game skill. The workflow of the user-defined hand gesture interface is shown in Fig. 5.4. Notice that the special mole sequence must match the last digit of the remaining play time for the user interface to be activated. In addition, the user must be able to recognise the special mole's number at a specific moment. When both requirements are satisfied, the user-defined interface will be displayed and the user will be awarded their special game skill.

5.4 Experiments

To evaluate the system design, we adopted the user survey scheme proposed by Schrepp *et al.* [58], which comprises six evaluation factors: **Attractiveness**, **Efficiency**, **Perspicuity**, **Dependability**, **Stimulation**, **Novelty**. Each factor is further divided into six or seven more specialised issues. Table 3.1 displays the spe-

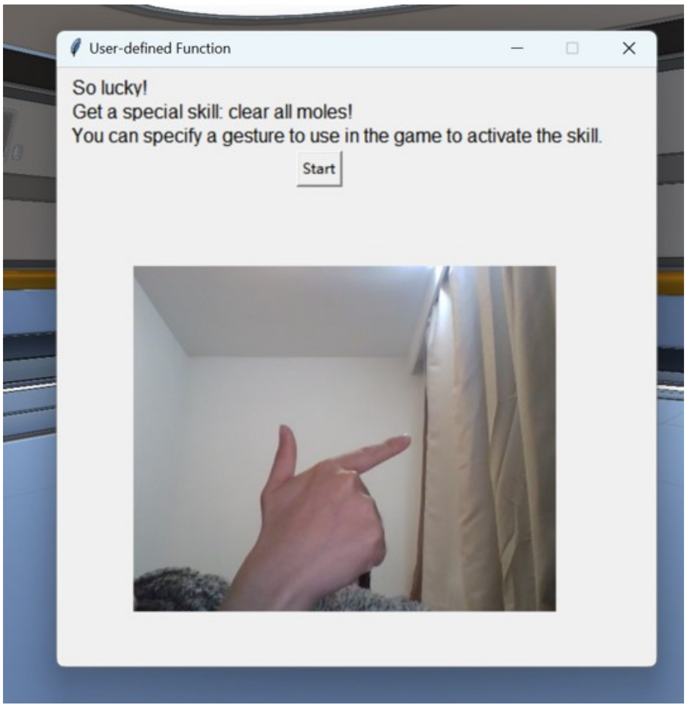

Figure 5.3: The user-defined interface.

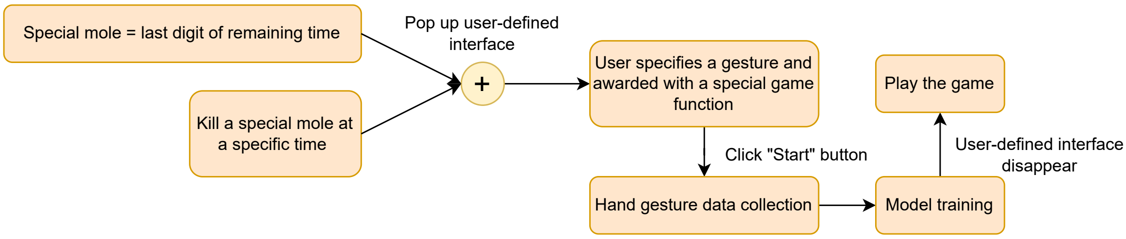

Figure 5.4: The workflow of the user-defined hand gesture interface.

cific issues associated with each factor. Based on the users' scores on a scale of 1.00 to 5.00 on particular issues, we assessed the merits of the system in each factor.

Participants were recruited through informal invitations shared via students' personal networks using social media platforms. The research assistants—who had no academic authority over participants—shared the invitation message within these networks to reach undergraduate students. All participants were aged between 19 and 21, had no prior formal training in ASL, and had limited prior exposure to immersive VR environments. Importantly, none of the participants who took part in the study described in Chapter 5 were involved in the studies described in Chapters 3 or 4. Each participant group was independent, ensuring there was no carry-over effect or novelty bias due to repeated exposure to similar environments or evaluation tools. The majority of the users had either very limited or no prior understanding of ASL, or any other sign language. They were instructed to explore the system, adhering to the instructions in order to learn ASL in three stages: learning signs from the dictionary interface; improving comprehension at the learning interface; and assessing their learning three times on the game interface. As it can be challenging for some beginners to pick up so many motions quickly, users were merely required to learn the ASL 0-9 numerals.

5.5 Result Analysis

The user evaluation is summarized in Fig. 5.5, the box-plots showing the Minimum, First Quartile, Median, Third Quartile, and Maximum, while the Mean is shown by an 'x'. The score distribution reflects generally positive feedback on the evaluation factors, all of which received mean scores greater than 3.00, while the overall system achieved a satisfactory average score of 3.42 (SD = 0.88) over the six factors. It is also interesting to note that some low scores (< 2.50) were given in all factors, the possible causes of which are discussed below.

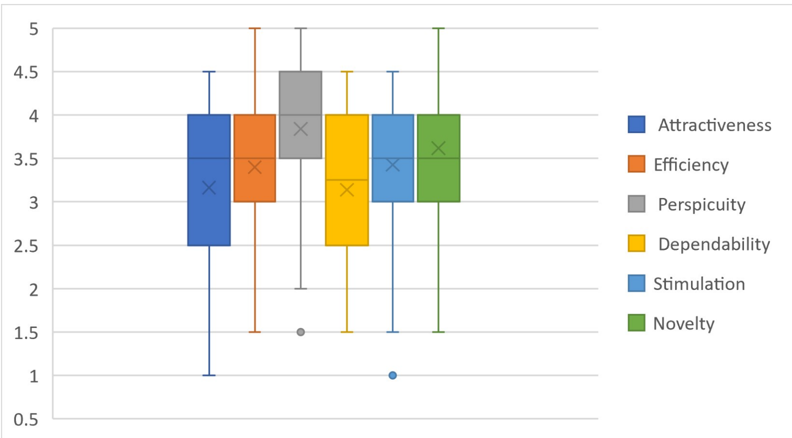

Figure 5.5: Box-plots of the score values for each of the user survey's factors.

Attractiveness: as shown in Fig. 5.6, the average score over the 7 questions is 3.16 (SD = 0.90). Each question has some scores lower than 2.50, possibly reflecting some lack of interaction with the users. For example, some users reported that animations should show up when the hand gesture was recognised correctly. Warnings should

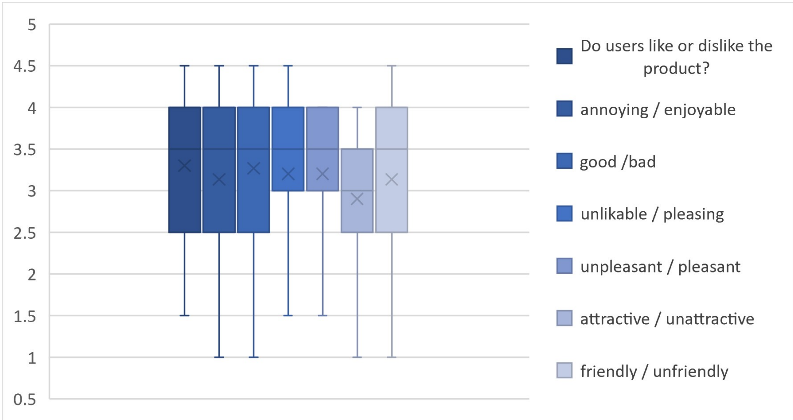

Figure 5.6: Box-plots of the scores for each subdivision of the attractiveness.

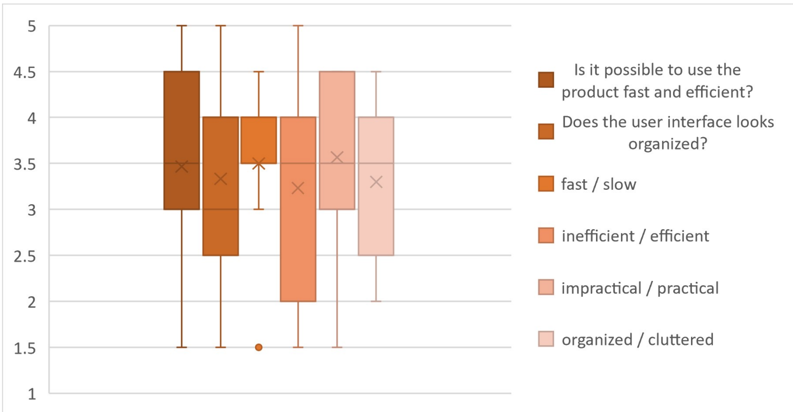

Figure 5.7: Box-plots of the scores for each subdivision of the efficiency.

also be shown if no hand was detected, or when the hands were too close to the camera. Future improvement plans include the addition of more interactive features, such as moving backgrounds, scene changes, and animations.

Efficiency: as shown in Fig. 5.7, the average score is 3.40 (SD = 0.93). There were 3 users who gave a score of 5.00 on some questions. However, nearly one-third of them gave scores lower than 2.50 on each question, indicating that there is still room for increasing the system's efficiency. According to user feedback, the practice interface was not so convenient to use, as the users had to click the button "Start" to check for correctness. Improvements could be made to automate this process, thus requiring less activity of this type from the users.

Perspicuity: as shown in Fig. 5.8, the average score of 3.84 (SD = 0.77) is the highest among the six factors, indicating that most users saw the system as easy to use, perhaps because of the simple design of the interface, which made it easy to use.

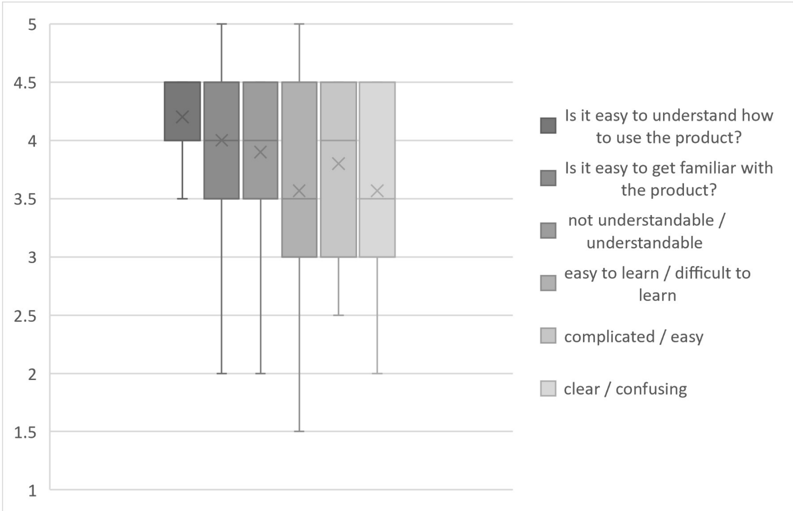

Figure 5.8: Box-plots of the scores for each subdivision of the perspicuity.

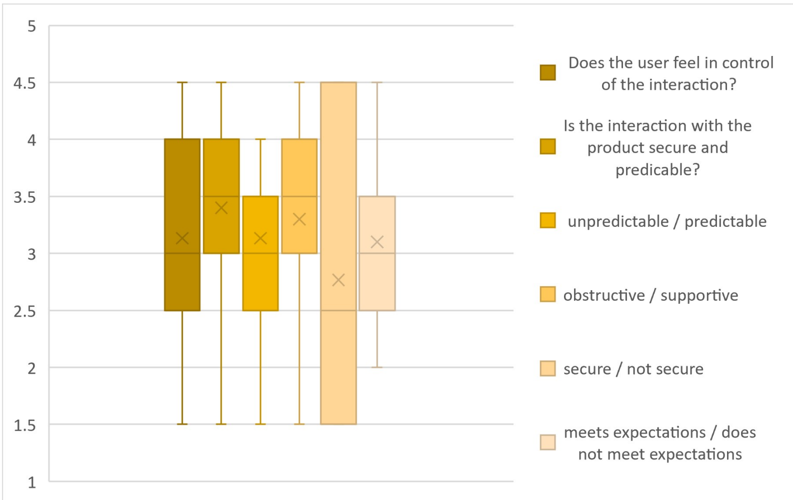

Figure 5.9: Box-plots of the scores for each subdivision of the dependability.

Nevertheless, one user complained about the scene navigation system, overlooking apparently the navigation instructions button of the main menu. Thus, when this particular user tried to move to the ASL 0-9 dictionary, they did not know how to do it until we told them to look at the button. In a future study, the users will first be taught how to navigate the scene, before going into the main study of gestures.

Dependability: as shown in Fig. 5.9, the average score is 3.14 (SD = 0.90). Although most users gave scores higher than 3.50, a small number of users gave scores of 1.50. Perhaps this was because, in the practice interface, some users thought that they did the correct gesture but were judged as being wrong. The reason behind this can be performance issues of the gesture recognition model in the back-end, perhaps

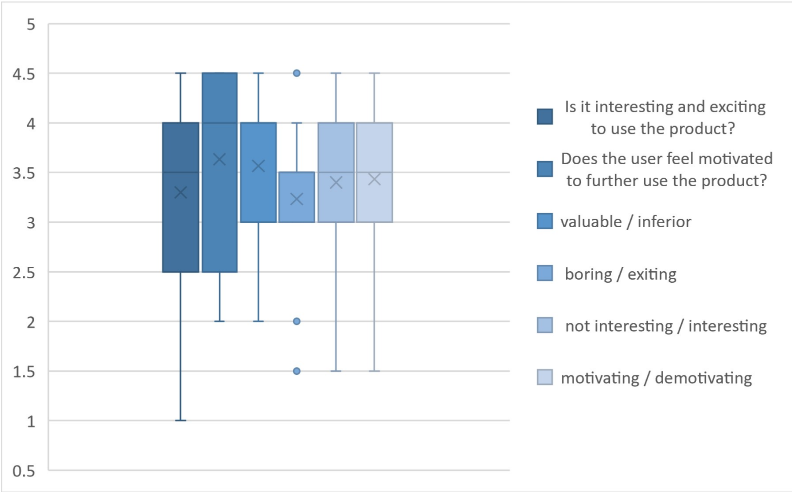

Figure 5.10: Box-plots of the scores for each subdivision of the stimulation.

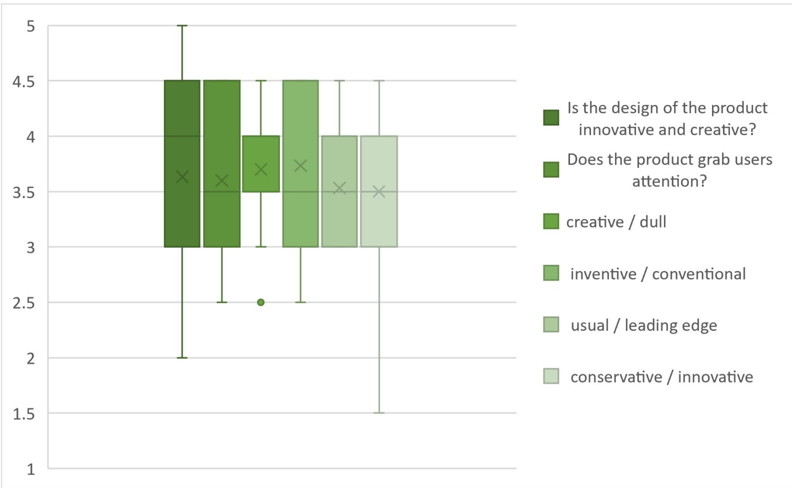

Figure 5.11: Box-plots of the scores for each subdivision of the novelty.

because the training data for some gestures might have been of poor quality, thus leading to low recognition accuracy rates for these gestures. Future improvements will aim at training recognition models that will be able to better generalize to natural physiological differences in users' hands.

Stimulation: as shown in Fig. 5.10, the average score for this factor is 3.43 (SD = 0.85), while the mean value of every question under this factor is above 3.25. It is also worthwhile to note that the first quartiles of all questions start from just below 3.00, and most of the users gave scores between 3.50 and 4.00. Nevertheless, were also some low scores in all questions, showing that there is still room for improvement. In particular, the interaction of the system could be designed more creatively, aiming at better inspiring the users.

Novelty: as shown in Fig. 5.11, the average score of this factor is 3.62 (SD = 0.70), suggesting a broadly positive reception, with the mean score in all questions at 3.50 or above. The second and the fourth questions have relatively tight distributions, with interquartile range between 3.00 and 4.50. Overall, the results on this factor suggest that most of the users regarded the system as being innovative, with only a few of them perceiving it differently.

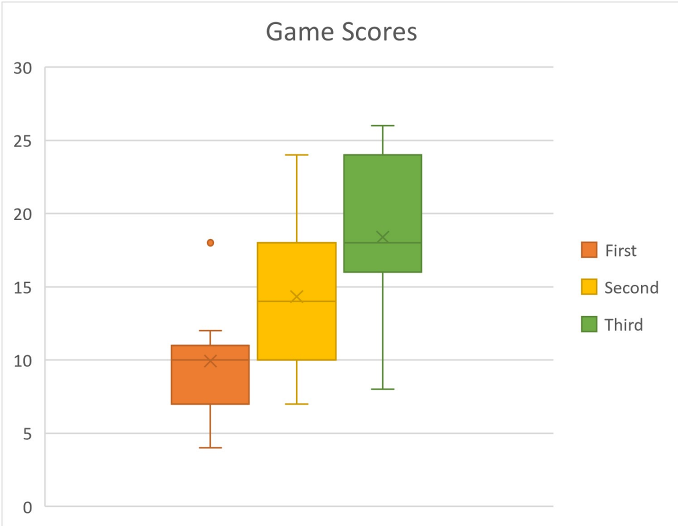

Figure 5.12: Box-plots of the game scores for each of the three attempts.

Fig. 7.11 shows that users typically had poor game scores at their initial attempt, with the exception of one outlier with a score of 18, who had the good fortune to activate the user-defined interface and obtain the hidden game skill on their first attempt. We note that the average, lowest, and highest user score all gradually increased at the second and third attempt, showing that the user's sign language proficiency increased. Additionally, several users claimed that this hidden feature might stimulate their interest in the game, and, implicitly, help them advance their sign language skills.

5.6 Discussion and Limitations

Using VR technology, we developed an immersive environment for learning ASL. We looked into whether a user-defined way of interaction could boost users' motivation to learn sign language. We evaluated this issue using the survey method proposed by Schrepp *et al.* [58].

For the user evaluation of the system we used six assessment factors. The survey's results on these six metrics demonstrate how well the user-defined interface for ASL learning operates and that it can genuinely satisfy user needs. Besides, because user-defined interactions are more experiential, most users seem to accept them, according to the analysis of environmental experience. However, the system still needs to be further optimised and adjusted for some functions to improve user experience, even though the majority of users are happy with user-defined interactions.

'Our work still has a number of limitations in terms of system design and implementation, as was already mentioned. Here, we summarise them for each assessment factor separately. **Attractiveness**: when sign language was correctly interpreted, some users complained that there weren't any animated clues. Animations in suggestions could enhance user experience. **Efficiency**: because users have to manually touch the start button to play the sign language game, some users complained that the user interface wasn't sufficiently automatic. **Perspicuity**: some users have complained that they were unsure of how to navigate the scenario. **Dependability**: while playing the game, some users claimed to have used the correct sign language, but the algorithm determined that they had not, giving them a lesser score. **Stimulation**: a small percentage of users gave low scores on the stimulation factor study, which suggests that there is still room for our system to be designed more creatively. **Novelty**: a small proportion of users felt that the system wasn't inventive enough, possibly because they thought the learning model was too simple.

On the other hand, our study has a number of methodological drawbacks. The user study and included only 15 participants. The invited people were between the ages of 19 and 21; there is no research on users in other age groups. To further test our methodology, we intend to enlist more individuals in future studies, who should come from a wider range of backgrounds (e.g., age). Ultimately, there is no specific evaluation indicator in the questionnaire survey on sign language acquisition. A uniform and standardised questionnaire-based assessment of sign language acquisition is needed for the next research.

We acknowledge the importance of clarifying participant independence for validity and interpretability of findings, especially in relation to dimensions like novelty and stimulation. Since all three studies (Chapters 3, 4, and 5) used different participant samples, comparisons across chapters are based on descriptive trends rather than direct statistical contrasts. Furthermore, the goal of each chapter was distinct: Chapter 3 focused on immersion vs. web-based learning for numeric ASL, Chapter 4 explored gamified learning of alphabetic ASL, and Chapter 5 investigated user-defined gesture learning. As such, each study addresses a different research question. Future work could benefit from a within-subject design to enable direct comparison under controlled conditions.

5.7 Conclusion

A virtual environment that allows users to learn ASL through the use of user-defined hand gestures has been developed by our team. The user interface that is embedded in the virtual environment made it possible to most of the users to readily comprehend the workflow of the system, as well as each stage of the ASL learning process. The results of a user questionnaire that we carried out (N = 15) revealed that participants were, in general, pleased with the digital ASL learning system that we developed. In conclusion, the overall results provide credence to our original hypothesis, which stated that an increase in users' motivation to learn can be attributed to the usage of user-defined interaction modalities.

In the future, we will include in the system more interactive components, such as backdrop movement, scene changes, and animation prompts. In order to reduce the amount of human involvement required to control the system, we will also add more automatic settings. To help the user understand how to manipulate the objects in the scene, a follow-through user interface will be developed. In addition, a stronger gesture recognition model will be developed, enabling the inclusion of more sophisticated sign language instruction materials.

CHAPTER 6

Comparing 2D and 3D Interfaces in an ASL Learning Environment

In this chapter, we build upon the foundational work presented in Chapter 3 by delving deeper into the comparative impact of 2D and 3D elements within virtual reality (VR) environments on the learning of American Sign Language (ASL). We investigate how different dimensional representations can influence learner engagement and the efficacy of learning sign language. The findings and methodologies of this research have been presented at the 31st IEEE Conference on Virtual Reality and 3D User Interfaces.

6.1 Introduction

Learning American Sign Language (ASL) has significant benefits for both the deaf community and hearing individuals. First and foremost, knowing ASL makes it easier to communicate with people who are deaf or hard of hearing. This promotes inclusion and lowers barriers to communication. Learning ASL also shows a dedication to building an inclusive society that values diversity and guarantees equal access for all. In addition to its social benefits, ASL also offers unique linguistic and cognitive benefits [88]. Research has shown that learning sign language improves cognitive functions [89] such as memory, attention, and problem-solving skills. It also enhances visual-spatial awareness and can help to improve reading and math skills [90]. Finally, being proficient in ASL provides access to prospective job opportunities in a variety of fields, including education, healthcare, and customer service. As the demand for ASL interpreters, educators, and other professionals continues to grow, learning ASL is becoming an increasingly important skill.

Virtual reality (VR) and gamification technologies are rapidly advancing, and their potential to revolutionise sign language learning is gaining attention from researchers. The integration of gamified approaches in sign language education offers a range of advantages, including enhanced interactivity, immersive practice experiences, and heightened learning motivation. Gamified sign language learning can provide learners with interactive and engaging practice opportunities that are more effective than traditional static materials. By practising their signing skills in a simulated environment, learners can develop confidence and fluency in using language in real-world situations [91–93]. Specifically, gamified sign language instruction can improve engagement and learning effectiveness by providing interactive and practical teaching methods. It can provide learners with opportunities to practice their sign language communication skills in a relaxed and enjoyable environment. Examples of gamified sign language learning activities include interactive exercises, immersive practice environments, and gamified learning activities such as sign language puzzles, challenges, and competitions [94–97].

As a result, researchers have conducted extensive research on gamifying sign language learning, primarily emphasising the creation and advancement of 2D and 3D games. For example, Zafrulla *et al.* [98] proposed CopyCat, a 2D game specifically designed to engage young children in interactive computer-based ASL learning, employing cutting-edge gesture recognition techniques. The game features a user-friendly interface with tutorial videos demonstrating accurate signing, a live video enabling gesture input for the recognition system, and real-time feedback on the child's progress. Additionally, the game features a character named Iris the cat, who carries out the child's instructions. Another prominent example of an educational game is MatLIBRAS Racing [99], which is designed to teach sign language for natural numbers from a cognitive perspective. The game received positive feedback from students in terms of its educational and gaming features, fostering social relationships among players and facilitating sign language learning. The results suggest that MatLIBRAS Racing holds promise as an effective educational tool, particularly in academic settings. Regarding 3D games, Economou *et al.* [100] presented a work-in-progress study evaluating the combined impact of scaffolded instruction and gamification in a 3D interactive game to support learning the sign language alphabet. Adamo *et al.* [101] presented the second iteration of the SMILE project, which introduces an immersive 3D game tailored for both deaf and hearing children. The game's enhanced design and user interaction had been carefully crafted to elevate motivation and appeal, fostering an engaging and captivating experience. Although the evaluations demonstrated that the application was enjoyable and user-friendly, the study did not assess the actual learning outcomes achieved by the participants. Collectively, the current proposals for gamified sign language learning have made valuable contributions to the field, advancing our understanding and paving the way for further developments. However, most of these approaches have focused on either 2D or 3D game formats, with limited exploration of user experience differences between these two genres.

This study builds on theories from human-computer interaction and learning psychology. Research in multimedia learning, including Mayer's Cognitive Theory of Multimedia Learning [102], highlights that interface representation can impact engagement, cognitive load, and comprehension. Similarly, studies on educational affordances of 3D environments have shown that dimensionality influences learner motivation, presence, and spatial reasoning [11, 103]. Human-computer interaction literature also suggests that 2D and 3D interfaces offer different trade-offs in terms of usability and immersion [104]. Therefore, comparing 2D and 3D layouts within an otherwise identical learning platform is theoretically justified to determine which better supports user satisfaction and engagement in the context of ASL learning.

To address this gap, we conducted user studies based on the methodologies proposed in Bangor *et al.* [105], Schrepp *et al.* [58], and performance scores, to answer the following research question: ***"How does the mode of interface representation (2D vs. 3D) influence user experience in a gamified ASL learning environment?"***. Our main contribution is a user study that compares the potential of 2D and 3D games in enhancing user experience and improving learning outcomes. Empirical studies have shown that 3D game environments can positively impact learning outcomes by increasing learner engagement, motivation, and spatial understanding [11,106]. However, user feedback from our study indicated that while the 3D version of the ASL learning environment was generally well-received, some users reported challenges related to system responsiveness and visual clarity. These findings suggest that dependability and clarity are key areas for improvement in our implementation of 3D learning interfaces.

6.2 Learning Environment and Games

The immersive learning environment for learning ASL numerals from 0 to 9 is shown in Figure 3.2. It was created in Unity (version 2020.3.32f1). Users interact with the scene utilising the eye-tracking functionality of HTC Vive Pro. When a user's attention is fixed for three seconds, the system allows them to click and choose an object.

For image acquisition, a built-in camera was utilised, connected to a PC that employed openCV (version 3.4.2) [77]. Gesture recognition was implemented using Mediapipe [64] because of its accessibility as an open-source platform and its intrinsic compatibility with the built-in camera on PCs. When Mediapipe detects a user's hand, it returns a sequence of 21 feature points (p_0, p_1, p_2,..., p_{20}) representing landmarks on that hand. The coordinate frame's origin was set at p_0, the point near the user's wrist at the bottom of the palm. The classifier employed was a multilayer perceptron consisting of three fully connected layers, implemented in Python 3.6 [80] and Tensorflow 2.6.0 [81], yielding recognition accuracy rates exceeding 90%. This level of accuracy was considered satisfactory for the study's objectives, ensuring a smooth user experience.

To enhance the effectiveness of the learning process, a question-answer module was integrated into the system. This module allows users to evaluate their proficiency level and practice their signing skills by responding to randomly generated questions from a database. In Figure 3.5, an example is shown where the system presents the question "Can you sign for 9?". The user has the option to refer to the dictionary or use their already acquired skills to sign the number '9'. Alternatively, they can choose the "I don't know" option, prompting the system to demonstrate the correct expression. In this case, the user is encouraged to continue practising until they feel confident enough to sign the digit independently, pressing the relevant button.

6.2.1 2D Game and 3D Game

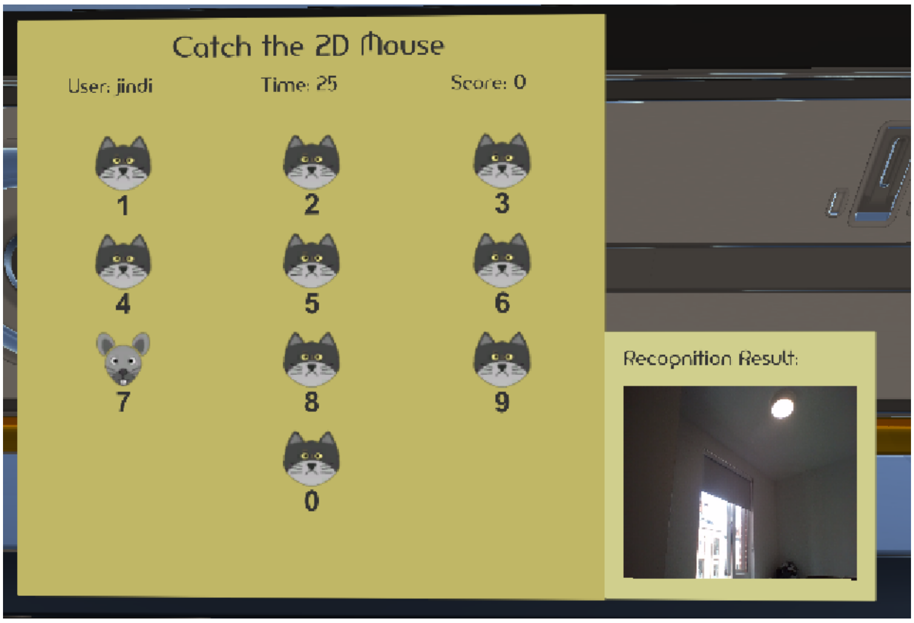

Figure 6.1: 2D Whack-a-Mole Game

To explore potential differences in user experiences during ASL learning, we developed a 2D and a 3D version of our Whack-a-Mole game. The objective of the game remains the same in both versions: players must promptly identify and sign the current location of the gopher within a designated time limit.

In both 2D and 3D versions, the game interface consists of a grid-like layout where each location is uniquely identified. Players earn one point for correctly signing the position of the gopher, indicating their understanding of ASL gestures.

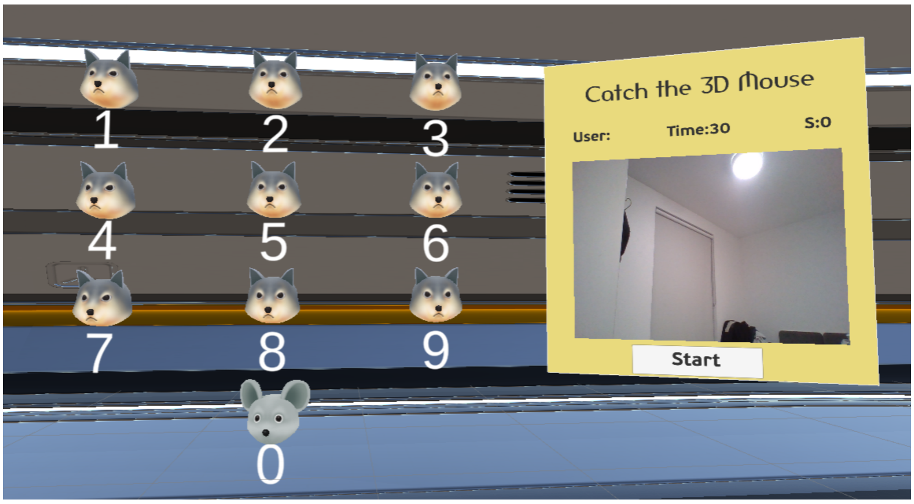

Figure 6.2: 3D Whack-a-Mole Game

No points are awarded if the player fails to accurately sign the gopher's location. To maintain an engaging experience, we have implemented a time limit of 3 seconds for each round. If the player does not sign the correct location within the given time, a new gopher automatically appears, providing the player with another opportunity to score points. The game duration is set at a total of 30 seconds, creating a challenging yet manageable time frame for players to showcase their ASL recognition and signing skills.

By offering both 2D and 3D versions of the game, we aimed to investigate potential variations in user experiences and learning outcomes across the two interfaces. This allowed us to explore the potential influence of immersion and visual depth within the 3D environment on players' engagement, motivation, and overall ASL learning, compared to the 2D version. Figure 6.1 and Figure 6.2 show the interfaces of the 2D and 3D versions of the Whack-a-Mole game, showcasing the grid layout and the gopher's locations.

6.3 Methodology

6.3.1 Participants and Procedure

Participants were recruited through informal invitations shared via student personal networks using social media platforms. A research assistant—who had no academic authority over participants—shared the invitation message within these networks to reach undergraduate students and other people. All participants were aged 23-34 years (M = 28.13, SD = 2.55), had no prior formal training in ASL, and limited prior exposure to immersive VR environments. Importantly, none of the participants who took part in the study described in Chapter 6 were involved in the studies described in Chapters 3, 4, or 5. Each participant group was independent, ensuring there was no carry-over effect or novelty bias due to repeated exposure. After being briefed on the purpose of the study, participants signed consent forms to participate in the experiment. The whole procedure followed the Ethics Guidance of our University ethics committee.

To enhance the learning process by combining instructional scaffolding approaches, we divided the learning process into three stages: **Learn**, **Practice**, and **Assess**. It is a recommended strategy for ASL learning, entailing the creation of a teaching tool that combines innovative technology with gamification components [107]. The details of the learning process are described below:

Learn: Participants were encouraged to explore and navigate the virtual learning platform at their own pace, familiarising themselves with the layout and function of each user interface. Following this, they were asked to spend three minutes familiarising themselves with the 0-9 ASL expressions, utilising the 0-9 ASL dictionary interface.

Practice: To enhance comprehension of 0-9 ASL, users were asked to engage in a Q&A session on the Quiz interface, for a period of three minutes, during which they were presented with randomly generated questions to respond to.

Assess: After the users have demonstrated their proficiency in 0-9 ASL by successfully completing the initial two stages, we presented them with the Whack-a-Mole game. This stage aims to compare the effects of the 2D and 3D versions of the game. First, we asked the participants (2D and 3D, each group has 12 users) to play the game once (30s) and to complete the system usability scale questionnaire for around 1 min and the user experience questionnaire for around 3 mins. In the second attempt at the game, half of the users continued in the same environment, while the other half swapped between 2D and 3D. That is, we had four groups in total, of six participants each: **2D-2D**, **2D-3D**, **3D-2D**, and **3D-3D**. This experimental design was inspired by the replicate cross-over designs used in medical research [108]. This design allowed us to distinguish between the learning effect on the sign language and the learning effect on the interface. We note that because of the symmetry of the design, any possible learning effect does not skew the results. Moreover, the decision to limit each gameplay session to 30 seconds was a deliberate methodological choice. The aim was to ensure that all participants experienced a controlled and equal exposure to the core gameplay loop—identifying a target, performing the sign, and receiving feedback. Longer durations could have introduced variability due to fatigue or reduced motivation, especially in crossover conditions. The goal of this study was to evaluate users' initial impressions of usability and user experience based on their interaction with the interface and game mechanics, rather than long-term engagement. The usability (SUS) and UX (UEQ) questionnaires were thus applied immediately after interaction to capture these early impressions while the experience was fresh in memory.

6.3.2 Evaluation Methods

Three dimensions of the VR learning environments were evaluated and compared: *usability*, *user experience*, and *user performance*.

For **usability**, we used the well-established SUS (System Usability Scale) questionnaire [105]. Participants were asked to rate the 2D and 3D versions of the VR learning environments on a scale of 1 (strongly disagree) to 5 (strongly agree) after using them.

For **user experience**, we employed the user survey method proposed by Schrepp *et al.* [58]. This approach utilises six scales, each representing a distinct aspect of the user experience: *Attractiveness*, *Efficiency*, *Perspicuity*, *Dependability*, *Stimulation*, *Novelty*. These scales provide a comprehensive framework for evaluating the users' perceptions and attitudes. Each scale comprises several specific items that capture various dimensions of the user experience, as outlined in Table 3.1. To collect data on the user experience, we used a 7-point Likert scale for each of the 26 items in the questionnaire. Participants were asked to rate their level of agreement with each statement, ranging from 1 (strongly agree with a negative statement) to 7 (strongly agree with a positive statement). This rating scale allowed for a quantitative analysis of the participants' perceptions across the various dimensions of the user experience.

Finally, for **user performance**, i.e., how well they learned using the VR environments, we collected and analysed the players' game scores.

6.4 Results

6.4.1 Comparing Usability

On a scale of 0 to 100, with 0 representing the lowest usability and 100 representing the highest usability, both the 2D and 3D game environments achieved high usability ratings on the SUS scale. The 2D game received an average score of 76.25 (SD = 5.15), while the 3D game scored 80.63 (SD = 3.41), a two-tailed independent samples t-test was conducted to evaluate whether the difference in usability scores between the two interfaces was statistically significant. The resulting p-value was 0.103, indicating no significant difference in usability ratings between the 2D and 3D game environments. These scores indicate that both game environments received "good" usability ratings [105]. The low standard deviations observed for both games suggest that the majority of participants had consistent and positive experiences with the usability of the interfaces. These findings highlight the success of creating user-friendly interfaces for both 2D and 3D games, with the transition to 3D not significantly impacting usability. This suggests that both 2D and 3D games can be equally effective in terms of usability, and it is important to consider the specific needs and preferences of the target audience when designing game interfaces.

The comparable high usability scores achieved by both the 2D and 3D game environments demonstrate the success of our design in terms of intuitive controls, clear visuals, and engaging gameplay mechanics. The positive SUS ratings indicate that users found both interfaces easy to navigate and efficient.

6.4.2 Comparing User Experience

Table 6.1 presents Guttman's lambda-2 coefficients for both 2D and 3D game environments, assessing the internal consistency of each user experience dimension. In the 3D condition, the dimensions of **Attractiveness**, **Efficiency**, and **Stimulation** demonstrate strong internal consistency ($\lambda > 0.7$), suggesting that participants responded consistently across items within these constructs. **Perspicuity** shows moderate consistency, while **Novelty** displays some variability, indicating room for improvement in how this dimension is measured. **Dependability**, however, reveals poor internal consistency, suggesting the items under this scale may require refinement or that users had differing interpretations of system dependability.

In the 2D condition, **Attractiveness** and **Efficiency** also demonstrate good reliability. However, **Perspicuity** ($\lambda = -0.01$) and **Stimulation** ($\lambda = -0.30$) show negative lambda-2 values, indicating weak or no internal reliability for these scales. These negative values suggest that participants responded inconsistently to items within these dimensions, which may be due to ambiguous questionnaire phrasing, reduced salience of these constructs in the 2D interface, or variability in user perception of clarity and engagement. **Dependability** shows modest consistency in the 2D condition, while **Novelty** remains moderately consistent.

These reliability issues highlight the need for future refinement of the measurement instrument—either by improving the clarity of questionnaire items or by enhancing the interface design to better convey concepts like perspicuity and stimulation, particularly in 2D environments.

Table 6.1: Guttman's lambda-2 coefficient values

Dimension Factor	**3D**	**2D**
Attractiveness	0.78	0.79
Perspicuity	0.54	-0.01
Efficiency	0.78	0.87
Dependability	0.38	0.47
Stimulation	0.75	-0.30
Novelty	0.46	0.52

Figure 6.3 visualises a comparison of user experience between the two game environments. In summary, the user experience analysis reveals that both 2D and

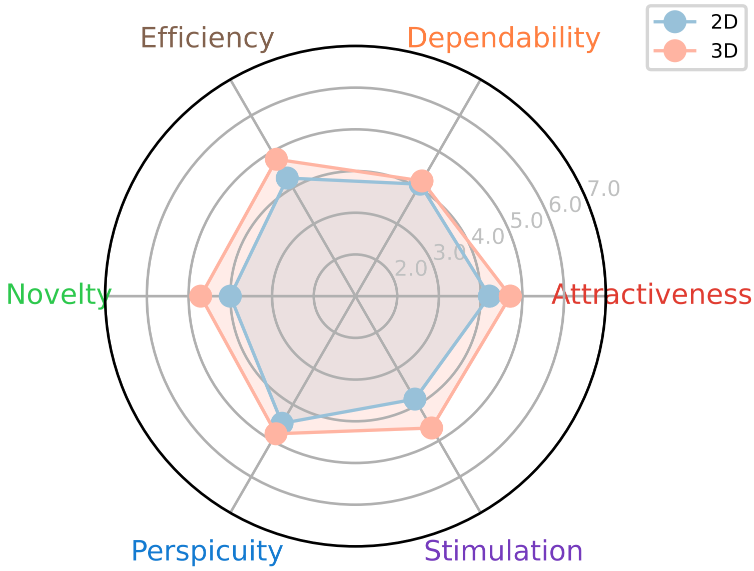

Figure 6.3: Six scales of user experience

3D games received positive ratings across various scales. However, the 3D game outperformed the 2D game in most aspects, including visual attractiveness, efficiency, novelty, perspicuity, and stimulation. The users found that the visuals of the 3D game were appealing and innovative, providing a fresh and immersive gaming/learning experience, as shown in Figure 6.4.2 and Figure 6.4.2. Additionally, the 3D game demonstrated higher efficiency, offering streamlined user and interface interactions, but in terms of "organized", it is similar to those of the 2D game, as shown in Figure 6.4.2. The interface of the 3D game was also perceived as easier to learn and understand, enhancing the overall user experience, but in terms of "easy", it is lower than that of the 2D game, as shown in Figure 6.4.2, which means users found that the 2D game is easier to operate. Moreover, the 3D game succeeded in delivering a more stimulating and engaging user experience, eliciting higher levels of excitement and immersion from the users, as shown in Figure 6.4.2.

Overall, the findings suggest that the added dimensionality and immersive elements of the 3D game contributed to a superior user experience compared to the 2D game. While the 3D game excelled in several areas, Figure 6.4.2 indicates that both 2D and 3D games were perceived as having low dependability by the users, as they thought that both games met their basic needs, but there are areas where they would like to see improvements, such as the complexity of the game, the difficulty of completing the level, and so on. Despite the strengths of the 3D game, it is worth noting that the 2D game still received positive ratings across all scales, indicating that it provided a satisfactory user experience as well. The results emphasise the importance of considering various factors, including visuals, efficiency, novelty, perspicuity, and stimulation, when designing game environments to enhance user experience and engagement.

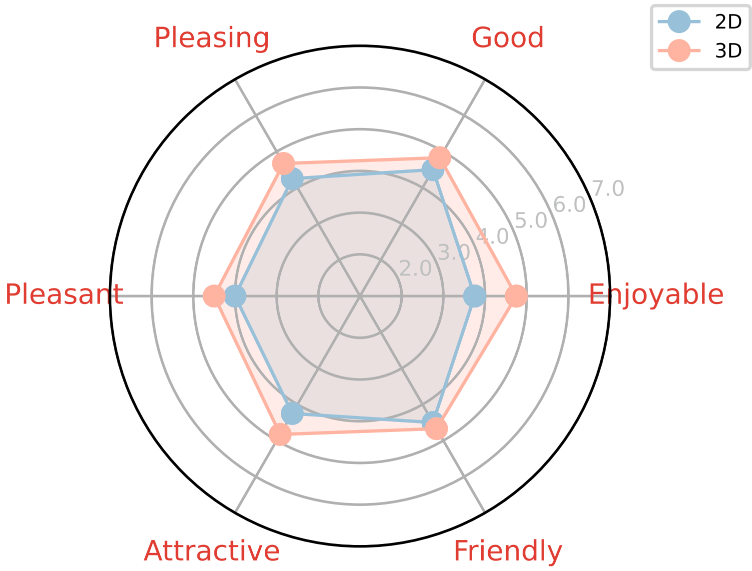

Figure 6.4: Mean scores for each item of the Attractiveness of user experience.

Table 6.2: Means and p-values for the three groups of scales.

Pragmatic and Hedonic Quality	**2D**	**3D**	**P-Value**
Attractiveness	4.21	4.71	1.304E-03
Pragmatic Quality	4.30	4.60	1.150E-02
Hedonic Quality	3.93	4.68	5.056E-09

In the study by Schrepp *et al.* [78], the scales of the user experience questionnaire were categorised into two groups: pragmatic quality and hedonic quality. Pragmatic quality includes perspicuity, efficiency, and dependability, which are related to task performance. Hedonic quality includes stimulation and novelty, which are related to non-task-related aspects. Attractiveness, on the other hand, is considered a pure valence dimension. Table 6.2 presents the mean scores for these grouped scales for the two different games, along with the corresponding p-values. The results indicate

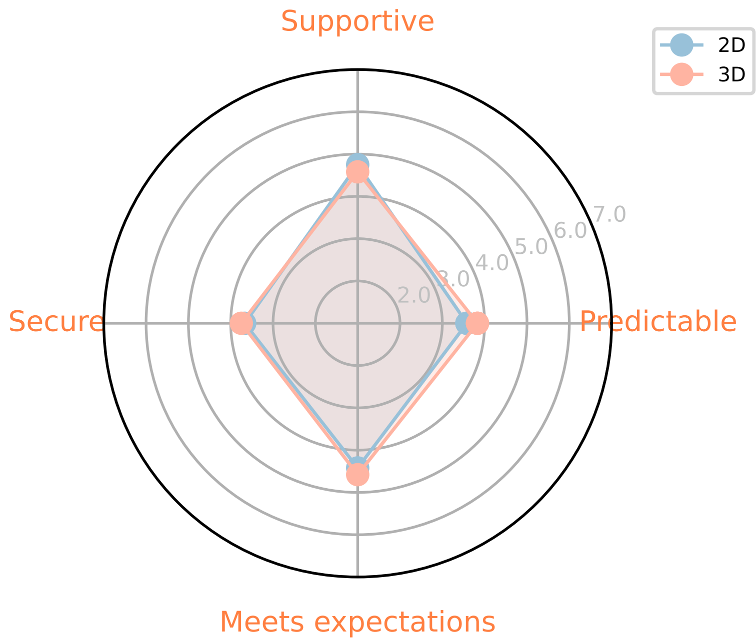

Figure 6.5: Mean scores for each item of the Dependability of user experience.

that the user preference for the 3D game was statistically significant, especially regarding hedonic quality.

6.4.3 Comparing User Performance

Table 6.3 summarises the results, for each of the four user groups and for each of their two attempts on the game. First, we conducted the Shapiro-Wilk test to see if the data distribution in the 2D and 3D environments matched the normal distribution. All datasets we found to follow the normal distribution, $p > 0.05$.

Table 6.3: Score means and standard deviations, for the four groups, for the first (I) and the second (II) attempts.

Group	**mean (I)**	**s.d. (I)**	**mean (II)**	**s.d (II)**
2D-2D	10.50	2.22	17.33	2.92
2D-3D	9.83	1.34	14.67	2.49
3D-2D	9.33	1.70	14.83	1.95
3D-3D	9.67	1.49	16.00	2.58

To measure the effect of the choice between 2D and 3D, we compare scores over the two environments, on the first attempt only. These scores correspond to the first column of the table, and we compare the first two rows against the last two.

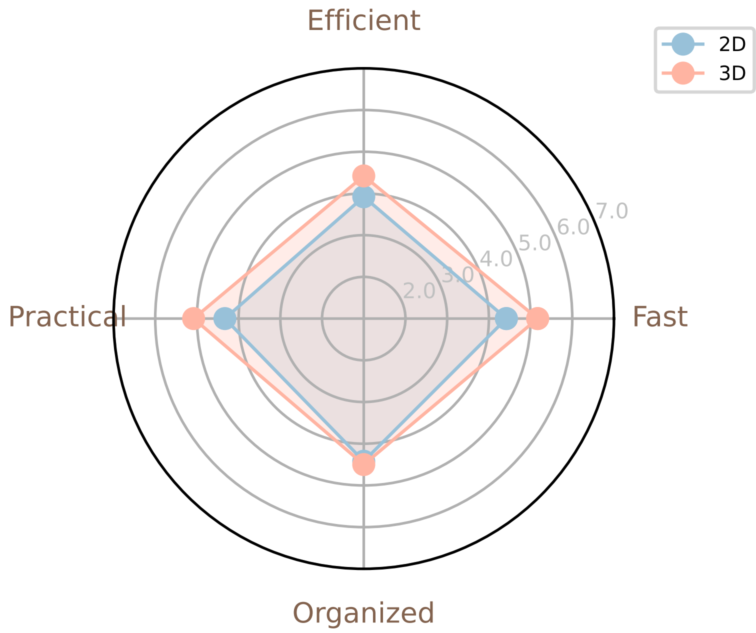

Figure 6.6: Mean scores for each item of the Efficiency of user experience.

The t-test returned a p-value of 0.43, indicating no statistical significance between the two environments, even though the 2D environment had a slightly higher mean score, 10.16 against 9.50.

To measure the general learning effect, we compared the scores between the first and the second attempt, which correspond to the first and third columns of the table. The t-test returned a p-value of 2.48E-14, with means of 9.83 and 15.70 for the first and the second attempt, respectively, indicating a very strong learning effect. Notice that the design of the experiment is symmetric, and thus, the result was not affected by the order in which the 2D and 3D games were played. Moreover, by comparing the second column of the table with the fourth, we notice that in the second attempt the standard deviations increased within all groups, indicating that the learning effect was not uniform across all users. This observation requires a larger-scale experiment to assess its statistical significance.

Finally, to specifically detect user familiarisation with the VR environment, as opposed to a general learning effect, we compared the second attempt scores between users who swapped environments between attempts (groups 2D-3D and 3D-2D) and those who did not (groups 2D-2D and 3D-3D). These scores correspond to the third

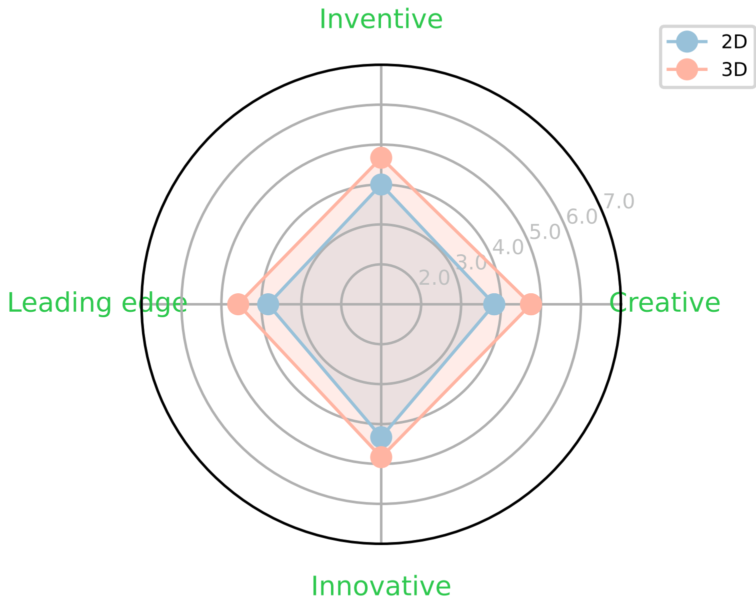

Figure 6.7: Mean scores for each item of the Novelty of user experience.

column of the table, and we compare rows 1 and 4, against rows 2 and 3. The corresponding means were 14.75 and 16.66, and a t-test p-value of 0.043 indicates that the detrimental effect of swapping VR environments was small but statistically significant.

6.5 Discussion

The findings of this study reveal several important insights into how the dimensionality of interface design—2D versus 3D—affects usability, user experience, and learning performance in gamified ASL learning environments.

From a theoretical perspective, our results contribute to ongoing research in the domains of human-computer interaction (HCI), educational psychology, and multimodal learning. Drawing on Mayer's Cognitive Theory of Multimedia Learning [102], the comparison of 2D and 3D interfaces helps highlight the role of spatial representation in cognitive processing. The improved ratings for the 3D interface—particularly in stimulation, attractiveness, and novelty—align with findings from prior studies suggesting that immersive environments can enhance engagement and motivation [11, 106]. However, the lower dependability scores for 3D and

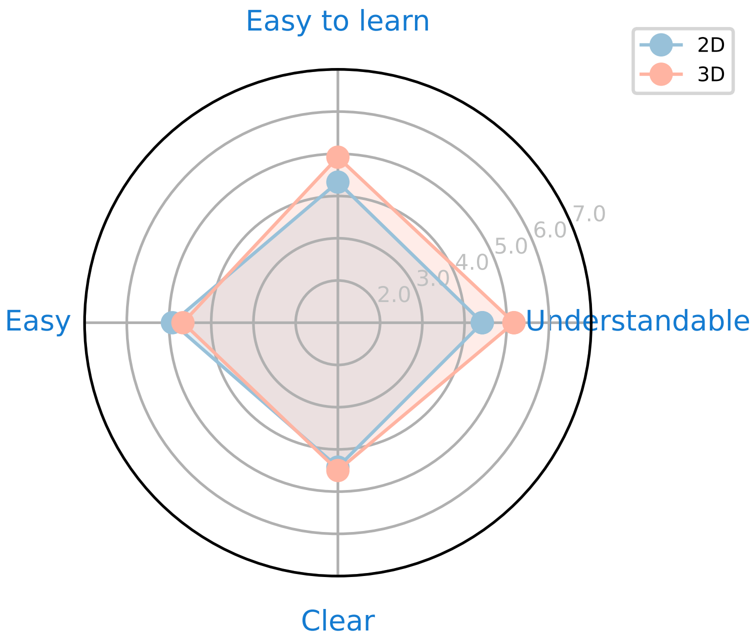

Figure 6.8: Mean scores for each item of the Perspicuity of user experience.

inconsistent responses in the 2D stimulation and perspicuity dimensions suggest that visual richness alone is not sufficient. Users must also be able to clearly interpret and interact with the environment for the experience to be meaningful.

Our study also informs the design of affective and hedonic aspects of learning environments. Following Schrepp *et al.*'s [78] framework, hedonic qualities like stimulation and novelty were more pronounced in 3D, suggesting that such environments may better satisfy users' emotional and experiential needs. However, the lack of a significant performance difference and the low reliability in some scales indicate that further exploration is required to fully understand the cognitive load introduced by added dimensionality.

Practically, the results suggest that while 3D environments are generally more engaging, they require careful design to ensure usability and clarity. Developers of ASL learning platforms may benefit from leveraging 3D interaction when aiming to enhance engagement, especially for novice users who may find traditional methods unappealing. However, the lower reliability in some user experience scales—particularly stimulation and perspicuity in 2D—also highlights the need for clearer user interface design, especially in lower-dimensional systems where visual

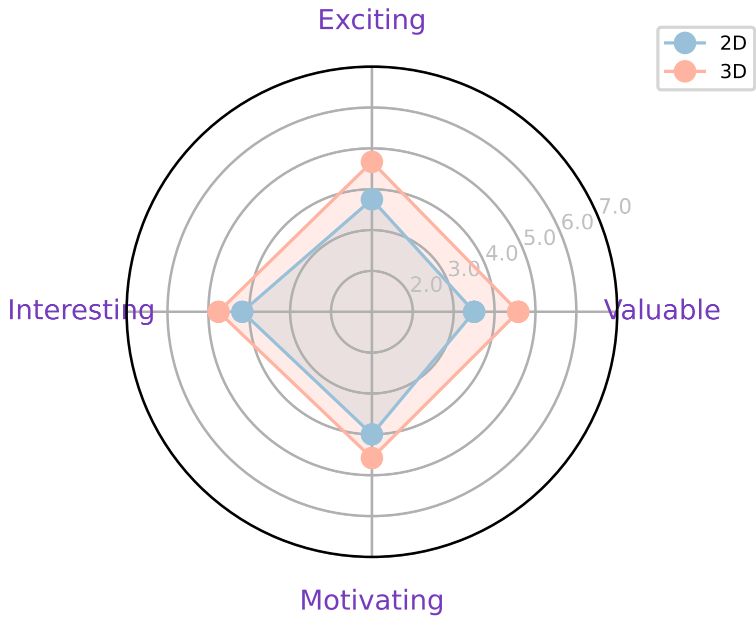

Figure 6.9: Mean scores for each item of the Stimulation of user experience.

cues may be limited.

The finding that both 2D and 3D interfaces received high SUS scores underscores that 2D is still a viable choice when system simplicity, accessibility, or hardware limitations are constraints. The learning curve associated with 3D VR and the hardware requirements may pose barriers in real-world deployments such as classrooms or public libraries. Thus, hybrid designs that adapt dimensionality based on user needs or learning context could offer an optimal solution.

Based on the findings, several design recommendations emerge:

- For 3D interfaces: Prioritise visual clarity and responsiveness to improve dependability and reduce possible confusion.

- For 2D interfaces: Introduce more interactive feedback mechanisms or visual effects to enhance stimulation and maintain engagement.

- In both: Consider implementing adaptive interfaces that adjust dimensionality or feedback based on learner behaviour or preferences.

It is important to note that the short interaction time may have amplified the effect of rapid user familiarisation. Even within a 30-second session, participants had the opportunity to adjust to the interface and mechanics, leading to potentially significant short-term learning effects. This is reflected in the overall improvement in scores during the second attempt, regardless of whether users switched environments. While part of this improvement may reflect improved ASL recognition, a portion is likely due to increased comfort with the interface or system feedback mechanisms. Future research should employ longer and repeated exposure designs to decouple usability and learning adaptation more clearly.

6.5.1 Limitations and Future Work

It is important to acknowledge that 2D and 3D games have their respective limitations. In the case of the 2D game, one area that warrants future work is the enhancement of the visual experience, employing advanced graphics techniques, which narrow the gap between 2D and 3D games and allow more detailed and visually appealing 2D environments. Additionally, further exploration of the gameplay mechanics and interactivity could lead to the creation of a more engaging and immersive 2D environment. Regarding the 3D game, a primary limitation relates to hardware requirements. By optimising the game engine and the graphics pipeline, we can further improve performance and enhance the user experience on the dependability scale.

Additionally, we acknowledge that this study has several limitations. The participant sample was small, and the session length was brief. As such, the findings should be considered exploratory and used to inform future, more comprehensive research. In future work, we recommend increasing the participant pool, diversifying user backgrounds, and employing within-subject comparisons to better isolate interface effects. Another limitation is the potential for rapid learning effects due to the short and simple interaction task. Although the short duration helped ensure standardised exposure, it also likely enabled users to adapt quickly, influencing both usability and user experience evaluations. Future studies should consider longer interaction periods and repeated trials to better understand how learning curves influence performance and perception, and to separate interface usability from familiarity effects.

Finally, while the platform and UEQ were reused from earlier chapters, this was an intentional choice to control for confounding variables and focus exclusively on the role of 2D versus 3D presentation. The research question, scope, and participant group were all unique to this chapter.

6.6 Conclusion

In conclusion, our study provides preliminary evidence of the potential of both 2D and 3D games in enhancing the user experience of learning ASL. The findings highlight the positive impact of 3D game environments on user engagement and overall experience, as evidenced by their higher ratings in attractiveness, usability, and efficiency, compared to 2D games. However, there is room for improvement in ensuring the dependability and clarity of 3D game environments. These results contribute to our understanding of the benefits of incorporating game-based approaches, particularly 3D environments, into ASL learning.

Future research can build upon these findings by delving deeper into the specific elements and design features that contribute to the positive user experience in both 2D and 3D games. Additionally, exploring strategies to enhance the dependability and clarity of 3D game environments can further optimise learning outcomes in ASL education. It would also be beneficial to increase the number and diversity of participants in future user studies to strengthen the generalisability of the findings.

CHAPTER 7

AI-Assisted Versus Non-AI Sign Language Learning: A Human-Centred Comparative Analysis

Chapters 3–6 examined core levers for entry-level ASL learning in isolation: VR vs. web (Ch. 3), gamification (Ch. 4), user-defined interaction (Ch. 5), and 2D vs. 3D representation (Ch. 6). Across these studies, engagement and UX often improved, but robust learning gains consistently hinged on the *clarity and specificity of corrective guidance* rather than immersion, points, or dimensionality alone. Single-session designs also limited what we could conclude about *retention*, *fluency*, and how performance evolves across time.

This chapter addresses those gaps with a head-to-head comparison of **AI-assisted diagnostic feedback** versus a **matched non-AI** pathway under identical content, timing, and scoring. We use repeated rounds to trace trajectories in *accuracy* and *response time*, examine item-level change, and pair quantitative results with brief learner reflections to explain *why* effects appear.

In short, Chapter 7 tests the mechanism implied by earlier chapters: whether *real-time, diagnostic feedback* turns practice into reliable gains more effectively than well-designed but non-adaptive materials.

7.1 Introduction

American Sign Language (ASL) proficiency is central to inclusion, accessibility, and equitable participation for Deaf communities and hearing interlocutors alike [109–111]. Over the last decade, digital learning platforms have broadened access to ASL instruction across formal and informal settings, enabling learners to engage anywhere and at flexible times [112–114]. These technologies promise scalability and reduced geographic and scheduling constraints, thereby addressing long-standing resource and instructor bottlenecks [112, 115, 116]. Yet the core design problem remains unsettled: how to balance accessibility and engagement with demonstrable learning effectiveness in a modality that combines handshape, orientation, location, and movement in space [117, 118].

Mainstream digital ASL tools typically deliver predefined curricula through videos and static multimedia with limited, delayed, or non-specific feedback [112, 119, 120]. While these systems have expanded reach, they often do not adapt to heterogeneous learner profiles (e.g., prior experience, motor control, visual–spatial reasoning) or to evolving performance trajectories within and across sessions [118, 121]. In particular, fixed-angle demonstrations can obscure subtle motion transitions, depth cues, and wrist orientations that are crucial for intelligibility and fluency, leaving learners to self-diagnose errors without expert guidance.

Advances in artificial intelligence (AI)—notably large language models (LLMs) and deep-learning pipelines for perception and feedback—have enabled systems that generate personalised explanations, adapt content to learner needs, and support dialogic practice [122–124]. In spoken-language education, such systems have improved feedback timeliness, reduced cognitive load, and supported stepwise mastery [125–129]. However, applications to ASL remain comparatively sparse and under-evaluated, despite promising technical advances in sign recognition and feedback generation [130–133]. Consequently, the empirical evidence base for AI-assisted ASL learning—particularly under controlled, repeated measures—remains thin.

Existing work in digital sign language education has largely emphasised aggregate performance outcomes (e.g., accuracy, completion) while paying limited attention to efficiency metrics such as response time and to item-level learning dynamics (e.g., how specific signs improve across sessions) [117, 134]. Without such measures, it is difficult to determine whether instructional gains reflect deeper procedural fluency or merely increased familiarity with task structure. Moreover, a purely quantitative focus can obscure the learner experience—how feedback is perceived, which design elements are motivating or frustrating, and what features learners request to advance their own practice.

To address these gaps, we conduct a controlled, repeated-measures comparison of two digital learning modes for ASL: a *non-AI* mode and an *AI-assisted* mode. We evaluate effects on both overall and item-level performance across initial, practice, and final assessments, with a joint focus on *accuracy* and *response time*. To contextualize behavioral outcomes, we complement the quantitative analyses with a qualitative examination of learner reflections, enabling us to characterize perceived clarity of instruction, engagement, and desired features in each mode.

The study is guided by the following question: *To what extent does AI-assisted learning—comprising real-time sign recognition and adaptive, model-generated feedback—compared with a non-AI mode, influence learner accuracy and response time across repeated and final ASL assessment sessions?*

This work advances ASL learning research in three ways that, to our knowledge, have not been jointly examined in prior ASL studies [120, 135]:

1. **Controlled, repeated-measures evaluation in ASL context.** We contrast an AI-assisted system (real-time recognition + adaptive feedback) with a non-AI mode across three rounds, tracking *both* accuracy and response time at each assessment. To our knowledge, this is the first controlled, repeated-measures comparison in the *ASL learning* context that models accuracy *and* timing across multiple sessions.

2. **Item-level (sign-level) learning dynamics.** We move beyond aggregate scores by analysing how specific signs evolve within and across rounds, separating accuracy gains from efficiency (response-time) changes - an aspect rarely reported in ASL learning evaluations.

3. **Learner-centered qualitative evidence** linking perceived clarity, interac-

tivity, and feedback usefulness to observed performance patterns; this includes structured thematic summaries that surface design requirements for future systems.

By combining repeated behavioral measures with qualitative insights, the study informs the design of next-generation digital ASL platforms that are both *effective* (improving accuracy and speed) and *usable* (delivering timely, actionable feedback and sustaining motivation). In doing so, it seeks to advance evidence-based practice for accessible language instruction and to narrow the evaluation gap between AI-driven and conventional approaches.

7.2 Context and Prior Work

7.2.1 Technology-Assisted Sign Language Learning

Sign language learning has been supported by diverse technologies, including static dictionaries, video-based repositories, and mobile applications [119, 136, 137]. Early systems emphasised access to sign banks, whereas recent approaches incorporate computer vision and machine learning for recognition and feedback [138–141]. Deep neural networks, particularly CNNs and RNNs, have achieved state-of-the-art recognition for static and continuous signing [142–145]. Tools such as MediaPipe have enabled real-time tracking of hand landmarks for accessible applications [146, 147].

Immersive learning via VR/AR has been explored to enhance motivation and authenticity in sign practice [148–150], though scalability and accessibility remain challenges. Mobile-based training platforms have demonstrated positive impacts on engagement and vocabulary retention [151, 152]. However, most systems present a predefined learning path with limited adaptability. Our study builds on these advances by embedding gesture recognition into an interactive interface, while extending prior work through a controlled comparison of AI-assisted versus non-AI predefined learning.

7.2.2 AI in Education and Adaptive Feedback

Intelligent tutoring systems (ITS) have long demonstrated the value of adaptive support across domains [153–155]. Adaptive feedback can enhance learner autonomy, improve error correction, and accelerate knowledge retention [156–158]. Natural language processing has broadened the scope of adaptive feedback in second-language learning, where chatbots and automated writing evaluators provide personalised scaffolds [159, 160].

The emergence of large language models (LLMs) has expanded possibilities for personalised, context-aware feedback [161–163]. Studies highlight their role in generating adaptive hints, conversational practice, and learning materials [161, 164]. Research in motor learning emphasises that immediate, adaptive error correction improves performance [165, 166], suggesting synergies with gesture-based education. Our AI-assisted interface operationalises these insights by delivering real-time recognition and adaptive feedback (with feedback text instantiated via a large language model), and we compare its effectiveness against a non-AI, predefined-materials condition.

7.2.3 Gamification and Engagement in Learning

Gamification is widely recognised for increasing motivation and retention in learning environments [167–169]. Serious games in language learning demonstrate benefits for vocabulary acquisition, engagement, and learner persistence [170–172]. Time-constrained challenges and repetition-based tasks foster recall and fluency [173,174].

Gesture-based games have been shown to reinforce kinesthetic learning, particularly in sign language and motor skill acquisition contexts [21,175,176]. Kinect-based and motion-controlled games highlight how embodied interaction supports memory consolidation [177]. In our study, the Whack-a-Mole–style evaluation extends this tradition, assessing immediate recall and fluency through playful but rigorous tasks.

7.2.4 Human-AI Collaboration and Trust

The effectiveness of AI in learning is mediated by user trust and acceptance. Human-centred AI frameworks emphasise augmentation of human capability rather than replacement [178]. Amershi et al. [179] propose guidelines for interactive AI design, highlighting transparency and error tolerance as critical for trust. In education, interpretability and learner control are essential for sustained engagement [180–182].

Studies have shown that poorly explained AI feedback may reduce learner agency [183]. Conversely, adaptive systems that communicate their reasoning foster learner confidence [184]. Our qualitative analysis complements this line of research by examining how learners perceived AI feedback in comparison to non-AI instruction, offering insights into acceptance and perceived reliability of LLM-generated guidance.

7.2.5 Summary

Prior research has established the feasibility of recognition-based systems for sign learning, the benefits of adaptive AI feedback, the motivational value of gamified practice, and the importance of trust in human-AI collaboration. However, few studies have experimentally compared AI-assisted versus non-AI pathways in ASL learning. By integrating gesture recognition, LLM-generated feedback, and game-based evaluation, our study advances understanding of how AI support influences skill-based, multimodal learning experiences.

7.3 Gesture Recognition Models

Two neural network classifiers were developed to recognise static ASL gestures: one for digits (*0–9*) and one for 10 common words (*airplane*, *father*, *hello*, *I'm*, *love*, *mother*, *ok*, *sorry*, *water*, *yes*), selected from Lean *et al.* work [185] as shown in Figure 7.1. Both models used MediaPipe to extract 21 landmark coordinates per frame, which were transformed into relative joint vectors and normalised before training. The architecture consisted of four fully connected layers (128, 64, 32, 16 neurons), each with ReLU activation and dropout ($p = 0.2$), followed by a softmax output

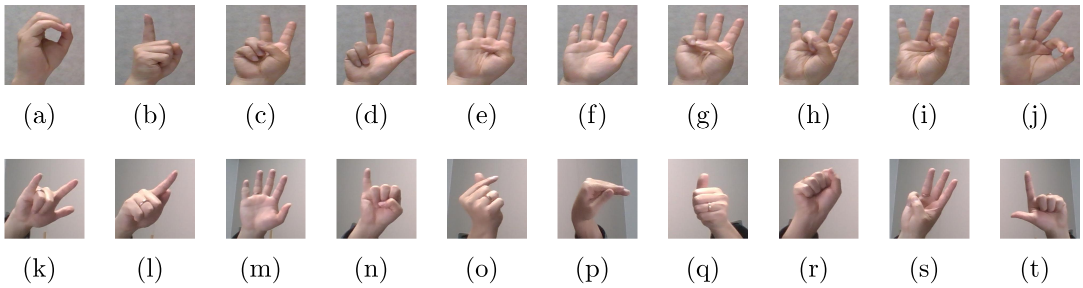

Figure 7.1: Illustration of the 20 American Sign Language (ASL) signs used in the study. Subfigures (a)–(j) show the digits 0–9, respectively: (a) 0, (b) 1, (c) 2, (d) 3, (e) 4, (f) 5, (g) 6, (h) 7, (i) 8, (j) 9. Subfigures (k)–(t) show the selected words: (k) *airplane*, (l) *father*, (m) *hello*, (n) *I'm*, (o) *love*, (p) *mother*, (q) *ok*, (r) *sorry*, (s) *water*, and (t) *yes*. These signs were used for assessment and learning across both the traditional and LLM-assisted conditions.

layer. Training employed the Adam optimiser, sparse categorical cross-entropy loss, and early stopping (patience = 10). Complete architecture/training details and full confusion matrices are provided in Appendix 10.2.

7.4 LLM feedback generation: differences and memory tips

To provide real-time, formative feedback, we integrated the hand-tracking recogniser with a large language model (LLM) assistant. The goal of this assistant was twofold: (i) to highlight the differences between the learner's current gesture and the intended ASL gesture, and (ii) to provide short, actionable memory tips to guide learners toward a correct performance on the next attempt. The full prompt template, JSON schema, and examples are included in Appendix 10.3.

7.5 Learning Interfaces

To investigate the role of AI assistance in American Sign Language (ASL) learning, we designed and implemented two platforms: a **Non-AI (traditional)** system and an **AI-assisted** system. Both platforms were integrated into the same experimental framework, comprising three components: assessment sessions, learning sessions, and interactive reinforcement games. The distinction between the two platforms lies in the type of instructional support and feedback provided during learning.

7.5.1 Assessment Session

. Both platforms employed an identical assessment interface for pre- and post-learning evaluations. As shown in Figure 7.2, participants completed a multiple-choice task in which they selected the correct ASL sign from four candidate images for each prompted item. Each correct response was awarded *one* point, and points were deducted for errors; the total score was the *number of correct responses*. Items were presented in randomised order, and the post-assessment used *equivalent but non-identical* items to the pre-assessment within the same round to minimise recall effects. For each item, response time (from prompt onset to selection) was recorded to enable efficiency analyses; unanswered items (if any) were scored as zero. This format ensured a controlled and consistent evaluation of learning outcomes across groups.

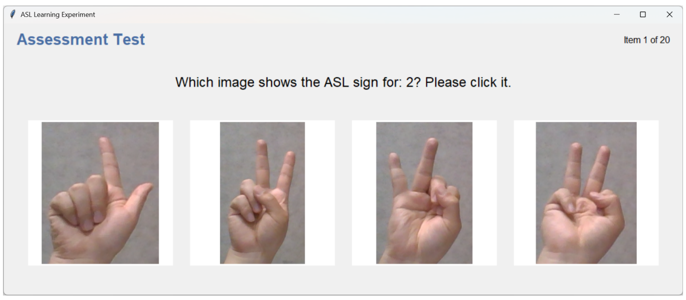

Figure 7.2: Assessment interface used in both platforms. Participants were asked to identify the correct ASL sign from four candidate images. This format was applied in both pre- and post-tests to measure learning outcomes consistently.

7.5.2 Non-AI Learning Session

In the Non-AI (traditional) learning session (Figure 7.3 left), participants studied a fixed sequence of static ASL sign images accompanied by text descriptions and optional sign demonstration videos. Each sign included a memory aid (e.g., associating the digit *0* with a closed circle) to support recall. The order of presentation was personalised: signs misclassified in the pre-assessment were prioritised before correctly identified ones.

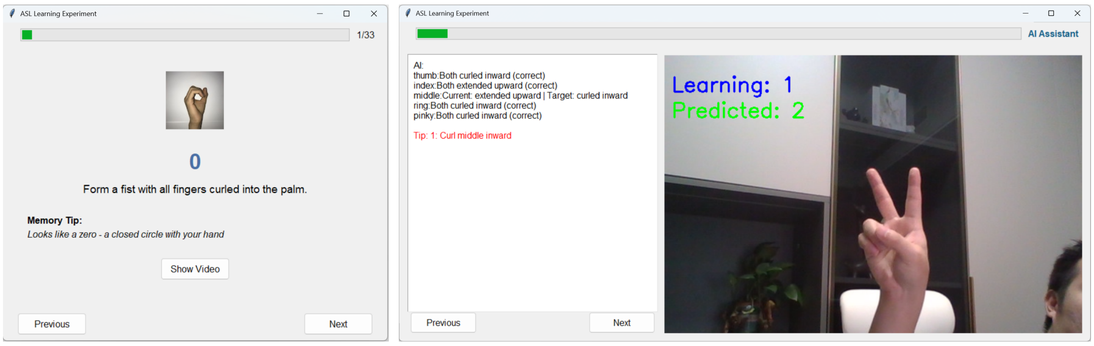

Figure 7.3: Learning sessions across the two platforms. (left) Non-AI interface: participants studied static sign images with descriptions and memory tips. (right) AI-assisted interface: participants received adaptive feedback from the recognition model and pre-generated LLM prompts.

7.5.3 AI-Assisted Learning Session

The AI-assisted learning session (Figure 7.3 right) incorporated a gesture recognition model and LLM-generated instructional prompts. MediaPipe extracted 21 hand landmarks per frame, which were classified using a neural network (Section 7.3). With over 95% recognition accuracy (Figures 10.1), the model enabled reliable real-time feedback. The AI assistant highlighted differences between the current gesture of the learner and the target gesture (e.g. 'curl middle inward'), and provided pre-generated memory tips from GPT-4o detailed in Section 7.4. This adaptive feedback allowed participants to iteratively correct their signs until successfully matched.

Figure 7.4 shows a representative participant engaging in the AI-assisted ASL learning experiment. The participant is seated in front of a laptop running the custom training software, with an external monitor displaying the development environment for system monitoring. During the task, the participant performs an ASL sign (here, the digit "6"), which is captured in real time by the laptop camera. The recognition system overlays the participant's live video feed with feedback on the detected gesture, indicating both the current input and whether it matches the expected target sign.

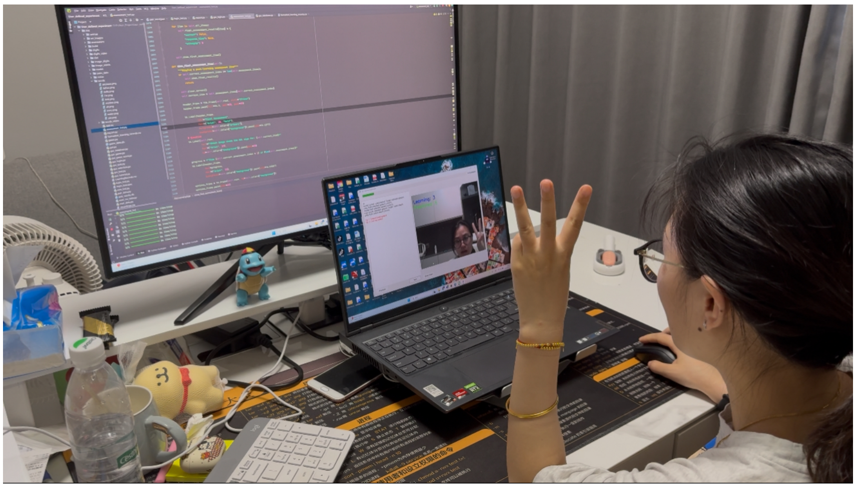

Figure 7.4: Example of a participant performing the ASL digit "6" during the learning experiment.

7.5.4 Reinforcement Game

Both platforms included a Whack-a-Mole style reinforcement game lasting 60 seconds (Figure 7.5). On each trial, a single on-screen stimulus appeared as a *target cue*, represented by a mouse icon corresponding to a specific ASL item. The game was implemented for both digits (0–9) and vocabulary words, using the same interface; only the labels differed.

Participants were instructed to produce the ASL sign that matched the target cue as quickly as possible. Participants first completed the digit-based game and then proceeded to the word-based version. The system monitored the pose of the hand in real time and registered a *hit* when the prompted sign was recognised above an acceptance threshold. Scores were accuracy-based: one point was awarded for each correctly recognised target; missing a target did not deduct points. Stimuli were presented in randomised order throughout the 60-second round, allowing multiple target opportunities; the interface and scoring logic were identical across the AI-assisted and non-AI conditions. The gesture recognition system provided real-time classification, updating scores accordingly. This gamified component was designed to strengthen retention and maintain learner engagement.

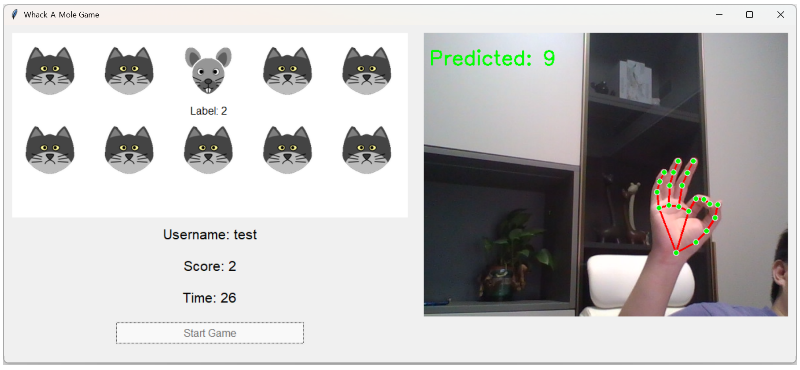

Figure 7.5: Whack-a-Mole style reinforcement game. Participants performed ASL gestures in response to target stimuli, with scores updated in real time based on recognition model predictions.

7.5.5 Summary

Both learning interfaces followed the same cycle: pre-assessment, learning session, post-assessment and reinforcement game. While the Non-AI system relied on predefined static materials, the AI-assisted system integrated gesture recognition and LLM-generated prompts to provide adaptive feedback. This design enabled a controlled comparison between non-AI and AI-supported ASL learning approaches. Design choices address single-session limitations noted in Chapters 3–6 and isolate diagnostic feedback as the active ingredient under matched content and timing.

7.6 User Study

7.6.1 Participants

A total of 46 participants completed the study, with 23 randomly assigned to the AI-assisted condition and 23 assigned to the non-AI condition. Participants were aged 22–38 years ($M = 28.5$, $SD = 4.2$), with a balanced gender distribution (24 male, 22 female).

All participants were Chinese nationals. The decision to recruit only Chinese participants was deliberate: ASL is not part of the local curriculum or daily communication environment in China, which ensured participants had no prior exposure to ASL. This homogeneity reduced potential confounds from pre-existing sign language knowledge, making observed performance differences more attributable to the experimental interventions.

Participants were recruited through on-site publicity at a public library. Recruitment posters with a QR code linked to the sign-up form were displayed at the library entrance. Adults from diverse professional fields volunteered and completed consent in accordance with Institutional Review Board (IRB) guidelines at Durham University. After eligibility screening, participants were *randomly assigned* in a 1:1 ratio to the AI-assisted or non-AI condition. Recruitment did *not* target any profession or demographic subgroup; backgrounds were heterogeneous, including Engineering, Design, IT, Teaching, Business, Arts, Management, and Education.

Table 7.1: Participant demographics by experimental condition

Condition	**N**	**Age Range**	**Gender (M/F)**	**Background Examples**
AI-assisted	23	22–37	12/11	Various professional fields
Non-AI	23	23–38	12/11	Various professional fields
Total	**46**	**22–38**	**24/22**	Various professional fields

7.6.2 Study Design and Procedure

The study followed a repeated-measures experimental design across three learning rounds. Each round consisted of four activities:

1. *Pre-assessment*: baseline ASL recognition task.
2. *Learning session*: AI-assisted (gesture recognition feedback) or non-AI (predefined static content).
3. *Post-assessment*: immediate evaluation using equivalent but non-identical items.
4. *Game session*: Whack-a-Mole–style evaluation where participants performed ASL signs in a 60s timed task.

The complete flow of the three rounds is illustrated in Figure 7.6.

Before starting the experiment, all participants were given a standardised introduction outlining the study procedure and ethical considerations. Participants

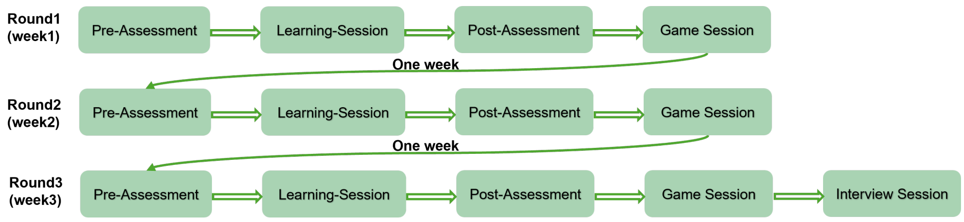

Figure 7.6: Three-round ASL learning and evaluation process used in the user study.

were informed that their task was to learn and practice a set of American Sign Language (ASL) digits and words across three rounds, with assessment and game-based activities after each learning session.

AI-assisted condition: Participants in the AI-assisted group were shown how to interact with the system interface, which included real-time hand tracking via webcam and immediate feedback from the gesture recognition model. They were instructed to perform each sign in front of the webcam, observe the system's feedback (e.g., correctness indicators, error highlighting), and repeat gestures until the system confirmed accurate recognition. A brief demonstration was provided to ensure they understood how to interpret the AI feedback.

Non-AI condition: Participants in the non-AI group received instructions on how to navigate the predefined learning interface. Instead of AI feedback, they were guided through structured instructional materials, such as static images, textual descriptions, and example videos of each sign. They were instructed to imitate the provided demonstrations and self-check their accuracy using the comparison resources (e.g., video replay, step-by-step illustrations).

Both groups were reminded that they would complete three activities in each round: (1) a pre-session assessment, (2) a learning session, and (3) a post-session assessment, followed by a game-based evaluation. Participants were explicitly told that the study compared two different learning approaches, but the exact hypotheses were not disclosed to minimise bias. To ensure comprehension, participants practised one sample sign under supervision before beginning the formal study rounds.

Each participant completed all three rounds over a two-week period. The assessments required selecting the correct ASL sign from four alternatives, scored using a predefined rubric. The learning sessions differed by condition: participants in the AI-assisted platform received personalised, real-time feedback generated by the gesture recognition model, whereas participants in the non-AI platform accessed predefined learning materials without adaptive feedback. The post-assessment was structurally identical to the pre-assessment. Finally, the Whack-a-Mole game reinforced recall in an engaging, time-constrained format.

7.6.3 Interview Process

Following completion of the third round's game session, participants engaged in a semi-structured interview lasting 10-20 minutes. The interview was organised into four thematic sections: *learning experience*, *general engagement*, *assessment & game reflection*, and *hypothetical comparisons*, as shown in Table 10.1. The complete semi-structured interview guide appears in Appendix 10.4.

7.7 Results

7.7.1 Quantitative analysis

Pre- and Post-Session Assessment Performance

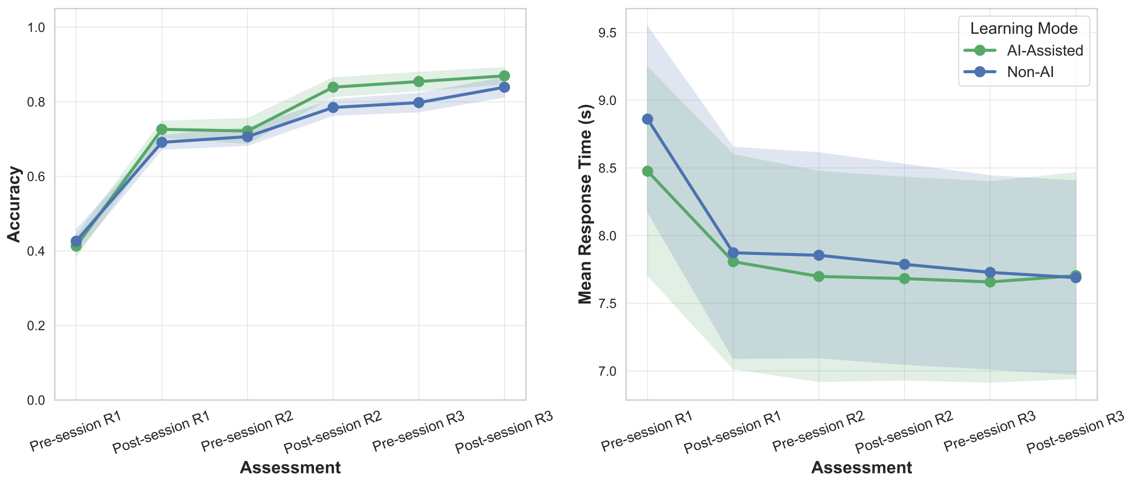

Figure 7.7: Mean accuracy (left) and mean response time (right) across pre- and post-session assessments in three consecutive learning rounds for AI-assisted and Non-AI participants. Error bands represent the standard error of the mean.

Figure 7.7 depicts participants' mean *accuracy* (top panel) and *response time* (bottom panel) in pre- and post-session assessments across three consecutive learning rounds, comparing the AI-assisted and Non-AI conditions. The horizontal axis represents the assessment phase in chronological order (*Pre 1*, *Post 1*, *Pre 2*, *Post 2*, *Pre 3*, *Post 3*). The dual perspective allows evaluation of not only participants' accuracy in producing correct signs, but also the efficiency of retrieval and execution, captured through response time.

Across all rounds, both groups demonstrated substantial improvement from pre- to post-session assessments, reflecting effective learning within each round. However, the AI-assisted group consistently achieved larger accuracy gains. In Round 1, the two groups started at comparable pre-session accuracy levels, but AI-assisted participants showed sharper post-session increases, pointing to the immediate effectiveness of adaptive feedback. This advantage persisted into Rounds 2 and 3: the AI-assisted group not only maintained higher post-session scores but also entered each new round with elevated pre-session baselines relative to the Non-AI group. This pattern indicates that AI-supported learning facilitated better retention between sessions and cumulative consolidation across rounds. In contrast, both groups improved across rounds; although the AI-assisted group often showed higher means, the between-group gap did not reliably widen by the final round.

A complementary trend emerged in response times. Both groups exhibited reductions across rounds, consistent with automatization of sign recall and production. However, the AI-assisted group consistently reached lower (faster) mean response times in both pre-session and post-session evaluations. In Round 1, post-session gains were modestly faster in the AI-assisted condition, but by Rounds 2 and 3 the difference widened, with AI learners performing more efficiently from the outset. The narrowing variability in later rounds suggests that practice reduced individual differences, though the AI-assisted group retained an edge in both speed and stability. Notably, the simultaneous increase in accuracy and decrease in response time reflects not a speed–accuracy trade-off but a genuine efficiency gain, with AI feedback promoting both correctness and fluency of execution.

Together, the accuracy and response time trajectories highlight the dual benefits of AI assistance. Non-AI learners demonstrated improvement, but their progress was characterised by slower, less stable gains and greater reliance on repeated exposure. In contrast, AI-assisted participants achieved steeper within-session improvements, higher cross-session retention, and faster response times, suggesting a more efficient internalisation of the target signs. These converging patterns underscore that AI-supported feedback not only accelerated acquisition but also optimised the quality and efficiency of learning, leading to durable advantages over time.

Table 7.2: Mixed-effects model predicting ASL assessment scores. Reference category: Post-session, Round 1, Non-AI. Negative coefficients for *Pre-session* indicate lower scores than the reference; positive coefficients for *Round* terms indicate higher scores than Round 1 in the same phase. Interaction terms show differences in phase effects across rounds and AI conditions.

Predictor	**Coef.**	**SE**	z	p	**95% CI**
Intercept (Post, Round 1, Non-AI)	13.826	0.497	27.82	$< .001$	[12.852, 14.800]
Pre-session (vs. Post-session)	-5.304	0.372	-14.24	$< .001$	[-6.034, -4.574]
Round 2 (Post vs. Round 1 Post)	1.870	0.372	5.02	$< .001$	[1.140, 2.599]
Round 3 (Post vs. Round 1 Post)	2.957	0.372	7.94	$< .001$	[2.227, 3.686]
AI-assisted (Post, Round 1 vs. Non-AI Post, Round 1)	0.696	0.703	0.99	0.322	[-0.682, 2.073]
Pre × Round 2 (vs. Round 1 Pre–Post gap)	3.739	0.527	7.10	$< .001$	[2.707, 4.771]
Pre × Round 3 (vs. Round 1 Pre–Post gap)	4.478	0.527	8.50	$< .001$	[3.446, 5.511]
Pre × AI-assisted (vs. Non-AI Pre–Post gap in Round 1)	-0.957	0.527	-1.82	0.069	[-1.989, 0.076]
Round 2 × AI-assisted (Post vs. Non-AI Post)	0.391	0.527	0.74	0.458	[-0.641, 1.424]
Round 3 × AI-assisted (Post vs. Non-AI Post)	-0.087	0.527	-0.17	0.869	[-1.119, 0.945]
Pre × Round 2 × AI-assisted (gap change vs. Non-AI Round 2)	0.174	0.745	0.23	0.815	[-1.286, 1.634]
Pre × Round 3 × AI-assisted (gap change vs. Non-AI Round 3)	1.478	0.745	1.99	0.047	[0.018, 2.938]
Random effects: Participant intercept variance = 4.087					

Mixed-effects model We analysed assessment scores with linear mixed-effects models (LMMs) appropriate for repeated measures with observations nested within participants [186, 187]. Fixed effects included *Phase* (Pre vs. Post), *Round* (1–3), *AI Condition* (AI-assisted vs. non-AI), and their interactions. To account for between-participant heterogeneity, we specified random intercepts for participants; where justified by model fit and convergence, we considered adding by-participant random slopes for Phase [188]. Models were estimated by (restricted) maximum likelihood; inference used Wald tests on fixed effects with two-sided $\alpha = .05$, and 95% confidence intervals are reported alongside coefficient estimates.

Mixed-effects models are well suited to longitudinal/within-subject designs because they (i) model the correlation of repeated observations within individuals, (ii) retain all available data without listwise aggregation, (iii) accommodate unequal trial counts or occasional missingness, and (iv) allow individual differences in baselines (intercepts) and, when supported, in change over time (slopes) [186, 189]. This makes them preferable to repeated-measures ANOVA for our setting with multiple rounds and phase contrasts, and aligns with best practice studies in psycholinguistics, HCI, and education using similar designs [188, 190].

The mixed-effects model (Table 7.2) confirmed several key trends:

- **Phase effect (reference = Round 1).** In Round 1, pre-session scores were, on average, 5.30 points lower than post-session ($p < .001$). This pre–post gap was markedly smaller in later rounds, as indicated by positive Phase×Round interactions: the gap was reduced by 3.74 points in Round 2 and by 4.48 points in Round 3.

- **Round effects:** Post-session scores were significantly higher in Round 2 (+1.87, $p < .001$) and Round 3 (+2.96, $p < .001$) compared to Round 1, suggesting cumulative skill gains over time.

- **AI main effect:** When collapsing across all phases and rounds, the AI-assisted group's post-session scores were not significantly different from the Non-AI group's (+0.70, $p = .322$). Group means were similar at baseline (Pre–1). This baseline similarity does not imply equivalence across later phases or rounds; condition differences are therefore interpreted within each round and phase.

- **Phase × Round interactions:** The pre–post improvement was smaller in Round 1 than in later rounds, with significant increases in Round 2 (+3.74, $p < .001$) and Round 3 (+4.48, $p < .001$), indicating that as participants became more experienced, they made larger within-session gains.

- **Phase × AI (Round 1).** In Round 1, the AI-assisted group showed a marginally larger pre–post gain ($\Delta \approx 0.96$ points, $p = .069$), indicating an *early-round advantage*; differences were smaller in later rounds (cf. Figure 7.8, Table 7.2).

- **Phase × Round × AI.** In Round 3, the AI-assisted group's pre–post gain was statistically larger than the Non-AI group's ($\Delta = 1.48$, $p = .047$). This *round-specific* effect does not imply a steadily increasing between-group gap; absolute differences in final scores did not consistently widen across rounds and may reflect ceiling constraints (see Figure 7.8 and Discussion).

In summary, both groups improved within and across rounds, but the AI-assisted group achieved consistently greater post-session performance and, by Round 3, significantly outpaced the Non-AI group in within-session improvement. This pattern indicates that AI support may be especially valuable in sustaining learning momentum and amplifying gains as learners approach higher levels of proficiency.

Improvement Across Rounds

Figure 7.8 presents the distribution of participants' improvement scores (*Post-session* minus *Pre-session*) in each round, separated by *AI-assisted* and *Non-AI* conditions. Across all rounds, the AI-assisted group tended to achieve greater within-session gains in accuracy, although the size and even the direction of this difference varied over time.

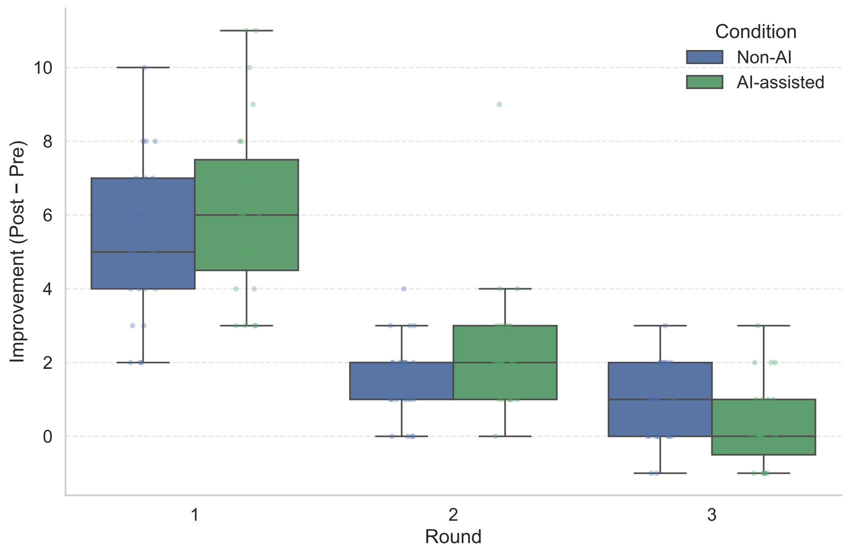

Figure 7.8: Distribution of improvement scores (Post–Pre) across rounds for AI-assisted and Non-AI groups. Error bars represent interquartile ranges; circles denote outliers.

Welch's independent t-tests for each round revealed that none of the between-group differences reached statistical significance individually, though the effect in

Round 2 approached significance ($t = 1.807,\ p = 0.079$). The smallest difference appeared in Round 1 ($t = 1.397,\ p = 0.170$), while in Round 3 the direction reversed slightly, with the Non-AI group showing a marginally higher median improvement ($t = -1.595,\ p = 0.118$). Per-round variability indicates round-specific differences rather than a monotonic trend; we therefore refrain from ranking stages and defer interpretation to the Discussion (Section 7.8).

A two-way mixed ANOVA, with *AI Condition* as a between-subjects factor and *Round* as a within-subjects factor, showed a significant main effect of *AI Condition* ($F(1, 45) = 4.014,\ p = 0.048$), indicating that AI-assisted learners achieved greater overall pre–post gains when averaging across all rounds. There was also a robust main effect of *Round* ($F(2, 90) = 157.630,\ p < 0.001$), with improvement scores declining steadily from Round 1 to Round 3. This decline likely reflects a ceiling effect: as participants' baseline skills rose, there was less room for large within-session gains. Importantly, the *AI Condition* × *Round* interaction was significant ($F(2, 90) = 3.524,\ p = 0.034$), showing that the AI advantage was not constant—its impact shifted depending on the stage of learning.

A mixed-effects model with random intercepts for participants largely corroborated the ANOVA findings. The main effect of *AI Condition* was marginally non-significant ($\beta = -0.957,\ p = 0.053$ for Non-AI relative to AI-assisted), again suggesting a general trend toward greater improvement for AI-assisted learners. Relative to Round 1, within-session gains decreased by 3.91 points in Round 2 ($p < 0.001$) and by 5.96 points in Round 3 ($p < 0.001$). The model also identified a significant interaction for Round 3 in the Non-AI group ($\beta = 1.478,\ p = 0.013$), indicating that the late-stage drop in improvement was less pronounced for Non-AI learners than for AI-assisted learners.

Taken together, these results suggest that AI-assisted feedback provided a notable boost in early- and mid-stage learning, enabling larger within-session gains when learners still had substantial room for improvement. However, as proficiency increased and error rates dropped, the relative advantage of AI narrowed, and by Round 3, Non-AI participants retained more consistent improvement. This shift may indicate that the Non-AI group's reliance on deliberate, self-paced correction fostered steadier incremental progress in later stages, while AI learners—having already captured many quick wins earlier—experienced diminishing returns on additional feedback. The interaction pattern supports a nuanced view: AI support accelerates early skill acquisition but may converge with, or even be matched by, strong self-regulated practice in advanced phases.

Learning Session Analysis

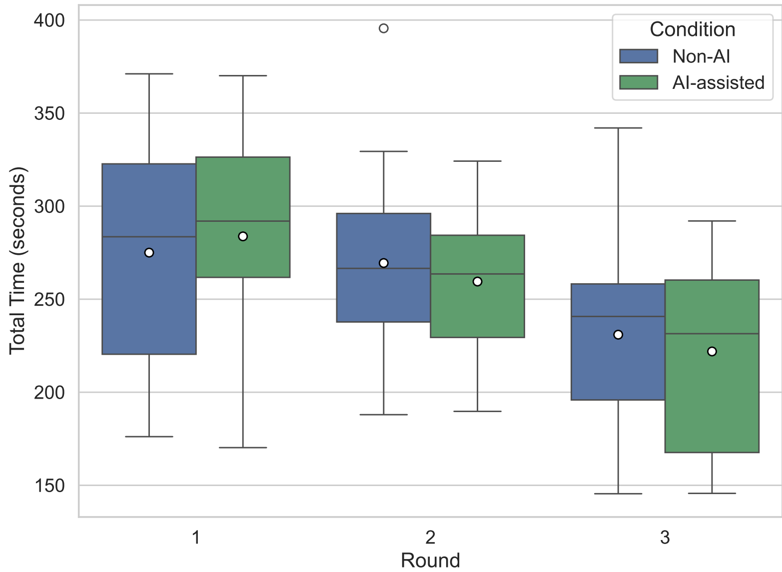

Figure 7.9: Distribution of total learning time across rounds for *AI-assisted* and *Non-AI* participants. Boxplots show the median (horizontal line), interquartile range (box), and range excluding outliers (whiskers), with outliers plotted as individual points.

Figure 7.9 shows the distribution of total time participants spent in each learning round, from pre-session to post-session completion, for both the AI-assisted and Non-AI groups. Median learning times declined markedly from Round 1 to Round 3 in both groups, reflecting increased familiarity with the task, more efficient recall. In the AI-assisted condition, shorter times likely reflect faster uptake of corrective feedback; in the non-AI condition, reductions are attributable to streamlined navigation and selective replay, not in-session practice of signing.

A two-way mixed ANOVA with *AI Condition* (AI-assisted vs. Non-AI) as a between-subjects factor and *Round* (1–3) as a within-subjects factor revealed a significant main effect of *Round*, $F(2, 88) = 59.66$, $p < 0.001$, confirming that learning times decreased significantly in later rounds. There was no significant main effect of *AI Condition*, $F(1, 45) = 0.36,\ p = 0.545$, indicating that, on average, the two groups spent similar amounts of time learning. The *AI Condition* $\times$ *Round* interaction was not significant ($p = 0.115$), suggesting that both groups showed broadly similar patterns of time reduction across rounds.

For robustness, we fitted a mixed-effects model with total learning time as the dependent variable, *AI Condition* and *Round* as fixed effects, and random intercepts for participants. The model specification and results are presented in Table 7.3, where the reference category is the AI-assisted group in Round 1.

Table 7.3: Mixed-effects model predicting total learning time (in seconds). Reference category: AI-assisted group, Round 1. Negative coefficients indicate less time than the reference; positive coefficients indicate more time. Interaction terms show how differences between AI and Non-AI groups vary across rounds.

Predictor	**Coef.**	**SE**	z	p	**95% CI**
Intercept (AI-assisted, Round 1)	283.766	10.220	27.77	< 0.001	[263.735, 303.797]
Non-AI (vs. AI-assisted, Round 1)	-8.738	14.454	-0.61	0.545	[-37.066, 19.591]
Round 2 (AI-assisted vs. Round 1)	-24.279	6.916	-3.51	< 0.001	[-37.833, -10.724]
Round 3 (AI-assisted vs. Round 1)	-61.844	6.916	-8.94	< 0.001	[-75.399, -48.290]
Non-AI $\times$ Round 2 (vs. AI-assisted Round 2)	18.679	9.780	1.91	0.056	[-0.489, 37.848]
Non-AI $\times$ Round 3 (vs. AI-assisted Round 3)	17.757	9.780	1.82	0.069	[-1.411, 36.926]
Random intercept variance (Participant)	1852.462	22.168			

The mixed-effects model confirmed the ANOVA's findings:

- **Main effect of round:** AI-assisted participants reduced their learning time by an estimated 24.28 seconds from Round 1 to Round 2, and by 61.84 seconds from Round 1 to Round 3 ($p < 0.001$ for both). This pattern indicates substantial efficiency gains over repeated practice cycles.

- **Main effect of AI Condition:** The Non-AI group's total time in Round 1 was slightly lower than that of the AI-assisted group, but the difference was not significant ($p = 0.545$).

- **Interaction effects:** The Non-AI group showed a tendency to spend more time than the AI-assisted group in Rounds 2 and 3 (about +18 seconds in each case), with p-values near the conventional threshold for significance (0.056 and

0.069). While not statistically reliable, these patterns suggest that the Non-AI group may have required slightly longer to complete later sessions.

Both groups became faster at completing the learning sessions over time, reflecting procedural fluency gains as sign recognition and production became more automatic. The absence of a significant overall difference between AI-assisted and Non-AI participants suggests that AI feedback did not reduce the total time needed to complete a round—likely because participants in the AI-assisted group spent additional moments responding to feedback and making corrections in real time. The marginal trend toward longer times for the Non-AI group in later rounds could reflect their need for more self-directed checking and repetition before moving on, whereas AI-assisted learners could rely on the system to flag errors and confirm correctness. In this sense, efficiency gains appear to be driven primarily by practice effects common to both groups, with only minor time savings potentially attributable to AI assistance in the later stages.

Correlation between Learning Time and Accuracy

To investigate the relationship between the total time spent in the learning session and the final assessment accuracy, Pearson's correlation coefficients were calculated separately for each round and AI condition (Figure 7.10).

In **Round 1**, a significant positive correlation was observed for the Non-AI group ($r = 0.49$, $p = 0.017$), suggesting that participants who spent more time learning tended to achieve higher final accuracy scores. In contrast, the AI-assisted group showed a negligible and non-significant correlation ($r = -0.08$, $p = 0.710$), indicating no clear relationship between time spent and performance in this round.

In **Round 2**, the correlations for both groups were positive but non-significant: Non-AI ($r = 0.24$, $p = 0.263$) and AI-assisted ($r = 0.28$, $p = 0.202$). This suggests that, while there was a weak tendency for longer learning times to be associated with higher accuracy, the effect was not statistically reliable.

In **Round 3**, the Non-AI group again showed a moderate positive correlation ($r = 0.39$, $p = 0.065$), approaching significance, whereas the AI-assisted group maintained a weak and non-significant association ($r = 0.11$, $p = 0.610$).

Overall, the results suggest that the relationship between time spent learning and subsequent performance is more consistent in the Non-AI group, particularly in the early round, whereas AI-assisted participants did not show a clear link between time investment and accuracy. The pattern suggests more efficient time-to-information in the AI-assisted condition—minutes are focused on corrective actions rather than undirected review—whereas non-AI practice shows diminishing returns. Participant reports align with this interpretation (Section 7.7.2); mechanisms are detailed in the Discussion (Section 7.8).

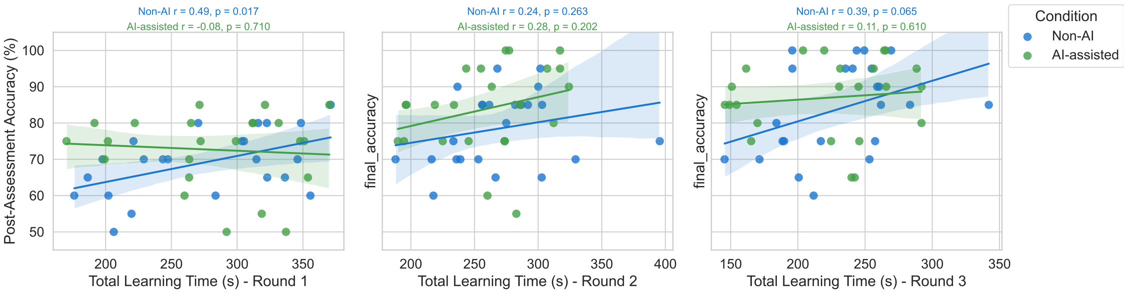

Figure 7.10: Scatterplots showing the correlation between total learning time and final assessment accuracy for each round, separated by AI condition. Lines represent the best-fitting regression for each group, with Pearson's r and p-values reported in the plot.

Game Performance

Figure 7.11 presents the distribution of Whack-a-Mole game scores for AI-assisted and Non-AI participants across three rounds. Several clear patterns emerge from the boxplot.

In **Round 1**, both groups displayed relatively wide score distributions, indicating varied initial performance levels. The AI-assisted group's median score was moderately higher than that of the Non-AI group, and the upper quartile extended further upward, suggesting that the top performers in the AI-assisted condition achieved noticeably higher scores. The presence of fewer low outliers in the AI-assisted group points to greater consistency, with fewer participants struggling substantially in the first exposure.

By **Round 2**, median scores rose for both groups, accompanied by narrower interquartile ranges. This reduction in variability suggests a consolidation effect,

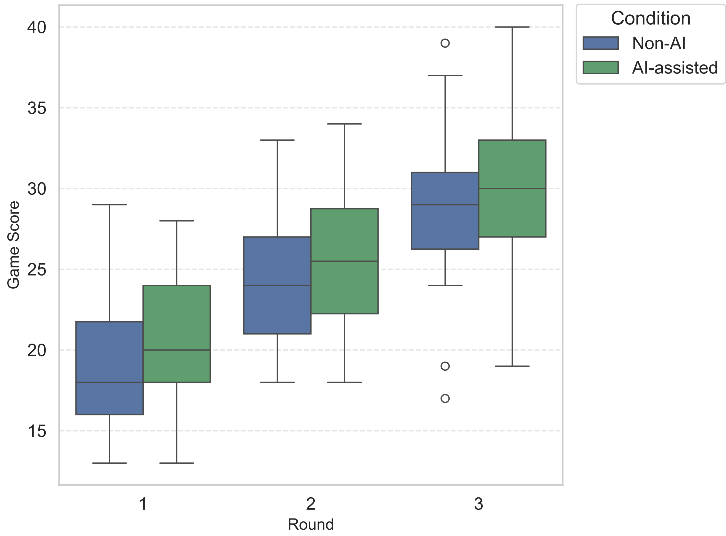

Figure 7.11: Boxplot of Whack-a-Mole game scores across three rounds, comparing AI-assisted and Non-AI participants. The central line represents the median, the box shows the interquartile range (IQR), and whiskers extend to 1.5 × IQR. Points outside the whiskers indicate outliers.

where repeated gameplay led to more uniform performance across participants. The AI-assisted group maintained a higher median and upper quartile, and again exhibited fewer extreme low scores, reinforcing the impression of more stable and reliable performance.

In **Round 3**, the median scores for both groups increased further, indicating continued skill development. Variability was smallest in this round, especially for the AI-assisted group, whose interquartile range compressed around higher score values. The Non-AI group also demonstrated strong improvement, but their distribution retained slightly more spread, hinting at greater individual differences in end-stage performance. By Round 3, outliers (circles) are visible in the non-AI condition, while in the AI-assisted condition the whiskers (and, where visible, the IQR) are wider than in earlier rounds, indicating greater dispersion despite comparable or higher medians.

Taken together, the boxplots show that while both groups improved with practice, the AI-assisted group consistently achieved higher central scores, tighter clustering, and fewer extreme lows across rounds. This pattern points to a potential advantage in both efficiency and consistency, with AI feedback possibly accelerating convergence toward high, stable performance levels.

7.7.2 Qualitative analysis

Hybrid qualitative pipeline (single-analyst, assistive)

Dataset size and structure. All participants ($N = 46$) completed a post-study interview organised into **4 thematic blocks** (*Learning Experience*, *General Engagement*, *Assessment & Game Reflection*, *Hypothetical Comparison & Open Feedback*), each with **2-5 short answers**, yielding up to 506 brief responses (median one–two sentences). Interview notes were captured verbatim and immediately *member-checked* by reading them back to participants.

Methods rationale. Because the corpus contains *many very short answers*, beginning from purely manual open coding risks inconsistent initial buckets and analyst drift across passes. We therefore used **assistive** methods to *surface* candidate structure reproducibly and to *flag* potential misfits—*not* to automate labelling or to increase apparent methodological complexity.

Pipeline.

1. **Pre-processing.** Lowercasing, punctuation removal, stop-word filtering; domain tokens (e.g., *ok*, *hello*, *water*) preserved.

2. **Topic surfacing (TF–IDF + NMF).** For each block, we built TF–IDF matrices with unigrams/bigrams (min_df/max_df tuned per block) and applied Non-negative Matrix Factorisation (NMF). We explored $k \in \{2, \dots, 6\}$ and selected k by balancing topic *coherence* and *diversity* via a simple harmonic score (see Fig. 7.13). The output at this stage is a set of *candidate clusters* only.

3. **LLM label suggestions.** For each candidate cluster, we provided the top TF–IDF terms and a small de-identified exemplar snippet to an LLM (gpt-4o-mini) to *suggest* a short, human-readable topic label. These suggestions were used as drafting aids *only*.

4. **Human adjudication and final coding. All** qualitative adjudication was performed by **the author** (single analyst). The author (i) re-read every re-

sponse in its original interview context, (ii) merged or split candidate clusters where automatic groupings or LLM label suggestions did not match the data or interview guide, and (iii) **wrote the final theme labels and assigned codes response-by-response**. LLM-suggested labels were accepted *only if* they accurately reflected the data after re-reading; otherwise they were edited or discarded. *No embedding or similarity methods (e.g., SBERT) were used for coding or theme assignment.* As detailed below, SBERT was employed strictly *after* coding to retrieve a single representative quote per theme (exemplar selection), not to influence labels or codes.

5. **SBERT-guided exemplar selection (sampling, not coding).** To anchor each theme with a representative piece of evidence, I used semantic similarity strictly as a *retrieval mechanism* for one exemplar quote per theme. All responses and the *human-written* theme descriptions were embedded with SBERT (`all-mpnet-base-v2`) and cosine similarity was computed [191].

   Specifically, let $\mathbf{r}_i \in \mathbb{R}^d$ denote the SBERT embedding of response i and $\mathbf{t}_j \in \mathbb{R}^d$ the embedding of topic j. After ℓ_2-normalization,

$$\hat{\mathbf{r}}_i = \frac{\mathbf{r}_i}{\|\mathbf{r}_i\|_2}, \quad \hat{\mathbf{t}}_j = \frac{\mathbf{t}_j}{\|\mathbf{t}_j\|_2} \tag{7.1}$$

   we computed the response–topic similarity matrix $S \in \mathbb{R}^{N \times K}$, with elements in [-1,1], via cosine similarity,

$$S_{ij} = (\hat{\mathbf{r}}_i)^\top \hat{\mathbf{t}}_j. \tag{7.2}$$

   For each theme, the *single highest-similarity* response was selected as the exemplar and then quoted.

Integrity safeguards and auditability. Credibility was supported via (i) *member-checking* at capture; (ii) a simple, reproducible surfacing step (TF–IDF/NMF); (iii) an *LLM-suggestions-only* label aid with human verification; (iv) a *flag-only* SBERT screen to catch off-theme items; and (v) an *audit log* of merges/splits, label edits, and codebook changes. Given the single-analyst design, we prioritised procedural transparency over inter-rater statistics.

Scope and limitations. This assistive pipeline suits corpora of numerous short utterances; it is not a substitute for hermeneutic analysis of long narratives. AI components (NMF, LLM label suggestions, SBERT flags) were introduced only where warranted by corpus characteristics; **all** reported themes and labels are the result of *human* interpretation.

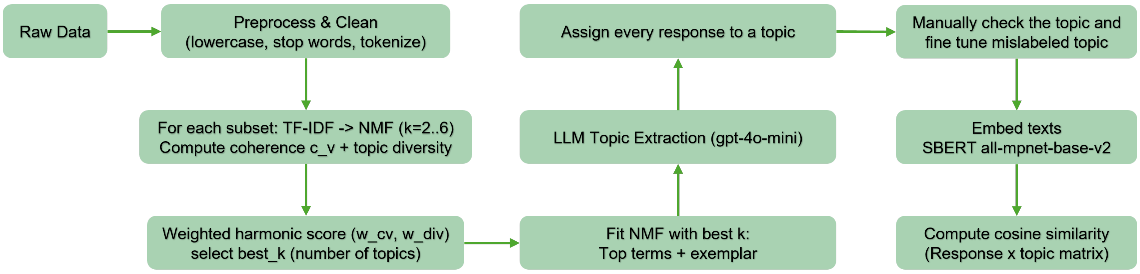

Figure 7.12: Assistive qualitative workflow. TF–IDF + NMF *surface* candidate clusters for short answers; the author conducts all adjudication and final coding. SBERT similarity is used only as a *diagnostic flag* to prompt re-inspection; it does not assign labels.

Selecting the Number of Topics

Figure 7.13 reports a weighted harmonic score that jointly captures *topic coherence* and *topic diversity* for candidate topic counts $k \in \{2, \ldots, 6\}$, estimated separately for the Non-AI and AI-Assisted groups within each section. Across three weighting schemes for coherence vs. diversity (0.5/0.5, 0.6/0.4, 0.7/0.3), we selected k by maximising the weighted harmonic score

$$H_w(k) = \left(\tfrac{w}{C(k)} + \tfrac{1-w}{D(k)}\right)^{-1} \tag{7.3}$$

where $C(k)$ is c_v coherence and $D(k)$ is topic diversity computed from the top-8 terms per topic (average pairwise Jaccard distance/proportion unique); higher H_w indicates a better balance (Fig. 7.13).

In all sections,the k that maximised the harmonic score did not depend on the weighting scheme (Fig. 7.13):

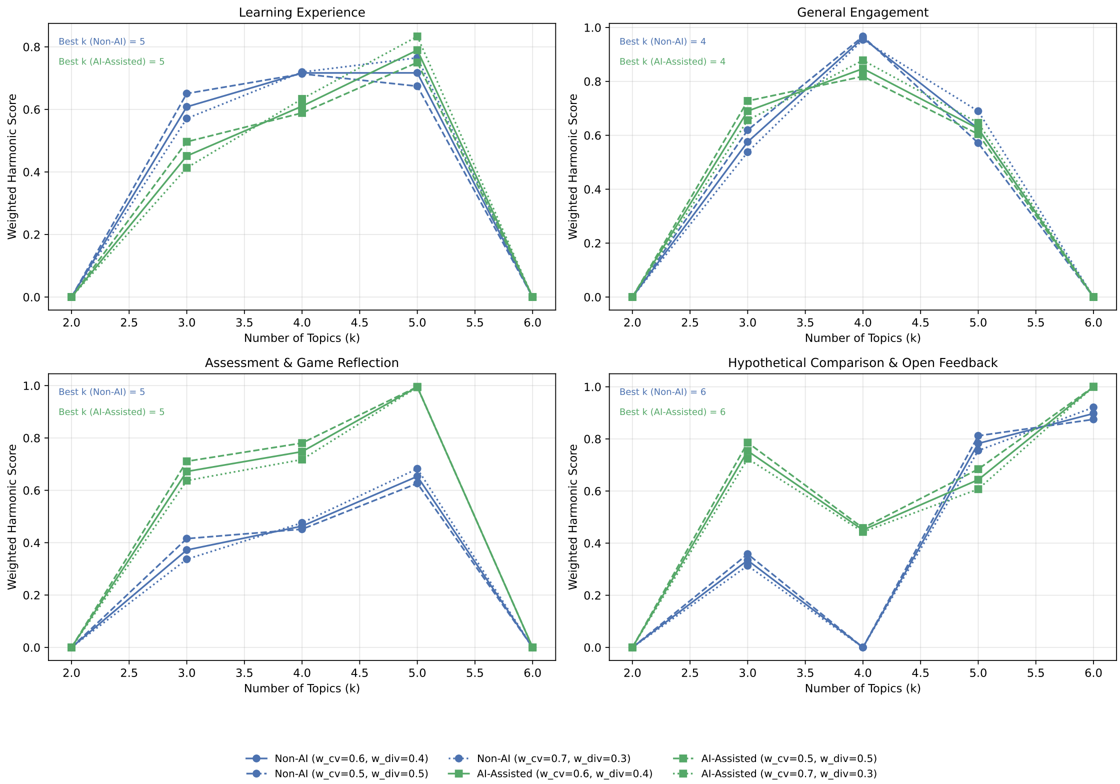

Figure 7.13: Selecting k per interview block by a simple coherence–diversity harmonic score for *topic surfacing*. This step is heuristic and non-binding; final themes are set by the author after re-reading the data.

- *Learning Experience*: k=5;
- *General Engagement*: k=4;
- *Assessment & Game Reflection*: k=5;
- *Hypothetical Comparison & Open Feedback*: k=6.

These settings suggest that the underlying thematic structure requires moderate granularity for learning processes and assessment reflections ($k=5$), a more compact representation for engagement dynamics ($k = 4$), and a finer partition for open, forward-looking feedback ($k=6$). Because the harmonic mean penalises imbalances between coherence and diversity, the observed peaks mark configurations where within-topic consistency and between-topic separation are jointly optimised. Moving to larger k risks fragmentation without substantive gain; reducing k compresses distinct topics and blurs interpretability.

Altering the weights placed on coherence vs. diversity shifts absolute score magnitudes but does not change the maximising k in any panel. This invariance supports the *robustness* of the selected topic counts and reduces concern that model selection is an artefact of a particular weighting. Adopting $k = \{5, 4, 5, 6\}$ for the four sections (Non-AI and AI-Assisted alike) yields like-for-like comparisons while preserving interpretability.

Learning Experience

Table 7.4: Learning Experience — Top Quotes per Topic (Non-AI on top; AI-Assisted below)

Non-AI		
Topic	**Quote**	**Similarity**
Difficulty with Complex Signs	... follow, but the word signs, especially mother and airplane, were more challenging because the videos didn't ...	0.7273
Clarity of Materials	... about the exact movement speed or finger angles.Yes I sometimes the hand in the video was at a different ...	0.7114
Self-Correction Challenges	... straightforward, but I had a hard time with mother and sorry because the materials didn't show the hand orientation ...	0.5686
Structured Learning Approach	... repeating the same small orientation error without noticing. Yes — the videos were short, so I sometimes had to ...	0.5359
Need for Additional Feedback	... video at key frames and matching my hand to them. A clear visual guide that's great for independent learners ...	0.4500
AI-Assisted		
Topic	**Quote**	**Similarity**
Immediate Improvement	... finger positioning helped me eventually correct them.A precise and tireless guide that doesn't let you ...	0.7900
Clarity of Instructions	... my hand because of the lighting in my room, so I had to adjust the angle.Water and ok were tricky - for ...	0.7166
Precision in Feedback	... I realized I was the one in the wrong.Once or twice the hand tracking didn't catch my movement if I went ...	0.7024
Challenges with Speed	... hand height and wrist angle until they looked right.Like having a coach who never gets tired of pointing ...	0.6461
Supportive Coaching Style	... which part was wrong — for example, "thumb should point outward." I trusted ...	0.5634

Across Non-AI accounts, participants consistently reported *difficulty with complex signs* (0.73) and *clarity of materials* (0.71) as their primary obstacles. These difficulties stemmed largely from the static, single-angle nature of the instructional videos, which limited visibility of fine-grained articulations such as subtle palm rotations, shifts in wrist tilt, and precise finger curvature. Signs involving compound or multi-directional movements (e.g., "mother", "airplane") were particularly prone to misinterpretation, as critical features often occurred outside the camera's angle or were not shown in slowed, segmented form. As one non-AI participant put it, "*... the word signs, especially mother and airplane, were more challenging because the videos didn't show the movement clearly ...*". Consequently, learners often defaulted to a cognitively demanding *trial-and-error* cycle: observe, attempt, self-evaluate through visual comparison and proprioception, and retry. This workflow increased the risk of unnoticed errors becoming entrenched over time.

Reports of *self-correction challenges* (0.57) highlight the mental effort required to identify error sources without external feedback. Participants frequently noted that they could repeat the same small orientation or handshape error across several attempts without realising it. The topic of a *structured learning approach* (0.54) offers insight into how learners coped: many relied on the short, looped nature of the videos, breaking tasks into repeated micro-practice cycles that allowed some degree of consistency and rhythm. While this structured repetition created stability, it also revealed a structural limitation—without external validation, learners risked reinforcing inaccuracies rather than refining them. This explains the recurring calls for *additional feedback* (0.45), such as requests for error highlighting or frame-by-frame comparisons, which would provide a mechanism to confirm whether corrections were accurate. Taken together, the Non-AI experience was shaped by heavy reliance on perception and memory, combined with individually improvised routines to manage error detection and correction.

In contrast, AI-Assisted narratives clustered more tightly around *immediate improvement* (0.79), *clarity of instructions* (0.72), and *precision in feedback* (0.70), reflecting a narrower but more focused set of experiences. Here, clarity did not refer to the comprehensibility of the demonstration, but to the *explicitness of diagnostic feedback*. Participants valued instructions such as "straighten your index finger" or "thumb should point outward," which pinpointed the exact biomechanical adjustment required. This diagnostic specificity effectively externalised error detection, shifting learners' cognitive load away from identifying mistakes toward simply executing corrections. As a result, we infer that the feedback loop was compressed—recognition and correction often occurred within a single cycle—which may have accelerated visible progress and reinforced confidence. Minor frictions—such as sensitivity to movement speed (0.65) or occasional tracking glitches under poor lighting—were acknowledged, but these were perceived as technical nuisances rather than structural barriers to learning. The system's *supportive coaching style* (0.56) further strengthened its perceived reliability, framing feedback as constructive guidance rather than evaluative judgment.

A key comparative insight is that while both groups emphasised "clarity" as central to effective learning, the meaning diverged sharply. In the Non-AI mode, clarity concerned the *representation of the model sign*—whether learners could perceive and interpret the intended movement from video materials. In the AI-Assisted mode, clarity referred to *diagnostic specificity*—how precisely the system identified and communicated learners' own errors. This distinction reflects a broader pedagogical shift: Non-AI participants engaged in **perception-based learning**, relying on self-directed observation and pattern recognition [192, 193], whereas AI-Assisted participants engaged in **correction-based learning**, where error detection was largely externalised [194, 195]. This structural difference in feedback flow reduced uncertainty and cognitive load in the AI-Assisted condition, positioning it as a more efficient pathway for developing motor accuracy and learner confidence.

General Engagement

Table 7.5: General Engagement — Top Quotes per Topic (Non-AI vs AI-Assisted)

Non-AI		
Topic	Quote	Similarity
Motivation from Progress	... but I liked the flexibility to practice when I wanted. ...	0.6505
Self-Paced Learning Benefits	... motivated because I enjoy structured skill building.Less engaging than interactive feedback but good for self-paced ...	0.6326
Preference for Structured Learning	... but I appreciated being able to replay sections as many times as I needed. ...	0.6147
Challenges with Engagement	... Moderately motivated. I enjoyed the structured approach, but I think live feedback would have made it more engaging. ...	0.4864
AI-Assisted		
Topic	Quote	Similarity
Interactive Feedback	... Seeing my improvement in both digits and words reinforced the effort.Definitely. ...	0.7035
Progress Tracking	... The interactive feedback kept me alert. ...	0.6797
Game Motivation	... Highly motivated - I kept wanting to beat my own game scores and fix old mistakes.Yes, ...	0.6534
Conversational Learning	... I was pretty motivated because I could see a visible difference in my scores each round.Yes, ...	0.6066

Non-AI engagement patterns were characterised by a steady, structure-driven orientation. Topics of *motivation from progress* (0.65), *self-paced learning benefits* (0.63), and *preference for structured learning* (0.61) point to a mode that rewards persistence and autonomy. Participants valued the ability to *control pacing*—for instance, replaying sections until movements felt correct—framing repetition as both a confidence-builder and a safeguard against forgetting; as one non-AI learner put it, "*... I could replay sections as many times as I needed ...*". Flexibility was also emphasised: "*... I liked the flexibility to practice when I wanted ...*". This approach aligned well with learners who favour deliberate, methodical skill accumulation. However, *challenges with engagement* (0.49) emerged when the absence of interactive prompts led to disengagement, especially for those who respond better to dynamic feedback. In such cases, structure alone maintained compliance but struggled to generate the kind of affective investment that keeps attention high over extended sessions.

In contrast, AI-Assisted engagement was anchored in *interactive feedback* (0.70) and *progress tracking* (0.68), supplemented by motivational triggers such as *game motivation* (0.65) and a *conversational learning* style (0.61). Here, engagement was not merely sustained through self-direction, but actively fuelled by a continuous loop of challenge, feedback, and visible achievement. Interactive elements transformed practice into a quasi-dialogic exchange, where the system acted as both coach and performance monitor. Progress tracking reinforced this dynamic by providing quantifiable evidence of improvement, while gamified elements—like beating previous scores—introduced an element of competition, even in solo practice. These factors collectively shifted the learner's mindset from **structured compliance** to **active investment**, with visible progress serving as an intrinsic reward mechanism [196].

The thematic distribution also reflects a difference in engagement mechanics: in Non-AI learning, motivation was primarily *self-regulated* [197] and dependent on internal discipline, whereas in AI-assisted learning, motivation was *co-regulated* [198]—emerging from a system-learner interaction that continually stimulated attention and rewarded effort. This distinction suggests that while both approaches can sustain engagement, AI-assisted modes are more likely to convert participation into sustained, high-energy involvement by coupling structure with interactive, personalised reinforcement.

Assessment & Game Reflection

Table 7.6: Assessment & Game Reflection — Top Quotes per Topic

Non-AI		
Topic	Quote	Similarity
Reduced Errors	... and water.It made practice more exciting and pushed me to react faster. ...	0.6310
Improved Recall Speed	... and mother.It made recall much faster. ...	0.6156
Increased Confidence	... Proud - I cut my errors from 17 in round one to only 5 by the final assessment.All digits, ...	0.6147
Motivation and Enjoyment	... mistakes from 13 in the first round to only 4 in the last round.All digits, plus hello, ok, and water.It was ...	0.5202
Pressure Handling	... signs to just 3 in the last round.All digits, water, hello, and ok.It made me quicker and more accurate ...	0.4592

AI-Assisted		
Topic	Quote	Similarity
Reduction in Errors	... by the end I was down to one mistake, ...	0.6615
Game's Impact on Recall	... father, sorry, and yes still weren't correct.Definitely 0 and love - I can do them quickly and confidently.The ...	0.6325
Confidence in Performance	... Huge improvement. I nailed ok and love, which I missed in the first assessment, and my digits became almost automatic. ...	0.5857
Improved Reaction Speed	... and most digits except 6.It really pushed my reaction speed. ...	0.5672
Mixed Feelings About Pressure	... The speed element forced me to recall quickly, which carried over into the assessment. ...	0.4829

For the Non-AI group, reflections converged on steady gains achieved through repeated, structured practice. The dominant topics were *reduced errors* (0.63), *improved recall speed* (0.62), and *increased confidence* (0.61). Participants often described a progression in which accuracy improvements reinforced speed, creating a positive feedback loop where familiarity reduced hesitation—"*... I reduced my mistakes from 13 in the first round to only 4 in the last round ...*"; "*... it made recall much faster.*". These gains were attributed to the ability to rehearse at a controlled pace, allowing fine-tuning without the cognitive strain of time pressure. Confidence, in this context, was largely built through consolidation—each correct repetition served as a confirmation of skill stability. However, *pressure handling* scored lower (0.46), reflecting mixed experiences: "*... the game was both fun and stressful.*". This suggests that in the Non-AI mode, improvement was anchored in mastery through repetition rather than adaptive performance under dynamic conditions.

AI-assisted reflections revealed a parallel set of topics but with a more performance-oriented edge. Again, *reduction in errors* (0.66) and *game's impact on recall* (0.63) were prominent, alongside *confidence in performance* (0.59) and *improved reaction speed* (0.57). Here, the interplay between speed and feedback was central: the timed, game-like assessments demanded rapid retrieval, while immediate corrective prompts ensured that errors were addressed before becoming entrenched. Participants described a carry-over from the game to the assessment; for example, "... the speed element forced me to recall quickly, which *carried over into the assess-*

ment ...". We therefore refer to this as a carry-over effect in our narrative, where the heightened pace and focus of the game transferred directly to the formal assessment. The *mixed feelings about pressure* (0.48) topic indicates that while most participants experienced this as productive stress — aligning with the concept of "desirable difficulty" — a minority found it introduced performance anxiety. Nonetheless, the balance of challenge and support appears to have tilted the learning environment toward high engagement and skill generalisation, in contrast to the Non-AI mode's emphasis on controlled repetition.

Taken together, these patterns point to a functional difference in how each modality shaped assessment readiness. Non-AI practice reinforced accuracy and stability through self-paced consolidation, cultivating a strong baseline but less exposure to high-pressure retrieval. Although the only timed component was the Whack-a-Mole game (used in both conditions), the AI-assisted mode paired that common time pressure with a feedback-rich learning phase; this combination may have fostered faster, more accurate responding at test.

Hypothetical Comparison & Open Feedback

Table 7.7: Hypothetical Comparison & Open Feedback — Top Quotes per Topic

Non-AI		
Topic	Quote	Similarity
Multi-Angle Videos	... sorry sign would have improved much faster with real-time feedback.Show the hand movement from multiple ...	0.6796
Real-Time Feedback	... real-time correction would make a big difference for subtle errors.Include a "practice with timer" mode ...	0.6500
Video Demonstrations	... I think my tricky signs would have been corrected earlier, ...	0.6483
Slow-Motion Playback	... would have corrected my sorry sign much faster.Add animated examples and optional slow-motion playback for ...	0.6178
Error Highlighting	... repeated mistakes - especially for tricky finger positions.Include a "review mistakes" section after each ...	0.5914
Comparison Tools	... I probably would have made fewer small errors in the early stages, ...	0.4979
AI-Assisted		
Topic	Quote	Similarity
Challenges with Mistakes	...I'd probably learn the basics, but I wouldn't catch small mistakes that matter in sign language clarity....	0.6946
Need for Different Perspectives	...I'd get the general shapes right, but I'd keep subtle mistakes that would make my signing less clear. ...	0.6415
Focus on Specific Signs	... I'd probably still be stuck on sorry. I need interactive nudges, not just watching....	0.6081
Desire for Progress Tracking	... I'd probably still learn digits fine, but my tricky words would take much longer to master ...	0.5601
Need for Real-Time Feedback	...I'd likely keep some bad habits without realizing it.Maybe add a "challenge mode" where the AI speeds ...	0.5529
Need for Replay Features	... and I might not fix subtle errors.I'd love a side-by-side replay of my gesture next to the perfect one ...	0.4476

When reflecting on unmet needs, both groups converged on a desire for richer, more controllable feedback mechanisms that would allow them to diagnose and verify their own performance. In the Non-AI group, the most common requests were for *multi-angle videos* (0.68), *real-time feedback* (0.65), *video demonstrations* (0.65), *slow-motion playback* (0.62), *error highlighting* (0.59), and *comparison tools* (0.50). Learners articulated clear gaps in the materials: "*... show the hand movement* *from multiple angles ...*", "*... real-time correction would make a big difference for subtle errors ...*", and "*... add slow-motion playback for tricky signs ...*". Together, these features signal a preference for high-control learning environments, where users can pause, replay, and scrutinise details at will, reducing reliance on memory or guesswork during self-correction.

Even among AI-Assisted users — who already received immediate corrective prompts — similar needs surfaced, though the emphasis shifted toward edge-case handling and fine-grained control. Requests included addressing *challenges with mistakes* (0.69), offering *different perspectives* (0.64), enabling focused drills on *specific signs* (0.61), enhanced *progress tracking* (0.56), refining *real-time feedback* (0.55), and adding *replay features* (0.45). Participants pointed to the limits of in-the-moment guidance and asked for post-hoc verification: "*... I'd get the general shapes right, but I'd keep subtle mistakes that make my signing less clear ...*", "*... I'd probably still be stuck on sorry—I need interactive nudges, not just watching ...*", and "*... I'd love a side-by-side replay of my gesture next to the perfect one ...*". These suggestions indicate that while automated guidance improved baseline performance, learners still sought tools to inspect borderline cases and confirm corrections when recognition was ambiguous.

Across both groups, these priorities reveal a shared vision of "transparent correction pathways" [194, 199]: learners want not just to be told they are wrong, but to see and understand why, from multiple vantage points and at a level of detail they control. This points toward hybrid feedback designs that combine the immediacy of AI with the interpretive clarity of visual artefacts, enabling both real-time course correction and deliberate, self-paced skill refinement. Such designs could serve as a common ground for learners who thrive under guided instruction and those who prefer autonomous mastery.

7.8 Discussion

This study contrasted *AI-assisted* and *Non-AI* approaches to introductory ASL learning across three rounds, triangulating accuracy, response time, improvement, learning time, game performance, and qualitative accounts of experience, engagement, and needs. Overall, the pattern is consistent: AI assistance compresses the feedback loop by externalising error detection, yielding higher post-session accuracy, faster responses, and tighter performance distributions, while Non-AI practice relies more on structured repetition and time-on-task to consolidate skills [200–202].

7.8.1 Summary of Key Findings

Performance and efficiency The combined accuracy–response-time figure (Figure. 7.7) shows steeper pre–post gains and faster responding for the AI-assisted group across rounds, with higher pre-session baselines by Rounds 2–3, indicating cross-session carry-over. Mixed-effects modelling (Table 7.2) confirmed strong within-session learning (pre $<$ post) and cumulative gains over rounds, with a significant three-way interaction in Round 3 indicating larger pre–post improvements for AI-assisted learners. Improvement analyses revealed that AI's advantage was strongest early-to-mid study and narrowed later, consistent with diminishing returns as learners approach ceiling and with staging effects in skill acquisition [203,204]. Boxplots of game performance showed higher medians and fewer low-end outliers for AI-assisted participants, suggesting improved *consistency* in addition to mean benefits.

Learning time and time–accuracy coupling Total learning time decreased across rounds for both groups (Fig. 7.9), with no reliable overall difference between conditions (Table 7.3). Correlational analyses indicated that in Non-AI, time spent was more often positively related to post-session accuracy (notably Round 1), whereas in AI-assisted learning this coupling was weak or absent. This implies AI improves the *quality of practice* (diagnostic specificity; knowledge-of-performance feedback) rather than simply reducing duration [200, 205]. By contrast, Non-AI gains appear more dependent on practice time and repetition [206].

Learner experience and needs. Qualitative themes corroborate the mechanism-level account. Non-AI learners reported difficulties with complex signs and material clarity, adopting structured micro-practice routines to compensate, yet risking reinforcement of subtle errors. AI-assisted learners emphasised immediate improvement and precise, actionable feedback; minor frictions (e.g., lighting sensitivity) were treated as technical rather than pedagogical barriers. Both groups requested richer, controllable artefacts (multi-angle views, slow-motion, side-by-side replay), pointing to a shared desire for *transparent, self-verifiable* correction pathways [201, 202, 207].

7.8.2 Mechanisms: From Perception-Based to Correction-Based Learning

A central contribution is specifying how AI assistance reshapes the cognitive workflow. In the Non-AI setting, learning is *perception-based*: learners infer correctness by aligning perception and proprioception with a model video, bearing the cognitive load of diagnosis. In the AI-assisted setting, learning becomes *correction-based*: error detection is externalised and the learner's effort shifts from *finding* to *fixing* the error. This transformation shortens the feedback loop, increases trust in adjustments, and reduces uncertainty—in line with knowledge-of-performance feedback in motor learning and ITS evidence on step-level guidance [200–202]. Across rounds, accuracy rose while response time fell. Given that assessments did not impose speed incentives, we interpret this joint improvement as increased fluency (automatisation) rather than a speed–accuracy trade-off, consistent with accounts in which practice and specific feedback shift the speed–accuracy function outward [208].

7.8.3 When AI Helps Most—and Why the Gap Narrows

The AI advantage is greatest when there is ample room for correction: the early and mid rounds showed higher pre-post gains and higher post-session scores for AI-assisted learners. As baseline accuracy climbs and errors become rarer and more idiosyncratic, Non-AI learners' deliberate, self-paced routines can produce steady marginal gains, leading to partial convergence in late training. This pattern aligns with *desirable difficulties*: early bandwidth-type feedback accelerates initial skill acquisition; later, carefully dosed challenge and reflection sustain growth [203, 204]. Our game analyses show that medians increase from Round 1 to Round 3 in both groups. By Round 3, outliers (circles) are visible only in the Non-AI condition, whereas the AI-assisted condition exhibits wider whiskers, indicating greater dispersion rather than compression toward the median. We therefore refrain from characterising AI as 'tightening' the distribution and instead note practice-related median gains with round-specific variance patterns.

7.8.4 Implications for ASL Pedagogy and System Design

Design AI feedback for diagnostic specificity and control. Actionable, biomechanical feedback (e.g., thumb orientation, wrist angle) should remain core. To satisfy learners' need for self-verification, systems should offer *controllable artefacts*: multi-angle exemplars, slow motion, frame-by-frame scrubbing, and side-by-side overlays of learner vs. model [202, 207].

Fuse real-time coaching with reflective replay. Real-time prompts accelerate correction; reflective tools consolidate learning and support metacognition. A practical pattern is a *two-pass loop*: (1) live guidance during attempts; (2) optional replay with visual error highlights and uncertainty cues [201].

Adapt feedback intensity and difficulty over time. Given early advantages and later convergence, systems should *fade* feedback and increase task challenge (adaptive timing, variable angles) to maintain desirable difficulty [203,209]. A dashboard of *feedback bandwidth* enables personalisation.

Target the long tail and reduce variance. Because AI assistance reduced low-end outliers, designs should explicitly support learners at risk of lagging—e.g., uncertainty-aware prompts, micro-drills for persistent errors, and "second-chance" practice slots [201].

Support robustness in real settings. Provide setup diagnostics (lighting gauge, camera framing helper), tolerant thresholds, and confidence visualisations so learners know when to trust the model [210].

7.8.5 Implications for Practitioners

For instructors and programme designers:

- **Blend modalities.** Pair AI coaching with structured replay artefacts (multi-angle exemplars, slow-motion, side-by-side) in lab and homework to combine fast correction with reflective consolidation [202, 207].
- **Stage the difficulty.** Use AI more intensively early on; progressively fade hints and raise contextual challenge (e.g., timing pressure, angle variability) after basics stabilise [203, 209].
- **Monitor the floor as well as the mean.** Track distributional metrics (low-end outliers, IQR)—AI should reduce variance, not only raise averages.
- **Teach with transparency.** Make diagnostic rules visible (what constitutes an error and why), enabling students to build self-checking habits transferrable to non-AI contexts [210].
- **Plan for environment.** Standardise lighting/camera protocols or provide quick setup checks to minimise technical friction.

7.8.6 Limitations and Threats to Validity

Participants were beginners and tasks were limited to digits and static words; generalisation to coarticulated phrases requires further study. The recognition system's sensitivity to environmental factors (e.g., lighting) introduces ecological constraints. Although mixed-effects models account for repeated measures, between-subjects assignment may still carry residual confounds (e.g., motivation). In the qualitative study, topic selection was validated via a coherence–diversity harmonic score, but interpretation remains partly subjective. Finally, game-based measures may conflate motor fluency with task familiarity; transfer to unseen signs and spontaneous production should be tested [201, 211].

7.9 Conclusion

We compared an AI-assisted ASL learning pipeline—real-time recognition with concise, adaptive feedback (LLM used only for phrasing)—to a non-AI, predefined-materials mode across three rounds. A repeated-measures analysis was complemented by a human-in-the-loop qualitative pipeline (TF–IDF + NMF topic modelling, coherence/diversity–guided k selection, SBERT-based quote selection with human verification).

Quantitatively, both groups improved across rounds; the pre–post gap was largest in Round 1 and smaller thereafter, with a round-specific advantage for the AI-assisted group in Round 3. Learning time declined in both groups; correlations between time-on-task and accuracy were clearer in the non-AI mode early on. Game medians rose for both groups. Variance patterns were round-specific rather than uniformly reduced in the AI condition. Accuracy increased while response time decreased without speed incentives, consistent with fluency gains rather than a speed–accuracy trade-off.

Qualitatively, 'clarity' diverged by condition: non-AI learners emphasised the *representation of the model sign* (perception-based learning), whereas AI-assisted learners emphasised *diagnostic specificity* about their own errors (correction-based learning).

CHAPTER 8

Discussion

8.1 Introduction

This chapter synthesises findings across Chapters 3–7 to answer the thesis research questions and to advance a unified account of *when and why* emerging technologies improve American Sign Language (ASL) learning. The studies collectively examined immersive media (VR/3D), gamification, gesture personalisation, and AI-based diagnostic feedback under *matched content, timing, and scoring*.

> **Thesis argument.** *Diagnostic, timely, and interpretable feedback is the primary driver of dependable ASL learning gains. Immersion (VR/3D), gamification, and personalisation act chiefly as engagement and access multipliers that help deliver and sustain such feedback. Where learning objectives* require *spatial affordances, 3D adds value; otherwise, well-designed 2D can match 3D.*

We first integrate the empirical results across chapters, then provide self-contained answers to RQ1–RQ5 with supporting evidence, and finally show how these answers combine to address the overarching thesis question. We situate contributions relative to prior literature and reframe limitations as *applicability and boundaries of generalisation* with practical implications.

8.2 Integrated synthesis and critical argument

This section advances a unified account of *when and why* the technologies examined in Chapters 3–7 improve ASL learning. Rather than revisiting each study in isolation, the argument proceeds by drawing cross-chapter connections that explain both positive and null findings under a single mechanism.

Across heterogeneous manipulations—platform (VR versus web), learning mechanics (gamified versus non-gamified), interaction (predefined versus user-defined gestures), representational dimensionality (2D versus 3D), and feedback pathway (AI-assisted versus non-AI)—two regularities are evident. First, design choices that heighten engagement consistently raised presence (the subjective sense of "being there" in the virtual environment), motivation, or perceived autonomy. Immersive VR and 3D interfaces increased the sense of presence relative to matched 2D or web baselines (Chapters 3 and 6); a Whack-a-Mole style activity sustained motivation and within-session persistence (Chapter 4); and allowing user-defined gestures enhanced perceived control and satisfaction (Chapter 5). *Where matched comparisons were used (Chapters 3, 4, and 6), these shifts occurred under content-, timing-, and scoring-controlled conditions, indicating genuine affective or behavioural gains rather than dosage artefacts.*

Second, and more importantly for learning, robust improvements in accuracy and response efficiency tracked the *specificity and timeliness of corrective guidance.* When feedback was *diagnostic*—identifying concrete biomechanical adjustments to handshape, orientation, or wrist configuration—and delivered promptly, learners achieved steeper within-session gains and faster responses under otherwise identical content and timing (Chapter 7). By contrast, when guidance was presentation-only (as in matched VR versus web demonstrations), or when motivation was raised without coupled correction (as in gamification without added diagnosticity), engagement increased but item-level gains did not reliably exceed the matched baseline (Chapters 3–4, 6). *In Chapter 5, user-defined gestures improved autonomy and satisfaction but introduced recognition ambiguity in some cases; absent a matched comparator, this is best interpreted as a design trade-off and a potential risk to the fidelity of corrective signals unless guardrails (guided capture, calibration, intelligibility checks) are in place.*

Taken together, these regularities support a mediator account in which immersion, gamification, and personalisation operate primarily as *carriers* that attract and sustain attention and increase the amount and intensity of practice (time on task, repetitions, spacing, task difficulty), while *diagnostically specific feedback* acts as the *active ingredient* that converts practice into dependable improvement. The same account explains the conditional value of 3D: when learning objectives intrinsically require spatial affordances such as multi-view reasoning, occlusion handling, or classifier morphology, 3D offers instrumentally useful information; when objectives are entry-level lexical items that do not leverage those affordances, a well-designed 2D interface paired with specific guidance is empirically comparable (Chapter 6). This mechanism-level synthesis not only reconciles positive presence effects with modest accuracy differences but also clarifies why AI assistance that externalises error detection shifts performance curves even when exposure is held constant (Chapter 7).

8.3 Answering the research questions, then the main thesis question

The research questions can now be answered in a manner that is both self-contained and grounded in cross-chapter evidence, with each result motivating the next empirical step.

RQ1: Does immersive VR enhance ASL learning compared with non-immersive platforms? The controlled comparison of immersive VR with a matched web interface for numeral learning (Chapter 3) showed substantially higher presence and engagement in VR, yet only small and statistically non-reliable differences in post-test accuracy, with response times broadly comparable. Under matched pedagogy, exemplars, and exposure, immersion functioned as an affective scaffold rather than an independent driver of item-level gains. The implication is that VR is best understood as an attentional amplifier whose learning value depends on whether it helps to deliver—or focus attention on—diagnostic cues. *This gap between engagement and item-level gains directly motivated Chapter 4: if immersion alone does not convert attention into learning, can structured motivation via game mechanics increase practice and thereby improve outcomes under matched materials?*

RQ2: What is the role of gamification in ASL learning? The gamified condition (Chapter 4) produced clear increases in enjoyment and steadier within-session practice relative to non-gamified study with matched materials. However, improvements in retention and accuracy were heterogeneous across items and participants. Gamification therefore acts as a vehicle for practice and engagement; its learning benefit depends on whether gameplay is coupled to feedback that makes the next correction obvious and immediately testable. Points and time pressure, in themselves, are insufficient to guarantee durable representational change. *This result sharpened the focus from "more practice" to the* ***quality*** *of practice, prompting Chapter 5 to examine whether increasing learner control through user-defined gestures improves experience and how such personalisation interacts with the dependability of machine feedback.*

RQ3: How do user-defined gestures affect learning experiences? Allowing learners to define their own gestures (Chapter 5) increased perceived autonomy and satisfaction in a VR, game-supported setting, with positive patterns on *Stimulation*, *Novelty*, and especially *Perspicuity* (the highest of the six factors). At the same time, the *Dependability* factor and open-ended comments revealed recognition ambiguity in some cases (e.g., participants reporting that a gesture they believed correct was judged incorrect), despite the recogniser's overall accuracy exceeding 90% in our setting. Chapter 5 did not run a direct predefined-versus–user-defined recogniser comparison, so we do not claim a performance delta between these mappings; rather, we infer a plausible tension: personalisation can enhance experience while introducing uncertainty into the recogniser output that underpins corrective guidance. In practical terms, unless guardrails—guided capture, calibration, and intelligibility checks—are in place, personalisation risks *reducing the effective fidelity* *of the corrective signal. Having established that interactional choices can strengthen or weaken corrective information, Chapter 6 then examined whether the **representation** of that information (2D vs. 3D) alters outcomes when pedagogy and exposure are held constant, i.e., whether dimensionality adds value beyond feedback quality.*

RQ4: Are 3D interfaces more effective than 2D for ASL learning? In a matched comparison (Chapter 6), both interfaces achieved high usability (SUS: 2D M=76.25, 3D M=80.63, p=0.103). User experience favoured 3D on hedonic qualities (stimulation/novelty) and overall attractiveness, with mixed patterns on pragmatic qualities (e.g., some *perspicuity* items were easier for 2D). On performance, there was *no significant first-attempt difference* in game scores between 2D and 3D; both conditions showed a strong within-session learning effect across attempts, and swapping environments between attempts produced a small but significant decrement (p=0.043), consistent with interface re-familiarisation rather than representational advantage. Because pedagogy, exemplars, exposure, and timing were held constant, dimensionality alone did not yield additional item-level gains under the short (30s) gameplay window used here. The reading is therefore conditional: 3D increases engagement and presence-related qualities and is *instrumentally* advantageous when learning objectives require spatial affordances; otherwise, a well-designed 2D interface delivering clear, specific guidance achieves comparable immediate outcomes with lower complexity. *With medium, motivational scaffolds, interactional control, and representational format now characterised as **carriers**, Chapter 7 tested the hypothesised **active ingredient**: whether diagnostically specific, timely feedback—delivered via AI assistance—converts practice into dependable gains under identical content and timing.*

RQ5: Can AI provide effective feedback in ASL learning? The three-round, between-subjects study (Chapter 7) contrasted an AI-assisted pathway—real-time recognition with concise, per-feature tips (LLM used for phrasing)—with a non-AI pathway using predefined materials under identical content and timing. Both groups showed robust pre–post improvement each round. The AI-assisted group generally exhibited higher post-session accuracy and faster responses across rounds, with linear mixed-effects modelling confirming strong within-session learning overall and a *round-specific* AI advantage in Round 3 (significant Phase $\times$ Round $\times$ AI interaction), while earlier-round differences were smaller or marginal. Accuracy increased while response times decreased in both groups, indicating fluency gains rather than a speed–accuracy trade-off. Time-on-task correlated with accuracy in the non-AI condition (notably in Round 1, and again trending in Round 3) but not in the AI condition, suggesting more efficient minutes when feedback externalises error detection. Interviews corroborated this mechanism: non-AI learners emphasised clarity of the *model sign* (perception-based learning), whereas AI-assisted learners emphasised *diagnostic specificity* about their own errors (correction-based learning). *These findings complete the progression from Chapters 3–6 by isolating feedback specificity as the factor that reliably turns engagement and exposure into learning, while also clarifying that the AI advantage is largest when there is headroom for correction and narrows as proficiency rises.*

These answers compose a single, overarching conclusion to the main thesis question: *Across modalities and features, what chiefly drives dependable ASL learning gains?* The weight of evidence indicates that the decisive factor is the *diagnostic specificity of feedback*—timely, interpretable, per-feature guidance that tells the learner exactly what to change. Engagement-oriented features such as VR/3D, gamification, and personalisation are most effective when they deliver, focus, and sustain this guidance. Where tasks demand spatial inference, 3D is instrumentally valuable; where they do not, 2D with high-quality guidance suffices. This unifying mechanism reconciles the presence advantages of immersive media with the modest accuracy differences observed under matched pedagogy, explains the conditional benefits of gamification and personalisation, and accounts for the superior efficiency observed when AI assistance compresses the recognition–correction cycle.

8.4 Positioning within the literature: alignment, divergence, and advancement

The evidence assembled across Chapters 3–7 both confirms and qualifies prevailing claims about immersive media, gamification, gesture-based interaction, and AI-supported feedback in technology-enhanced learning. In what follows, I situate the findings within the wider literature, indicating where they align with established results, where they diverge or refine existing claims, and where they advance the state of knowledge in the specific context of ASL learning.

Alignment with prior work

Prior work on immersive and 3D learning consistently shows robust gains in presence and engagement, with learning advantages that are conditional on instructional design rather than guaranteed by medium alone [19, 38, 42]. Building on this base, the present thesis contributes controlled, *matched-pedagogy* contrasts that hold content, exposure, timing, and scoring constant. Under these conditions, Chapter 3 demonstrates that VR elevates presence without yielding superior item-level accuracy over a web baseline for entry-level ASL numerals, and Chapter 6 extends the result to 2D versus 3D representations within an otherwise identical game loop. These findings align with the literature's caution that dimensionality is not an active ingredient in itself unless the task genuinely requires spatial affordances; they add precision by showing this boundary under tightly controlled ASL tasks rather than heterogeneous materials.

Gamification research has repeatedly documented reliable effects on motivation, time-on-task, and persistence, with mixed or indirect effects on attainment unless game loops are coupled to high-quality practice and feedback [25, 28]. Chapter 4 replicates the motivational effect in a sign-learning context—enjoyment rises and within-session practice is steadier—while the heterogeneous retention and accuracy gains observed here qualify broad claims of efficacy. The innovation lies in demonstrating, under matched materials, that gamification's instructional value depends on whether gameplay actually *delivers diagnostic information* that makes the next correction obvious and testable, rather than merely increasing repetition.

Work on personalisation often assumes that greater learner control uniformly improves learning experience and outcomes. Chapter 5 aligns with the first part of this claim—user-defined gestures increased perceived autonomy and satisfaction—but advances the literature by showing the *mechanism of risk*: personalisation can degrade recogniser reliability and thus the *fidelity of the corrective signal* unless guardrails (guided capture, calibration, intelligibility checks) are in place. This extends prior UI-centric studies by linking interactional freedom directly to the quality of feedback that underpins skill acquisition in ASL.

Finally, decades of work in intelligent tutoring and formative assessment show that timely, specific, knowledge-of-performance feedback drives learning and fluency [154, 157, 202, 207]. Chapter 7 aligns with this result and extends it for ASL via a repeated-measures design that tracked both accuracy and response time across three rounds under identical content and timing, showing concurrent accuracy gains and faster responses (i.e., fluency rather than a speed–accuracy trade-off). Mixed-effects modelling further indicated a round-specific advantage for the AI-assisted condition (larger pre–post gain in Round 3), while interviews attributed progress to concise, biomechanical guidance. Together, these contributions move beyond aggregate accuracy to reveal efficiency gains and to link outcomes to the perceived clarity and actionability of feedback.

Across these areas, the thesis thus converges with prior evidence on the motivational strengths of immersion and gamification and the centrality of specific, timely feedback, while contributing tighter causal identification (via matched designs), mechanism-level clarification (feedback fidelity as the mediating pathway), and efficiency-sensitive outcomes (response time alongside accuracy) in the specific context of entry-level ASL learning.

Points of divergence and clarification

While broadly consistent with prior findings, this thesis adds nuance where claims in the literature are sometimes generalised beyond their evidential base. First, studies reporting global learning benefits of VR or 3D sometimes conflate increased *presence* with changes in pedagogy. By holding content, exposure, timing, and scoring constant, Chapters 3 and 6 show that immersion's affective advantages do not automatically yield higher item-level accuracy for entry-level ASL, clarifying that dimensionality is not an *active ingredient* in itself unless the task genuinely recruits spatial affordances [11, 19, 38, 42, 102]. The implication is caution against interpreting presence gains as a proxy for learning without attention to diagnostic guidance and task–affordance fit.

Second, personalisation is frequently framed as unconditionally desirable in HCI and learning technologies. Chapter 5 qualifies this view: user-defined gestures increased perceived autonomy and satisfaction but also introduced recognition ambiguity, which can degrade the *fidelity of the corrective signal* that formative feedback depends on. This mechanism-level clarification aligns with human–AI design guidance emphasising error tolerance and transparency [178, 179] and extends user-defined gesture work by showing that, in ASL learning, personalisation is *conditional* on guardrails (guided capture, calibration, intelligibility checks) to preserve recogniser reliability [82, 86, 87].

Third, claims that gamification improves attainment often do not specify whether the game loop *delivers diagnostic information* or merely increases repetition. Consistent with meta-analyses and syntheses that find reliable motivational effects but mixed attainment outcomes [25, 28, 167, 168], Chapter 4 demonstrates that enjoyment and within-session persistence increase under matched materials, while retention and accuracy gains remain heterogeneous unless gameplay is coupled to clear, testable corrections. This specifies the pedagogical precondition—timely, specific feedback—under which game mechanics translate practice into representational change [202, 207].

Taken together, these divergences refine broad claims in immersive, personalised, and gamified learning by identifying *feedback fidelity* and *task–affordance fit* as the mediating conditions under which design choices affect ASL learning outcomes.

Advancement and original contribution

Conceptual. This thesis advances a mechanism-centred account of technology-enhanced ASL learning by proposing and substantiating a *mediator* view in which engagement levers (VR/3D, gamification, personalisation) function primarily as *carriers* that attract and sustain attention and increase the amount and intensity of practice (time on task, repetitions, and spacing), whereas *diagnostically specific feedback* is the *active ingredient* that converts practice into dependable gains. This synthesis reconciles the well-documented motivational advantages of immersion and 3D with often modest accuracy differences under matched pedagogy [11, 19, 38, 42, 102], aligns with evidence that game mechanics raise engagement more reliably than attainment absent targeted guidance [25, 28, 167, 168], and locates the causal lever in the timing and specificity of formative feedback [154,202,207]. Concretely, the thesis explains (i) presence gains without accuracy advantages in Chapters 3 and 6 when content, exposure, and scoring are fixed; (ii) the conditional value of gamification in Chapter 4 when gameplay does or does not surface concrete, testable corrections; (iii) the trade-offs in user-defined interaction in Chapter 5 through their impact on the *fidelity of the corrective signal*; and (iv) the superior efficiency of AI-assisted feedback in Chapter 7 when error detection is externalised and guidance is per-feature and timely. By situating the *site of causality* in feedback specificity rather than medium or mechanics, the thesis integrates strands of research often treated separately and is consistent with broader instructional theories that emphasise the alignment of task demands, representations, and guidance [102, 154].

Methodological. The work contributes ASL-specific, controlled comparisons that hold pedagogy and exposure constant across modalities—an approach that reduces attribution ambiguities common in immersive learning studies where materials, time-on-task, and interface change simultaneously [19,38]. The matched designs in Chapters 3 and 6 isolate the effect of medium/representation; Chapter 4 holds content constant while varying game mechanics; Chapter 5 examines interactional control while tracking recognition reliability; and Chapter 7 employs a repeated-measures comparison of AI-assisted versus non-AI learning that models *both* accuracy and response time across three rounds. This yields a richer portrait of *fluency* (simultaneous accuracy gains and faster responses) rather than accuracy alone, and it links quantitative trajectories to learner accounts of feedback clarity—tightening the explanatory chain from intervention to mechanism to outcome in line with best-practice evaluation in ITS and learning sciences [154, 157].

Practical. The thesis distils evidence-based design guidance for ASL instruction: deploy 3D when objectives genuinely require spatial affordances; couple personalisation with capture and intelligibility guardrails to preserve recogniser reliability [178, 179]; treat gamification as a vehicle for delivering diagnostic information rather than repetition alone [167, 168]; and architect AI support to externalise error detection with concise, per-feature corrections, complemented by reflective artefacts (multi-angle exemplars, slow/segmented replay, side-by-side overlays) [202, 207]. These prescriptions translate general insights from formative assessment and intelligent tutoring to the concrete, multimodal constraints of sign language learning, where handshape, orientation, location, and movement interact in visually subtle ways.

Summary. Rather than replicating known effects, the thesis delineates the boundary conditions under which they hold in ASL learning and isolates a feedback-centric mechanism that unifies them. By integrating matched experimental contrasts with learner-centred qualitative evidence, it explains mixed findings in immersive and gamified learning and provides actionable principles for designing next-generation ASL learning systems grounded in diagnostic specificity and task–affordance fit.

8.5 Applicability, Ecological Validity, and External Validity (Reframed Limitations)

Rather than enumerating omissions, this section articulates where the findings of Chapters 3–7 are most credible, where caution is warranted, and where claims fall outside the validated space. The discussion is organised around scope of inference (construct validity), ecological validity (deployment context), and external validity (generalisation), so that readers can interpret the results in context and assess practical relevance.

Scope of inference (construct validity)

The causal claim examined in this thesis concerns the effect of diagnostically specific feedback—explicit, per-feature guidance on handshape, orientation, location, and movement—relative to presentation-only or non-adaptive materials. Immersion

(VR/3D) and gamification are treated as delivery contexts that shape engagement rather than as mechanisms that directly alter the learning representation. The outcomes analysed were item-level recognition accuracy and response time for entry-level ASL items. These measures capture accuracy and retrieval fluency; they do not index continuous signing, co-articulation, discourse-level production, or intelligibility to native signers. Claims about those competencies lie outside the validated construct space of the present studies.

Ecological validity (deployment context)

All studies were conducted in controlled or semi-controlled environments with guidance on camera framing and lighting. Recognition-based feedback presumes broadly similar conditions. In noisier settings—such as classrooms with variable backgrounds and movement, community spaces with inconsistent illumination, or mobile devices with constrained cameras—both tracking stability and user trust may change. The findings are therefore most applicable where basic environmental quality can be attained or instrumented through setup diagnostics that assist with framing and illumination. The experimental designs used matched dosage and neutral scripts to isolate mechanisms. In authentic classrooms, teacher mediation, peer dynamics, and curriculum pacing are likely to moderate effects by compensating for missing specificity in some cases and amplifying it in others. Translation to instruction should recognise this human layer as part of the causal pathway.

External validity (population and content)

Participants were novice adult learners drawn from a single national context. This design choice reduces confounds from prior ASL knowledge but narrows generalisability. The direction of effects—namely, the value of specific, timely feedback—is likely to be stable across groups, yet the magnitude of benefit and the optimal bandwidth of feedback may differ for younger learners, Deaf communities, mixed-ability cohorts, or professional interpreters. The task domain is centred on digits and frequent static words. Extending conclusions to classifier constructions, complex movement paths, and discourse requires feedback that treats motion and location as first-class targets and may benefit more from 3D affordances such as multi-view inspection and occlusion management. As a result, the present conclusions apply most directly to entry-level lexical items.

Measurement considerations

The joint pattern of increasing accuracy and decreasing response time is consistent with fluency gains rather than a speed–accuracy trade-off. Even so, response time conflates retrieval with motor execution, and additional measures are needed for outcomes that emphasise production naturalness or expressive clarity, including expert ratings or intelligibility to native signers. Equivalent-but-non-identical test forms and temporal spacing were used to mitigate recall, yet repeated testing can still contribute to general improvement and ceiling effects, particularly later in training, which renders between-condition contrasts conservative when baselines are high.

Statistical conclusion validity

Per-condition sample sizes, together with ceiling compression in later rounds, limit precision for small effects and higher-order interactions. Linear mixed-effects models leveraged repeated measures and provided interval estimates, but interaction contrasts should be replicated with larger cohorts before informing high-stakes decisions. Alternative specifications—such as log-transformed response times, tail trimming, and by-participant slopes where convergence permitted—preserved effect directions, although some model dependence remains and replication in independent samples is prudent.

Technology dependence and equity

Feedback quality depends on the stability of detection across hand morphology, skin tone, camera angle, and occlusion. Although offline accuracy exceeded 95% in the present setting and setup guidance was provided, real-time error rates can vary with environment. Systems should therefore surface confidence or uncertainty indicators, provide fall-back explanations when confidence is low, and include setup diagnostics to preserve trust. Differences in hand characteristics and lighting can bias recognition rates; before scale-up, fairness audits by subgroup are recommended to ensure equitable access to high-quality feedback.

Temporal scope

The analyses modelled immediate pre–post change within rounds and improvement across three rounds. Long-term retention, transfer to spontaneous production, and intelligibility to native signers were not directly measured. The claims made here therefore concern acquisition and short-horizon fluency rather than durable mastery. Delayed measures are required for curricular adoption.

Consequences for interpretation

The cross-study pattern supports a mediator view in which immersion and gamification expand access and attention, whereas diagnostically specific feedback is the causal lever that converts practice into dependable gains under the applicability conditions described above. General claims that exceed this scope—such as universal superiority of 3D, unqualified benefits of personalisation, or classroom-wide effects without orchestration—are not warranted by the current evidence.

CHAPTER 9

Conclusion and Future Directions

This chapter consolidates the thesis contributions and sets out a clear, staged trajectory for future research and impact. The argument proceeds from the validated scope of the present findings to a programme of work that addresses the applicability constraints articulated in Chapter 8.5. In doing so, it demonstrates how the current studies form the foundation for a continuing line of scholarly and practical development in sign language learning technologies.

9.1 Concluding synthesis

The empirical chapters have shown that immersive and gamified designs reliably elevate presence, motivation, and persistence, while dependable learning gains are associated with diagnostically specific and timely feedback. Dimensionality offers instrumental value when objectives require spatial affordances; otherwise, a carefully designed 2D interface combined with explicit guidance is sufficient. Personalisation increases autonomy but can degrade feedback fidelity in the absence of capture guardrails. The AI versus non-AI comparison established that externalising error detection and communicating per-feature corrections not only increases accuracy under matched exposure but also reduces response time without inducing a speed–accuracy trade-off. Together, these findings motivate a mechanism-centric perspective in which engagement features function as carriers that expand attention and practice dose, and diagnostic specificity acts as the active ingredient that converts practice into learning.

9.2 From limitations to a research programme

The limitations in Chapter 8.5 identify the precise boundaries of inference and thereby suggest a sequenced agenda for future work.

First, with respect to time, the present studies characterise acquisition and short-horizon fluency; the next step is to trace durability and generalisation. Longitudinal designs with delayed post-tests should assess retention at multiple lags and examine transfer to novel lexical items and spontaneous production [203,204]. To ensure that gains reflect communicative competence rather than test familiarity, future work should incorporate native-signer intelligibility ratings, consistent with formative assessment principles that prioritise performance beyond the immediate task [202,207].

Second, in terms of construct coverage, our item-level recognition and response-time measures index accuracy and retrieval fluency but do not capture co-articulation, movement timing, or discourse-level competence. Extending beyond static items requires feedback that treats motion and location as first-class targets. Methodologically, this implies moving from frame-wise handshape checks to sequence-level diagnostics that highlight path, aperture, contact points, and temporal profiles; practically, it calls for coupling real-time coaching with reflective artefacts—multi-angle exemplars, slow or segmented replay, and side-by-side overlays—so learners can verify corrections at the granularity demanded by continuous signing [210, 212].

Third, task–affordance alignment needs explicit testing. Our matched 2D/3D comparison was intentionally conservative and did not require multi-view reasoning or occlusion resolution. To determine when three-dimensional interfaces yield genuine learning advantages, future tasks should be constructed so that spatial affordances are necessary rather than incidental—for example, classifier morphology, occluded trajectories, and multi-view decision-making—while guidance focuses attention on the relevant spatial features [11, 102].

Fourth, isolating mechanisms warrants factorial ablations. Because the current systems bundle several supports (diagnostic tips, optional replay media, a timed game) to mirror real usage, controlled experiments should disentangle the marginal contribution of each element and their interactions. Contrasts between specific versus generic feedback, with versus without reflective replay, under timed versus untimed conditions, can reveal whether it is hint specificity, verification opportunity, time pressure, or a particular combination that most efficiently drives improvement [202].

Fifth, ecology and equity must be foregrounded. Classroom and community deployments introduce heterogeneous devices, variable lighting, and social dynamics that affect tracking stability, teacher orchestration, and learner trust. Field trials at the class or school level should therefore include teachers in the loop and dashboards that expose uncertainty cues, common confusions, and exemplar generation tools [210]. Robustness and fairness should be treated as first-order outcomes, with subgroup audits (e.g., skin tone, hand morphology, jewellery, tremor, camera quality) built into evaluation protocols and mitigation strategies—adaptive thresholds, environment diagnostics, user-visible confidence indicators—pre-specified and reported [213].

Sixth, statistical precision and distributional effects deserve targeted designs. Some contrasts approached ceiling and late-round variance patterns were non-uniform. Subsequent studies should be powered for interactions and planned to characterise not only mean changes but also dispersion and tails. Policies that adapt hint bandwidth, pacing, and targeted drills should be evaluated on their ability to lift the lower tail without suppressing the upper. Pre-registration, blinded rating procedures for production outcomes, and shared analysis plans will strengthen inference and support cumulative science [214–216].

Finally, open science and impact pathways can accelerate progress. Reference implementations of diagnostic-feedback primitives—APIs for per-feature hints, uncertainty exposure, and exemplar generation—together with de-identified interaction logs, item banks, and analysis notebooks would enable replication and fair comparison across laboratories and classrooms [217]. Partnerships with educators, Deaf community organisations, and public libraries can translate research prototypes into accessible learning kits and teacher-facing formative assessment tools, accompanied from the outset by ethical safeguards around privacy, consent, and appropriate use [210].

9.3 Anticipated contributions and milestones

The near-term outcome of the programme is an evidence-based account of retention and transfer under AI-assisted feedback, expressed through delayed accuracy, response time, and intelligibility measures. In the medium term, ablation studies and spatially demanding tasks will clarify when three-dimensional interfaces are warranted and how reflective replay should be sequenced with real-time coaching. In the longer term, classroom trials with teacher orchestration and fairness audits will establish ecological validity and equity, producing guidelines and open artefacts that others can adopt. Across these stages, the core theoretical contribution remains intact: diagnostic specificity is the principal lever for turning practice into progress; engagement features are most powerful when they help deliver and focus such feedback.

9.4 Closing statement

The present thesis establishes a mechanism-centred foundation for ASL learning technologies. By treating immersion and gamification as carriers and diagnostic specificity as the active ingredient, it reconciles mixed findings in the literature and yields actionable principles for design. The proposed trajectory extends this foundation to durable learning, continuous signing, ecological robustness, and equitable access. In doing so, it situates the work within a continuing programme of scholarly inquiry and practical development that can meaningfully advance accessible sign language education.

Bibliography

[1] Y. Chamekh and M. Hammami, "Impact of virtual reality on modern education," *International Journal of Sciences: Basic and Applied Research*, vol. 50, no. 2, pp. 1–8, 2020. 1.1

[2] F. Faiella and M. Ricciardi, "Gamification and learning: a review of issues and research," *Journal of e-learning and knowledge society*, vol. 11, no. 3, 2015. 1.1

[3] W. C. Stokoe Jr, "Sign language structure: An outline of the visual communication systems of the american deaf," *Journal of deaf studies and deaf education*, vol. 10, no. 1, pp. 3–37, 2005. 1.1

[4] H.-K. Wu, S. W.-Y. Lee, H.-Y. Chang, and J.-C. Liang, "Current status, opportunities and challenges of augmented reality in education," *Computers & education*, vol. 62, pp. 41–49, 2013. 1.1

[5] Y.-J. Lan, "Immersion into virtual reality for language learning," in *Psychology of learning and motivation*, vol. 72, pp. 1–26, Elsevier, 2020. 1.1

[6] M. Raja and G. Lakshmi Priya, "Using virtual reality and augmented reality with ict tools for enhancing quality in the changing academic environment in covid-19 pandemic: An empirical study," in *Technologies, artificial intelligence and the future of learning post-COVID-19: The crucial role of international accreditation*, pp. 467–482, Springer, 2022. 1.1

[7] J. Willcott, "Pandemic and post-pandemic use of immersive learning technology," in *Handbook of research on teaching with virtual environments and AI*, pp. 1–16, IGI Global, 2021. 1.1

[8] R. S. Alsawaier, "The effect of gamification on motivation and engagement," *The International Journal of Information and Learning Technology*, vol. 35, no. 1, pp. 56–79, 2018. 1.1

[9] R. Jackendoff, "Précis of foundations of language: Brain, meaning, grammar, evolution," *Behavioral and brain sciences*, vol. 26, no. 6, pp. 651–665, 2003. 1.1

[10] T. Govender and J. Arnedo-Moreno, "An analysis of game design elements used in digital game-based language learning," *Sustainability*, vol. 13, no. 12, p. 6679, 2021. 1.1

[11] B. Dalgarno and M. J. Lee, "What are the learning affordances of 3-d virtual environments?," *British journal of educational technology*, vol. 41, no. 1, pp. 10–32, 2010. 1.1, 6.1, 6.5, 8.4, 8.4, 9.2

[12] D. Bragg, N. Caselli, J. W. Gallagher, M. Goldberg, C. J. Oka, and W. Thies, "Asl sea battle: gamifying sign language data collection," in *Proceedings of the 2021 CHI conference on human factors in computing systems*, pp. 1–13, 2021. 1.1, 2.4, 3.1, 4.1, 5.1

[13] J. Wang, I. Ivrissimtzis, Z. Li, Y. Zhou, and L. Shi, "Exploring the potential of immersive virtual environments for learning american sign language," in *European Conference on Technology Enhanced Learning*, pp. 459–474, Springer, 2023. 1.3.1, 1.6

[14] J. Wang, I. Ivrissimtzis, Z. Li, Y. Zhou, and L. Shi, "Developing and evaluating a novel gamified virtual learning environment for ASL," in *INTERACT 2023*, LNCS, Springer, 2023. 1.3.1, 1.6, 3.1

[15] J. Wang, I. Ivrissimtzis, Z. Li, Y. Zhou, and L. Shi, "User-defined hand gesture interface to improve user experience of learning american sign language," in *International Conference on Intelligent Tutoring Systems*, pp. 479–490, Springer, 2023. 1.3.1, 1.6, 3.2

[16] J. Wang, I. Ivrissimtzis, Z. Li, and L. Shi, "Comparative efficacy of 2d and 3d virtual reality games in american sign language learning," in *The 31st IEEE Conference on Virtual Reality and 3D User Interfaces*, Newcastle University, 2024. 1.3.1, 1.6

[17] L. Quandt, "Teaching asl signs using signing avatars and immersive learning in virtual reality," in *Proceedings of the 22nd international ACM SIGACCESS conference on computers and accessibility*, pp. 1–4, 2020. 2.3, 2.6

[18] D. Kamińska, G. Zwoliński, and A. Laska-Leśniewicz, "Usability testing of virtual reality applications—the pilot study," *Sensors*, vol. 22, no. 4, p. 1342, 2022. 2.3

[19] X. Xu, J. Kang, and L. Yan, "Understanding embodied immersion in technology-enabled embodied learning environments," *Journal of Computer Assisted Learning*, vol. 38, no. 1, pp. 103–119, 2022. 2.3, 2.5, 8.4, 8.4, 8.4, 8.4

[20] T. Xie, Y. Li, and Y. Tang, "Effects of using immersive virtual reality for science education on learning outcomes: A randomized controlled pilot study,"

IEEE Transactions on Learning Technologies, vol. 16, no. 6, pp. 1045–1056, 2023. 2.3

[21] D. Economou, M. G. Russi, I. Doumanis, M. Mentzelopoulos, V. Bouki, and J. Ferguson, “Using serious games for learning british sign language combining video, enhanced interactivity, and vr technology,” *Journal of Universal Computer Science*, vol. 26, no. 8, pp. 996–1016, 2020. 2.3, 2.4, 3.2, 7.2.3

[22] P. Mafor and U. Ramnarain, “Investigating the role of gesture and embodiment in natural sciences learning using immersive virtual reality,” in *International Conference on Computers in Education*, pp. 472–480, 2022. 2.3

[23] E. Mohd Mokhtar, *Investigating User Experience Using Gesture-based and Immersive-based Interfaces on Animation Learners*. PhD thesis, Goldsmiths, University of London, 2023. 2.3

[24] A. Miglani, S. Kidambi, and P. Mareguddi, “Play to improve: Gamifying usability evaluations in virtual reality,” in *International Conference on Human-Computer Interaction*, pp. 248–266, Springer, 2020. 2.3

[25] C. Uz Bilgin and A. Gul, “Investigating the effectiveness of gamification on group cohesion, attitude, and academic achievement in collaborative learning environments,” *TechTrends*, vol. 64, no. 1, pp. 124–136, 2020. 2.4, 8.4, 8.4, 8.4

[26] L.-M. Putz, F. Hofbauer, and H. Treiblmaier, “Can gamification help to improve education? findings from a longitudinal study,” *Computers in Human Behavior*, vol. 110, p. 106392, 2020. 2.4

[27] I. Bouchrika, N. Harrati, V. Wanick, and G. Wills, “Exploring the impact of gamification on student engagement and involvement with e-learning systems,” *Interactive Learning Environments*, vol. 29, no. 8, pp. 1244–1257, 2021. 2.4

[28] A.-I. Zourmpakis, M. Kalogiannakis, and S. Papadakis, “Adaptive gamification in science education: An analysis of the impact of implementation and adapted game elements on students’ motivation,” *Computers*, vol. 12, no. 7, p. 143, 2023. 2.4, 8.4, 8.4, 8.4

[29] K. Ishaq, N. A. M. Zin, F. Rosdi, M. Jehanghir, S. Ishaq, and A. Abid, “Mobile-assisted and gamification-based language learning: a systematic literature review,” *PeerJ Computer Science*, vol. 7, p. e496, 2021. 2.4

[30] B. Spittle, M. Frutos-Pascual, C. Creed, and I. Williams, “A review of interaction techniques for immersive environments,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 29, no. 9, pp. 3900–3921, 2022. 2.5

[31] B. Joksimoski, E. Zdravevski, P. Lameski, I. M. Pires, F. J. Melero, T. P. Martinez, N. M. Garcia, M. Mihajlov, I. Chorbev, and V. Trajkovik, “Technological solutions for sign language recognition: a scoping review of research trends, challenges, and opportunities,” *IEEE Access*, vol. 10, pp. 40979–40998, 2022. 2.5

[32] S. Sharma and S. Singh, "Vision-based hand gesture recognition using deep learning for the interpretation of sign language," *Expert Systems with Applications*, vol. 182, p. 115657, 2021. 2.5

[33] M. Al-Hammadi, G. Muhammad, W. Abdul, M. Alsulaiman, M. A. Bencherif, and M. A. Mekhtiche, "Hand gesture recognition for sign language using 3dcnn," *IEEE access*, vol. 8, pp. 79491–79509, 2020. 2.5

[34] H. Xia, M. Glueck, M. Annett, M. Wang, and D. Wigdor, "Iteratively designing gesture vocabularies: A survey and analysis of best practices in the hci literature," *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 29, no. 4, pp. 1–54, 2022. 2.5

[35] L. Oppici, B. Mathias, S. Narciss, and A. Proske, "Benefits of enacting and observing gestures on foreign language vocabulary learning: A systematic review and meta-analysis," *Behavioral Sciences*, vol. 13, no. 11, p. 920, 2023. 2.5

[36] A. B. García-Gámez and P. Macizo, "Gestures as scaffolding to learn vocabulary in a foreign language," *Brain sciences*, vol. 13, no. 12, p. 1712, 2023. 2.5

[37] G. B. Petersen, V. Stenberdt, R. E. Mayer, and G. Makransky, "Collaborative generative learning activities in immersive virtual reality increase learning," *Computers & Education*, vol. 207, p. 104931, 2023. 2.6

[38] M. Thompson, C. Uz-Bilgin, M. S. Tutwiler, M. Anteneh, J. C. Meija, A. Wang, P. Tan, R. Eberhardt, D. Roy, J. Perry, *et al.*, "Immersion positively affects learning in virtual reality games compared to equally interactive 2d games," *Information and Learning Sciences*, vol. 122, no. 7/8, pp. 442–463, 2021. 2.6, 8.4, 8.4, 8.4, 8.4

[39] F.-C. Yang, C. Mousas, and N. Adamo, "Holographic sign language avatar interpreter: A user interaction study in a mixed reality classroom," *Computer Animation and Virtual Worlds*, vol. 33, no. 3-4, p. e2082, 2022. 2.6

[40] K. Secora and K. Emmorey, "Visual-spatial perspective-taking in spatial scenes and in american sign language," *The Journal of Deaf Studies and Deaf Education*, vol. 25, no. 4, pp. 447–456, 2020. 2.6

[41] J. Shen, Y. Wang, C. Chen, M. R. Nelson, and M. Z. Yao, "Using virtual reality to promote the university brand: When do telepresence and system immersion matter?," *Journal of Marketing Communications*, vol. 26, no. 4, pp. 362–393, 2020. 2.6

[42] T. Nagarajan and K. Grauman, "Learning affordance landscapes for interaction exploration in 3d environments," *Advances in Neural Information Processing Systems*, vol. 33, pp. 2005–2015, 2020. 2.6, 8.4, 8.4, 8.4

[43] M. Cabanillas-Carbonell, P. Cusi-Ruiz, D. Prudencio-Galvez, and J. L. Herrera Salazar, "Mobile application with augmented reality to improve the process of learning sign language," 2022. 2.6

[44] E. Kochmar, D. D. Vu, R. Belfer, V. Gupta, I. V. Serban, and J. Pineau, "Automated personalized feedback improves learning gains in an intelligent tutoring system," in *International conference on artificial intelligence in education*, pp. 140–146, Springer, 2020. 2.7

[45] T. Rasul, S. Nair, D. Kalendra, M. Robin, F. de Oliveira Santini, W. J. Ladeira, M. Sun, I. Day, R. A. Rather, and L. Heathcote, "The role of chatgpt in higher education: Benefits, challenges, and future research directions," *Journal of Applied Learning and Teaching*, vol. 6, no. 1, pp. 41–56, 2023. 2.7

[46] Z. Wei and M. Yuan, "Research on the current situation and future development trend of immersive virtual reality in the field of education," *Sustainability*, vol. 15, no. 9, p. 7531, 2023. 2.7

[47] J. Shin, A. S. Musa Miah, M. A. M. Hasan, K. Hirooka, K. Suzuki, H.-S. Lee, and S.-W. Jang, "Korean sign language recognition using transformer-based deep neural network," *Applied Sciences*, vol. 13, no. 5, p. 3029, 2023. 2.7

[48] N. Panchamia, J. Mehta, P. Ghosh, and J. Mehta, "Asl tutor using deep learning," in *2021 International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON)*, pp. 1–6, IEEE, 2021. 2.7

[49] D. Sharma, N. Reddy, P. Gupta, and R. Sharma, "Enhancing customer experience personalization through ai: Leveraging collaborative filtering, neural networks, and natural language processing," *Journal of AI ML Research*, vol. 11, no. 7, 2022. 2.7

[50] R. Zekaj, "Ai language models as educational allies: Enhancing instructional support in higher education," *International Journal of Learning, Teaching and Educational Research*, vol. 22, no. 8, pp. 120–134, 2023. 2.7

[51] G. Mustafa and M. Z. I. Chaudhary, "Enhancing educational experiences: Integrating virtual reality and artificial intelligence," *Jahan-e-Tahqeeq*, vol. 6, no. 4, pp. 592–598, 2023. 2.7

[52] N. A. A. Empe, R. C. L. Echon, H. D. A. Vega, P. L. C. Paterno, M. N. Jamis, and E. R. Yabut, "Simbowika: A mobile and web application to learn filipino sign language for deaf students in elementary schools," in *(R10-HTC)*, pp. 1–6, IEEE, 2020. 3.1, 3.2, 5.1

[53] I. Estrada-Cota, M. A. Carreño-León, J. A. Sandoval-Bringas, A. A. Leyva-Carrillo, and H. X. C. Quiroz, "Design of a web tool for teaching–learning of states and capitals of méxico through the mexican sign language," in *International Conference on Inclusive Technology and Education (CONTIE)*, pp. 174–179, IEEE, 2021. 3.1, 3.2, 5.1

[54] A. Patricks, "Developing an accessible learning application for sign language (c)," 2022. 3.1, 3.2, 5.1

[55] N. Adamo-Villani, E. Carpenter, and L. Arns, "An immersive virtual environment for learning sign language mathematics," in *ACM SIGGRAPH 2006 educators program*, pp. 20–es, 2006. 3.1, 5.1

[56] J. Schioppo, Z. Meyer, D. Fabiano, and S. Canavan, "Sign language recognition: Learning american sign language in a virtual environment," in *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, pp. 1–6, 2019. 3.1, 5.1

[57] H. D. Phan, K. Ellis, A. Dorin, and P. Olivier, "Feedback strategies for embodied agents to enhance sign language vocabulary learning," in *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents*, pp. 1–8, 2020. 3.1, 5.1

[58] M. Schrepp, A. Hinderks, and J. Thomaschewski, "Applying the user experience questionnaire (UEQ) in different evaluation scenarios," in *International Conference of Design, User Experience, and Usability*, pp. 383–392, Springer, 2014. 3.1, 3.4, 4.3, 5.1, 5.4, 5.6, 6.1, 6.3.2

[59] V. Bheda and D. Radpour, "Using deep convolutional networks for gesture recognition in american sign language," *arXiv:1710.06836*, 2017. 3.2, 5.2.1

[60] S. Kim, Y. Ji, and K.-B. Lee, "An effective sign language learning with object detection based ROI segmentation," in *Proc. IRC*, pp. 330–333, IEEE, 2018. 3.2, 5.2.1

[61] P. Battistoni, M. Di Gregorio, M. Sebillo, and G. Vitiello, "Ai at the edge for sign language learning support," in *Proc. HCC*, pp. 16–23, IEEE, 2019. 3.2, 5.2.1

[62] X. Jiang, B. Hu, S. Chandra Satapathy, S.-H. Wang, and Y.-D. Zhang, "Fingerspelling identification for chinese sign language via alexnet-based transfer learning and adam optimizer," *Scientific Programming*, vol. 2020, 2020. 3.2, 5.2.1

[63] N. C. Camgoz, O. Koller, S. Hadfield, and R. Bowden, "Sign language transformers: Joint end-to-end sign language recognition and translation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10023–10033, 2020. 3.2, 5.2.1

[64] F. Zhang, V. Bazarevsky, A. Vakunov, A. Tkachenka, G. Sung, C.-L. Chang, and M. Grundmann, "Mediapipe hands: On-device real-time hand tracking," *arXiv preprint arXiv:2006.10214*, 2020. 3.2, 5.2.1, 6.2

[65] T. Goswami and S. R. Javaji, "Cnn model for american sign language recognition," in *ICCCE 2020*, pp. 55–61, Springer, 2021. 3.2, 5.2.1

[66] P. Pallavi and D. Sarvamangala, "Recognition of sign language using deep neural network," *Int. J. of Advanced Research in Computer Science*, vol. 12, pp. 92–97, 2021. 3.2, 5.2.1

[67] S. S. Kumar, T. Wangyal, V. Saboo, and R. Srinath, "Time series neural networks for real time sign language translation," in *Proc. ICMLA*, pp. 243–248, IEEE, 2018. 3.2

[68] J. Joy, K. Balakrishnan, and M. Sreeraj, "Signquiz: a quiz based tool for learning fingerspelled signs in indian sign language using ASLR," *IEEE Access*, vol. 7, pp. 28363–28371, 2019. 3.2

[69] A. Vaitkevičius, M. Taroza, T. Blažauskas, R. Damaševičius, R. Maskeliūnas, and M. Woźniak, "Recognition of American sign language gestures in a virtual reality using leap motion," *Applied Sciences*, vol. 9, no. 3, p. 445, 2019. 3.2

[70] J. J. Bird, A. Ekárt, and D. R. Faria, "British sign language recognition via late fusion of computer vision and leap motion with transfer learning to american sign language," *Sensors*, vol. 20, no. 18, p. 5151, 2020. 3.2

[71] A. John, R. H. Krishnan, and A. M. Vinitha, *Language Recognition System: An Application Based Study with Special Reference to Sociolinguistics and Computational Linguistics.* 2021. 3.2

[72] K. Bantupalli and Y. Xie, "American sign language recognition using deep learning and computer vision," in *Proc. ICBD*, pp. 4896–4899, IEEE, 2018. 3.2

[73] J. Schnepp, R. Wolfe, G. Brionez, S. Baowidan, R. Johnson, and J. McDonald, "Human-centered design for a sign language learning application," in *Proc. PETRAE*, pp. 1–5, 2020. 3.2

[74] M. J. C. Samonte, "An assistive technology using fsl, speech recognition, gamification and online handwritten character recognition in learning statistics for students with hearing and speech impairment," in *Proc. ICFET*, pp. 92–97, 2020. 3.2

[75] J.-H. Park and H. J. Choi, "Factors influencing adult learners' decision to drop out or persist in online learning," *Journal of Educational Technology & Society*, vol. 12, no. 4, pp. 207–217, 2009. 3.2

[76] I. Reisoğlu, B. Topu, R. Yılmaz, T. Karakuş Yılmaz, and Y. Göktaş, "3d virtual learning environments in education: A meta-review," *Asia Pacific Education Review*, vol. 18, pp. 81–100, 2017. 3.2

[77] G. Bradski and A. Kaehler, "Opencv," *DDJ of software tools*, vol. 3, p. 120, 2000. 3.3.1, 4.2, 5.3.1, 6.2

[78] M. Schrepp, J. Thomaschewski, and A. Hinderks, "Construction of a benchmark for the user experience questionnaire (UEQ)," *International Journal of*

Interactive Multimedia and Artificial Intelligence, vol. 4, no. 4, pp. 40–44, 2017. 3.5, 3.5, 4.3, 6.4.2, 6.5

[79] E. Bialystok *et al.*, *Bilingualism in development: Language, literacy, and cognition.* Cambridge University Press, 2001. 4.1

[80] W. Python, "Python," *Python Releases for Windows*, vol. 24, 2021. 4.2, 5.3.1, 6.2

[81] J. V. Dillon, I. Langmore, D. Tran, E. Brevdo, S. Vasudevan, D. Moore, B. Patton, A. Alemi, M. Hoffman, and R. A. Saurous, "Tensorflow distributions," *arXiv preprint arXiv:1711.10604*, 2017. 4.2, 5.3.1, 6.2

[82] T. Piumsomboon, A. Clark, M. Billinghurst, and A. Cockburn, "User-defined gestures for augmented reality," in *IFIP Conference on HCI*, pp. 282–299, Springer, 2013. 5.1, 8.4

[83] H. Wu, Y. Wang, J. Qiu, J. Liu, and X. Zhang, "User-defined gesture interaction for immersive vr shopping applications," *BIT*, vol. 38, no. 7, pp. 726–741, 2019. 5.2.2

[84] H. Wu, Y. Wang, J. Liu, J. Qiu, and X. L. Zhang, "User-defined gesture interaction for in-vehicle information systems," *Multimed. Tools Appl.*, vol. 79, no. 1, pp. 263–288, 2020. 5.2.2

[85] J. I. Koh, J. Cherian, P. Taele, and T. Hammond, "Developing a hand gesture recognition system for mapping symbolic hand gestures to analogous emojis in computer-mediated communication," *ACM TiiS*, vol. 9, no. 1, pp. 1–35, 2019. 5.2.2

[86] Q. Dai, X. Li, W. Geng, W. Jin, and X. Liang, "Capg-myo: A muscle-computer interface supporting user-defined gesture recognition," in *Proceedings of the 9th ICCCM*, pp. 52–58, 2021. 5.2.2, 8.4

[87] Y. Takayama, Y. Ichikawa, B. Shizuki, I. Kawaguchi, and S. Takahashi, "A user-based mid-air hand gesture set for spreadsheets," in *Asian CHI Symposium*, pp. 122–128, 2021. 5.2.2, 8.4

[88] E. L. Newport and R. P. Meier, "The acquisition of american sign language," in *The crosslinguistic study of language acquisition*, pp. 881–938, Psychology Press, 2017. 6.1

[89] B. L. Freel, M. D. Clark, M. L. Anderson, G. L. Gilbert, M. M. Musyoka, and P. C. Hauser, "Deaf individuals' bilingual abilities: American sign language proficiency, reading skills, and family characteristics," *Psychology*, vol. 2, no. 1, pp. 18–23, 2011. 6.1

[90] D. Quinto-Pozos, J. L. Singleton, and P. C. Hauser, "A case of specific language impairment in a deaf signer of american sign language," *The Journal of Deaf Studies and Deaf Education*, vol. 22, no. 2, pp. 204–218, 2017. 6.1

[91] D. Oluwajana, A. Idowu, M. Nat, V. Vanduhe, and S. Fadiya, “The adoption of students’ hedonic motivation system model to gamified learning environment,” *Journal of theoretical and applied electronic commerce research*, vol. 14, no. 3, pp. 156–167, 2019. 6.1

[92] P. Buckley and E. Doyle, “Gamification and student motivation,” *Interactive learning environments*, vol. 24, no. 6, pp. 1162–1175, 2016. 6.1

[93] F. Roosta, F. Taghiyareh, and M. Mosharraf, “Personalization of gamification-elements in an e-learning environment based on learners’ motivation,” in *2016 8th International symposium on telecommunications (IST)*, pp. 637–642, IEEE, 2016. 6.1

[94] S. M. Shohieb, “A gamified e-learning framework for teaching mathematics to arab deaf students: Supporting an acting arabic sign language avatar,” *Ubiquitous Learning: An International Journal*, vol. 12, no. 1, pp. 55–70, 2019. 6.1

[95] A. Antonaci, R. Klemke, and M. Specht, “The effects of gamification in online learning environments: A systematic literature review,” *Informatics*, vol. 6, no. 3, p. 32, 2019. 6.1

[96] L. Shi, A. I. Cristea, S. Hadzidedic, and N. Dervishalidovic, “Contextual gamification of social interaction – towards increasing motivation in social e-learning,” in *Advances in Web-Based Learning – ICWL 2014* (E. Popescu, R. W. H. Lau, K. Pata, H. Leung, and M. Laanpere, eds.), (Cham), pp. 116–122, Springer International Publishing, 2014. 6.1

[97] L. Shi and A. I. Cristea, “Motivational gamification strategies rooted in self-determination theory for social adaptive e-learning,” in *Intelligent Tutoring Systems* (A. Micarelli, J. Stamper, and K. Panourgia, eds.), (Cham), pp. 294–300, Springer International Publishing, 2016. 6.1

[98] Z. Zafrulla, H. Brashear, P. Presti, H. Hamilton, and T. Starner, “Copycat: an american sign language game for deaf children,” in *Face and Gesture 2011*, pp. 647–647, IEEE Computer Society, 2011. 6.1

[99] H. P. Pontes, J. B. F. Duarte, and P. R. Pinheiro, “An educational game to teach numbers in brazilian sign language while having fun,” *Computers in Human Behavior*, vol. 107, p. 105825, 2020. 6.1

[100] D. Economou, M. Mentzelopoulos, J. Ingram, T. Martinez-Mukimov, S. Primkulova, and S. Abduvaliev, “Work-in-progress-gamifying the process of learning sign language in vr,” in *2022 8th International Conference of the Immersive Learning Research Network (iLRN)*, pp. 1–3, IEEE, 2022. 6.1

[101] N. Adamo-Villani and K. Wright, “Smile: an immersive learning game for deaf and hearing children,” in *ACM SIGGRAPH 2007 educators program*, pp. 17–es, ACM Press, 2007. 6.1

[102] R. E. Mayer, "Multimedia learning," in *Psychology of learning and motivation*, vol. 41, pp. 85–139, Elsevier, 2002. 6.1, 6.5, 8.4, 8.4, 9.2

[103] M. Slater, S. Wilbur, *et al.*, "A framework for immersive virtual environments (five): Speculations on the role of presence in virtual environments," *Presence: Teleoperators and virtual environments*, vol. 6, no. 6, pp. 603–616, 1997. 6.1

[104] D. Benyon, *Designing interactive systems: A comprehensive guide to HCI, UX and interaction design*, vol. 3. Pearson Harlow, 2014. 6.1

[105] A. Bangor, P. Kortum, and J. Miller, "Determining what individual sus scores mean: Adding an adjective rating scale," *Journal of usability studies*, vol. 4, no. 3, pp. 114–123, 2009. 6.1, 6.3.2, 6.4.1

[106] T. A. Mikropoulos and A. Natsis, "Educational virtual environments: A ten-year review of empirical research (1999–2009)," *Computers & Education*, vol. 56, no. 3, pp. 769–780, 2011. 6.1, 6.5

[107] D. Wood, J. S. Bruner, and G. Ross, "The role of tutoring in problem solving," *Journal of child psychology and psychiatry*, vol. 17, no. 2, pp. 89–100, 1976. 6.3.1

[108] S. Senn, K. Rolfe, and S. A. Julious, "Investigating variability in patient response to treatment–a case study from a replicate cross-over study," *Statistical methods in medical research*, vol. 20, no. 6, pp. 657–666, 2011. 6.3.1

[109] O. M. Al-Hassan, K. E. Bani-Hani, and M. M. Al-Masa'deh, "Sign language as a means of inclusion: A case study," *International Journal of Disability, Development and Education*, vol. 71, no. 6, pp. 987–1001, 2024. 7.1

[110] M. De Meulder and H. Haualand, "Sign language interpreting services: A quick fix for inclusion?," *Translation and Interpreting Studies*, vol. 16, no. 1, pp. 19–40, 2021. 7.1

[111] J. K. Fernandes and S. S. Myers, "Inclusive deaf studies: Barriers and pathways," *Journal of Deaf Studies and Deaf Education*, vol. 15, no. 1, pp. 17–29, 2010. 7.1

[112] P. Martins, H. Rodrigues, T. Rocha, M. Francisco, and L. Morgado, "Accessible options for deaf people in e-learning platforms: technology solutions for sign language translation," *Procedia Computer Science*, vol. 67, pp. 263–272, 2015. 7.1

[113] K. Karpouzis, G. Caridakis, S.-E. Fotinea, and E. Efthimiou, "Educational resources and implementation of a greek sign language synthesis architecture," *Computers & Education*, vol. 49, no. 1, pp. 54–74, 2007. 7.1

[114] W. Aljedaani, R. Krasniqi, S. Aljedaani, M. W. Mkaouer, S. Ludi, and K. Al-Raddah, "If online learning works for you, what about deaf students? emerging challenges of online learning for deaf and hearing-impaired students during

covid-19: a literature review," *Universal access in the information society*, vol. 22, no. 3, pp. 1027–1046, 2023. 7.1

[115] K. Papadimitriou, G. Potamianos, G. Sapountzaki, T. Goulas, E. Efthimiou, S.-E. Fotinea, and P. Maragos, "Greek sign language recognition for an education platform," *Universal Access in the Information Society*, vol. 24, no. 1, pp. 51–68, 2025. 7.1

[116] R. McKee and S. Pivac Alexander, "Teaching sign language remotely in the covid-19 pandemic," *The Language Learning Journal*, vol. 52, no. 5, pp. 597–613, 2024. 7.1

[117] T. Haug, "Review of sign language assessment instruments," *Sign Language & Linguistics*, vol. 8, no. 1-2, pp. 61–98, 2005. 7.1

[118] J. E. Llamazares de Prado, "Sign language teaching technological advances and differences in international contexts," *The International Journal of Information and Learning Technology*, vol. 38, no. 5, pp. 433–453, 2021. 7.1

[119] D. David, A. H. Alamoodi, O. S. Albahri, S. Garfan, A. S. Albahri, B. Zaidan, and J. Chen, "Sign language mobile apps: a systematic review of current app evaluation progress and solution framework," *Evolving Systems*, vol. 15, no. 2, pp. 669–686, 2024. 7.1, 7.2.1

[120] I. Papastratis, C. Chatzikonstantinou, D. Konstantinidis, K. Dimitropoulos, and P. Daras, "Artificial intelligence technologies for sign language," *Sensors*, vol. 21, no. 17, p. 5843, 2021. 7.1

[121] R. I. Mayberry, "The importance of childhood to language acquisition: Evidence from american sign language," *The development of speech perception*, pp. 57–90, 1994. 7.1

[122] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, *et al.*, "Language models are few-shot learners," *Advances in neural information processing systems*, vol. 33, pp. 1877–1901, 2020. 7.1

[123] J. Achiam, S. Adler, S. Agarwal, L. Ahmad, I. Akkaya, F. L. Aleman, D. Almeida, J. Altenschmidt, S. Altman, S. Anadkat, *et al.*, "Gpt-4 technical report," *arXiv preprint arXiv:2303.08774*, 2023. 7.1

[124] W. X. Zhao, K. Zhou, J. Li, T. Tang, X. Wang, Y. Hou, Y. Min, B. Zhang, J. Zhang, Z. Dong, *et al.*, "A survey of large language models," *arXiv preprint arXiv:2303.18223*, vol. 1, no. 2, 2023. 7.1

[125] K. Guo, E. D. Zhang, D. Li, and S. Yu, "Using ai-supported peer review to enhance feedback literacy: An investigation of students' revision of feedback on peers' essays," *British Journal of Educational Technology*, vol. 56, no. 4, pp. 1612–1639, 2025. 7.1

[126] A. N. Ansari, S. Ahmad, and S. M. Bhutta, "Mapping the global evidence around the use of chatgpt in higher education: A systematic scoping review," *Education and Information Technologies*, vol. 29, no. 9, pp. 11281–11321, 2024. 7.1

[127] A. S. E. AbuSahyon, A. Alzyoud, O. Alshorman, and B. Al-Absi, "Ai-driven technology and chatbots as tools for enhancing english language learning in the context of second language acquisition: A review study," *International Journal of Membrane Science and Technology*, vol. 10, no. 1, pp. 1209–1223, 2023. 7.1

[128] H. S. Nwana, "Intelligent tutoring systems: an overview," *Artificial Intelligence Review*, vol. 4, no. 4, pp. 251–277, 1990. 7.1

[129] S. Pal Chowdhury, V. Zouhar, and M. Sachan, "Autotutor meets large language models: A language model tutor with rich pedagogy and guardrails," in *Proceedings of the Eleventh ACM Conference on Learning@ Scale*, pp. 5–15, 2024. 7.1

[130] R. Sreemathy, M. Turuk, I. Kulkarni, and S. Khurana, "Sign language recognition using artificial intelligence," *Education and Information Technologies*, vol. 28, no. 5, pp. 5259–5278, 2023. 7.1

[131] P. Barve, N. Mutha, A. Kulkarni, Y. Nigudkar, and Y. Robert, "Application of deep learning techniques on sign language recognition—a survey," *Data Management, Analytics and Innovation: Proceedings of ICDMAI 2021, Volume 1*, pp. 211–227, 2021. 7.1

[132] A. Coy, P. S. Mohammed, and P. Skerrit, "Inclusive deaf education enabled by artificial intelligence: The path to a solution," *International Journal of Artificial Intelligence in Education*, vol. 35, no. 1, pp. 96–134, 2025. 7.1

[133] P. Sharma, D. Tulsian, C. Verma, P. Sharma, and N. Nancy, "Translating speech to indian sign language using natural language processing," *Future internet*, vol. 14, no. 9, p. 253, 2022. 7.1

[134] D. Lillo-Martin and J. Henner, "Acquisition of sign languages," *Annual review of linguistics*, vol. 7, no. 1, pp. 395–419, 2021. 7.1

[135] K. Gournaris, "Adjusting to change: learning american sign language online during a global pandemic," *Northwest Journal of Teacher Education*, vol. 17, no. 2, p. 6, 2022. 7.1

[136] R. I. Mayberry and B. Squires, "Sign language acquisition," *Encyclopedia of language and linguistics*, vol. 11, pp. 739–43, 2006. 7.2.1

[137] F. I. Saman, N. F. M. Shariff, and N. I. S. Nasaruddin, "i-sign: Sign language learning application via gamification.," *Asian Journal of University Education*, vol. 15, no. 3, pp. 187–197, 2019. 7.2.1

[138] O. Koller, J. Forster, and H. Ney, "Continuous sign language recognition: Towards large vocabulary statistical recognition systems handling multiple signers," *Computer Vision and Image Understanding*, vol. 141, pp. 108–125, 2015. 7.2.1

[139] N. C. Camgoz, S. Hadfield, O. Koller, H. Ney, and R. Bowden, "Neural sign language translation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7784–7793, 2018. 7.2.1

[140] X. Chai, G. Li, Y. Lin, Z. Xu, Y. Tang, X. Chen, and M. Zhou, "Sign language recognition and translation with kinect," in *IEEE conf. on AFGR*, vol. 655, p. 4, 2013. 7.2.1

[141] H. R. V. Joze and O. Koller, "Ms-asl: A large-scale data set and benchmark for understanding american sign language," *arXiv preprint arXiv:1812.01053*, 2018. 7.2.1

[142] L. Pigou, S. Dieleman, P.-J. Kindermans, and B. Schrauwen, "Sign language recognition using convolutional neural networks," in *European conference on computer vision*, pp. 572–578, Springer, 2014. 7.2.1

[143] D. Kothadiya, C. Bhatt, K. Sapariya, K. Patel, A.-B. Gil-González, and J. M. Corchado, "Deepsign: Sign language detection and recognition using deep learning," *Electronics*, vol. 11, no. 11, p. 1780, 2022. 7.2.1

[144] M. J. Cheok, Z. Omar, and M. H. Jaward, "A review of hand gesture and sign language recognition techniques," *International Journal of Machine Learning and Cybernetics*, vol. 10, no. 1, pp. 131–153, 2019. 7.2.1

[145] M. Vázquez-Enríquez, J. L. Alba-Castro, L. Docío-Fernández, and E. Rodríguez-Banga, "Isolated sign language recognition with multi-scale spatial-temporal graph convolutional networks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 3462–3471, 2021. 7.2.1

[146] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime multi-person 2d pose estimation using part affinity fields," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7291–7299, 2017. 7.2.1

[147] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C.-L. Chang, M. G. Yong, J. Lee, *et al.*, "Mediapipe: A framework for building perception pipelines," *arXiv preprint arXiv:1906.08172*, 2019. 7.2.1

[148] E. Boers-Visker and B. Van Den Bogaerde, "Learning to use space in the l2 acquisition of a signed language: Two case studies," *Sign Language Studies*, vol. 19, no. 3, pp. 410–452, 2019. 7.2.1

[149] J. C. Salazar Guevara and J. C. Sánchez Ramírez, "Leap signvr: aplicación basada en realidad virtual para el apoyo de la enseñanza de la lengua de señas colombiana usando leap motion.," 2018. 7.2.1

[150] A. Rakhmadi, A. Yudhana, and S. Sunardi, "Virtual reality and augmented reality in sign language recognition: A review of current approaches," *International Journal of Informatics and Computation*, vol. 6, no. 2, pp. 64–83, 2024. 7.2.1

[151] B. Parton, "Facilitating exposure to sign languages of the world: The case for mobile assisted language learning," 2014. 7.2.1

[152] M. E. Ahmed and S. Hasegawa, "Development of new distance learning platform to create and deliver learning content for deaf students," *Education Sciences*, vol. 12, no. 11, p. 826, 2022. 7.2.1

[153] T. Heift and M. Schulze, *Errors and intelligence in computer-assisted language learning: Parsers and pedagogues.* Routledge, 2007. 7.2.2

[154] K. R. Koedinger, E. Brunskill, R. S. Baker, E. A. McLaughlin, and J. Stamper, "New potentials for data-driven intelligent tutoring system development and optimization," *AI Magazine*, vol. 34, no. 3, pp. 27–41, 2013. 7.2.2, 8.4, 8.4, 8.4

[155] I. Adeshola and A. P. Adepoju, "The opportunities and challenges of chatgpt in education," *Interactive Learning Environments*, vol. 32, no. 10, pp. 6159–6172, 2024. 7.2.2

[156] J. R. Anderson, A. T. Corbett, K. R. Koedinger, and R. Pelletier, "Cognitive tutors: Lessons learned," *The journal of the learning sciences*, vol. 4, no. 2, pp. 167–207, 1995. 7.2.2

[157] A. C. Graesser, S. Lu, G. T. Jackson, H. H. Mitchell, M. Ventura, A. Olney, and M. M. Louwerse, "Autotutor: A tutor with dialogue in natural language," *Behavior Research Methods, Instruments, & Computers*, vol. 36, no. 2, pp. 180–192, 2004. 7.2.2, 8.4, 8.4

[158] A. Mitrovic and S. Ohlsson, "Evaluation of a constraint-based tutor for a database language," 1999. 7.2.2

[159] W. Huang, K. F. Hew, and L. K. Fryer, "Chatbots for language learning—are they really useful? a systematic review of chatbot-supported language learning," *Journal of computer assisted learning*, vol. 38, no. 1, pp. 237–257, 2022. 7.2.2

[160] E. Verhelst, R. Janssens, T. Demeester, and T. Belpaeme, "Adaptive second language tutoring using generative ai and a social robot," in *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 1080–1084, 2024. 7.2.2

[161] E. Kasneci, K. Seßler, S. Küchemann, M. Bannert, D. Dementieva, F. Fischer, U. Gasser, G. Groh, S. Günnemann, E. Hüllermeier, *et al.*, "Chatgpt for good? on opportunities and challenges of large language models for education," *Learning and individual differences*, vol. 103, p. 102274, 2023. 7.2.2

[162] D. Wood and S. H. Moss, "Evaluating the impact of students' generative ai use in educational contexts," *Journal of Research in Innovative Teaching & Learning*, vol. 17, no. 2, pp. 152–167, 2024. 7.2.2

[163] D. Thüs, S. Malone, and R. Brünken, "Exploring generative ai in higher education: a rag system to enhance student engagement with scientific literature," *Frontiers in Psychology*, vol. 15, p. 1474892, 2024. 7.2.2

[164] J. H. Woo and H. Choi, "Systematic review for ai-based language learning tools," *arXiv preprint arXiv:2111.04455*, 2021. 7.2.2

[165] A. Moinuddin, A. Goel, and Y. Sethi, "The role of augmented feedback on motor learning: a systematic review," *Cureus*, vol. 13, no. 11, 2021. 7.2.2

[166] S. P. Swinnen, "for motor skill learning: A review," *Advances in motor learning and control*, vol. 37, 1996. 7.2.2

[167] J. Hamari, D. J. Shernoff, E. Rowe, B. Coller, J. Asbell-Clarke, and T. Edwards, "Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning," *Computers in Human Behavior*, vol. 54, pp. 170–179, 2016. 7.2.3, 8.4, 8.4, 8.4, 10.1

[168] J. L. Plass, B. D. Homer, and C. K. Kinzer, "Foundations of game-based learning," *Educational Psychologist*, vol. 50, no. 4, pp. 258–283, 2015. 7.2.3, 8.4, 8.4, 8.4, 10.1

[169] A. Domínguez, J. Saenz-de Navarrete, L. De-Marcos, L. Fernández-Sanz, C. Pagés, and J.-J. Martínez-Herráiz, "Gamifying learning experiences: Practical implications and outcomes," *Computers & education*, vol. 63, pp. 380–392, 2013. 7.2.3

[170] J. Ranalli, "Learning english with the sims: exploiting authentic computer simulation games for l2 learning," *Computer Assisted Language Learning*, vol. 21, no. 5, pp. 441–455, 2008. 7.2.3

[171] H.-T. Hung, J. C. Yang, G.-J. Hwang, H.-C. Chu, and C.-C. Wang, "A scoping review of research on digital game-based language learning," *Computers & Education*, vol. 126, pp. 89–104, 2018. 7.2.3

[172] G. Lampropoulos and A. Sidiropoulos, "Impact of gamification on students' learning outcomes and academic performance: A longitudinal study comparing online, traditional, and gamified learning," *Education Sciences*, vol. 14, no. 4, p. 367, 2024. 7.2.3

[173] J. Looyestyn, J. Kernot, K. Boshoff, J. Ryan, S. Edney, and C. Maher, "Does gamification increase engagement with online programs? a systematic review," *PloS one*, vol. 12, no. 3, p. e0173403, 2017. 7.2.3

[174] F. F.-H. Nah, Q. Zeng, V. R. Telaprolu, A. P. Ayyappa, and B. Eschenbrenner, "Gamification of education: a review of literature," in *International conference on hci in business*, pp. 401–409, Springer, 2014. 7.2.3

[175] T. Kamnardsiri, L.-o. Hongsit, P. Khuwuthyakorn, and N. Wongta, "The effectiveness of the game-based learning system for the improvement of american sign language using kinect," *Electronic Journal of e-Learning*, vol. 15, no. 4, pp. pp283–296, 2017. 7.2.3

[176] Y.-H. Chang, J.-H. Hwang, R.-J. Fang, and Y.-T. Lu, "A kinect-and game-based interactive learning system," *Eurasia Journal of Mathematics, Science and Technology Education*, vol. 13, no. 8, pp. 4897–4914, 2017. 7.2.3

[177] Y. S. Chee, "Embodiment, embeddedness, and experience: Game-based learning and the construction of identity," *Research and Practice in Technology Enhanced Learning*, vol. 2, no. 01, pp. 3–30, 2007. 7.2.3

[178] B. Shneiderman, *Human-centered AI.* Oxford University Press, 2022. 7.2.4, 8.4, 8.4

[179] S. Amershi, D. S. Weld, M. Vorvoreanu, A. Fourney, B. Nushi, P. Collisson, J. Suh, S. T. Iqbal, P. N. Bennett, K. Inkpen, *et al.*, "Guidelines for human-ai interaction," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–13, ACM, 2019. 7.2.4, 8.4, 8.4, 10.1

[180] B. D. Mittelstadt, P. Allo, M. Taddeo, S. Wachter, and L. Floridi, "The ethics of algorithms: Mapping the debate," *Big Data & Society*, vol. 3, no. 2, p. 2053951716679679, 2016. 7.2.4

[181] P. Topali, A. Ortega-Arranz, M. J. Rodríguez-Triana, E. Er, M. Khalil, and G. Akçapınar, "Designing human-centered learning analytics and artificial intelligence in education solutions: a systematic literature review," *Behaviour & Information Technology*, vol. 44, no. 5, pp. 1071–1098, 2025. 7.2.4

[182] K. Wang, C.-S. Chai, J.-C. Liang, and G. Sang, "Exploring teachers' behavioural intentions to design artificial intelligence-assisted learning in chinese k–12 education," *Technology, Pedagogy and Education*, vol. 33, no. 5, pp. 629–645, 2024. 7.2.4

[183] B. du Boulay, "Artificial intelligence in education and ethics," in *Handbook of open, distance and digital education*, pp. 93–108, Springer, 2023. 7.2.4

[184] A. Darvishi, H. Khosravi, S. Sadiq, and D. Gašević, "Incorporating ai and learning analytics to build trustworthy peer assessment systems," *British Journal of Educational Technology*, vol. 53, no. 4, pp. 844–875, 2022. 7.2.4

[185] L. K. S. Tolentino, R. O. Serfa Juan, A. C. Thio-ac, M. A. B. Pamahoy, J. R. R. Forteza, and X. J. O. Garcia, "Static sign language recognition using deep learning," *International Journal of Machine Learning and Computing*, vol. 9, no. 6, pp. 821–827, 2019. 7.3

[186] A. Gelman and J. Hill, *Data analysis using regression and multilevel/hierarchical models.* Cambridge university press, 2007. 7.7.1

[187] J. C. Pinheiro and D. M. Bates, *Mixed-effects models in S and S-PLUS*. Springer, 2000. 7.7.1

[188] D. J. Barr, R. Levy, C. Scheepers, and H. J. Tily, "Random effects structure for confirmatory hypothesis testing: Keep it maximal," *Journal of memory and language*, vol. 68, no. 3, pp. 255–278, 2013. 7.7.1

[189] R. Gueorguieva and J. H. Krystal, "Move over anova: progress in analyzing repeated-measures data andits reflection in papers published in the archives of general psychiatry," *Archives of general psychiatry*, vol. 61, no. 3, pp. 310–317, 2004. 7.7.1

[190] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *Journal of statistical software*, vol. 67, pp. 1–48, 2015. 7.7.1

[191] N. Reimers and I. Gurevych, "Sentence-bert: Sentence embeddings using siamese bert-networks," *arXiv preprint arXiv:1908.10084*, 2019. 5

[192] A. W. Salmoni, R. A. Schmidt, and C. B. Walter, "Knowledge of results and motor learning: a review and critical reappraisal.," *Psychological bulletin*, vol. 95, no. 3, p. 355, 1984. 7.7.2

[193] R. A. Schmidt, T. D. Lee, C. Winstein, G. Wulf, and H. N. Zelaznik, *Motor control and learning: A behavioral emphasis*. Human kinetics, 2018. 7.7.2

[194] J. Hattie and H. Timperley, "The power of feedback," *Review of educational research*, vol. 77, no. 1, pp. 81–112, 2007. 7.7.2, 7.7.2

[195] V. J. Shute, "Focus on formative feedback," *Review of educational research*, vol. 78, no. 1, pp. 153–189, 2008. 7.7.2

[196] R. M. Ryan and E. L. Deci, "Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being," *American Psychologist*, vol. 55, no. 1, pp. 68–78, 2000. 7.7.2, 10.1

[197] B. J. Zimmerman, "Becoming a self-regulated learner: An overview," *Theory into practice*, vol. 41, no. 2, pp. 64–70, 2002. 7.7.2

[198] A. F. Hadwin, S. Järvelä, and M. Miller, "Self-regulated, co-regulated, and socially shared regulation of learning," *Handbook of self-regulation of learning and performance*, vol. 30, pp. 65–84, 2011. 7.7.2

[199] D. Carless and D. Boud, "The development of student feedback literacy: enabling uptake of feedback," *Assessment & Evaluation in Higher Education*, vol. 43, no. 8, pp. 1315–1325, 2018. 7.7.2

[200] R. A. Schmidt and T. D. Lee, *Motor Control and Learning: A Behavioral Emphasis*. Champaign, IL: Human Kinetics, 5 ed., 2011. 7.8, 7.8.1, 7.8.2

[201] K. VanLehn, “The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems,” *Educational Psychologist*, vol. 46, no. 4, pp. 197–221, 2011. 7.8, 7.8.1, 7.8.2, 7.8.4, 7.8.4, 7.8.6

[202] V. J. Shute, “Focus on formative feedback,” *Review of Educational Research*, vol. 78, no. 1, pp. 153–189, 2008. 7.8, 7.8.1, 7.8.2, 7.8.4, 7.8.5, 8.4, 8.4, 8.4, 8.4, 9.2

[203] R. A. Bjork, “Memory and metamemory considerations in the training of human beings,” in *Metacognition: Knowing about Knowing* (J. Metcalfe and A. P. Shimamura, eds.), pp. 185–205, Cambridge, MA: MIT Press, 1994. 7.8.1, 7.8.3, 7.8.4, 7.8.5, 9.2

[204] E. L. Bjork and R. A. Bjork, “Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning,” in *Psychology and the Real World: Essays Illustrating Fundamental Contributions to Society* (M. A. Gernsbacher, R. W. Pew, L. M. Hough, and J. R. Pomerantz, eds.), pp. 55–64, New York, NY: Worth Publishers, 2011. 7.8.1, 7.8.3, 9.2

[205] R. A. Magill and D. I. Anderson, *Motor Learning and Control: Concepts and Applications*. New York, NY: McGraw-Hill Education, 11 ed., 2017. 7.8.1

[206] K. A. Ericsson, R. T. Krampe, and C. Tesch-Römer, “The role of deliberate practice in the acquisition of expert performance,” *Psychological Review*, vol. 100, no. 3, pp. 363–406, 1993. 7.8.1

[207] J. Hattie, *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement*. London, UK: Routledge, 2009. 7.8.1, 7.8.4, 7.8.5, 8.4, 8.4, 8.4, 8.4, 8.4, 8.4, 9.2

[208] R. P. Heitz, “The speed–accuracy tradeoff: History, physiology, methodology, and behavior,” *Frontiers in Neuroscience*, vol. 8, p. 150, 2014. 7.8.2

[209] K. R. Koedinger, A. T. Corbett, and C. Perfetti, “The knowledge–learning–instruction framework: Bridging the science–practice chasm to enhance robust student learning,” *Cognitive Science*, vol. 36, no. 5, pp. 757–798, 2012. 7.8.4, 7.8.5

[210] W. Holmes, M. Bialik, and C. Fadel, *Artificial Intelligence in Education: Promise and Implications for Teaching and Learning*. Boston, MA: Center for Curriculum Redesign, 2019. 7.8.4, 7.8.5, 9.2

[211] R. A. Schmidt, “A schema theory of discrete motor skill learning,” *Psychological Review*, vol. 82, no. 4, pp. 225–260, 1975. 7.8.6

[212] R. Sigrist, G. Rauter, R. Riener, and P. Wolf, “Augmented visual, auditory, haptic, and multimodal feedback in motor learning: A review,” *Psychonomic Bulletin & Review*, vol. 20, no. 1, pp. 21–53, 2013. 9.2

[213] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan, “A survey on bias and fairness in machine learning,” *ACM Computing Surveys*, vol. 54, no. 6, pp. 115:1–115:35, 2021. 9.2

[214] D. Lakens, “Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and anovas,” *Frontiers in Psychology*, vol. 4, p. 863, 2013. 9.2

[215] B. A. Nosek, C. R. Ebersole, A. C. DeHaven, and D. T. Mellor, “The preregistration revolution,” *Proceedings of the National Academy of Sciences*, vol. 115, no. 11, pp. 2600–2606, 2018. 9.2

[216] R. L. Wasserstein, A. L. Schirm, and N. A. Lazar, “Moving to a world beyond “p ¡ 0.05”,” *The American Statistician*, vol. 73, no. sup1, pp. 1–19, 2019. 9.2

[217] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, and et al., “The FAIR guiding principles for scientific data management and stewardship,” *Scientific Data*, vol. 3, p. 160018, 2016. 9.2

[218] J. D. Kirkpatrick and W. K. Kirkpatrick, *Kirkpatrick's four levels of training evaluation*. Association for Talent Development, 2016. 10.1

[219] S. Messick, “Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning,” *American Psychologist*, vol. 50, no. 9, pp. 741–749, 1995. 10.1

[220] L. M. Earl, *Assessment as Learning: Using Classroom Assessment to Maximize Student Learning*. Thousand Oaks, CA: Corwin, 2013. 10.1

[221] A. Pollitt, “Comparative judgement for assessment,” *International Journal of Technology and Design Education*, vol. 22, no. 2, pp. 157–170, 2012. 10.1

[222] D. A. Dillman, J. D. Smyth, and L. M. Christian, *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. Hoboken, NJ: John Wiley & Sons, 2014. 10.1

CHAPTER 10

Appendix

10.1 Raw Data

All raw data, experiment code, and supplementary materials supporting this thesis are openly available in a dedicated repository:

`https://github.com/WJD95/ASL-learning`

The repository contains the datasets, software implementations, and generated outputs from the studies reported in this thesis. A detailed description of the repository contents and instructions for reproduction are provided in the accompanying `README` file.

10.2 Gesture Recognition Models

Across both tasks, the models achieved over 95% test accuracy. Figure 10.1 shows the confusion matrices. The results demonstrate strong diagonal dominance, with only minor misclassifications between visually similar signs (e.g., *ok* vs. *sorry*, or digit *6* vs. *9*). This level of accuracy was sufficient to support the learning experiment, ensuring that system feedback was generally dependable and not confounded by large-scale recognition errors.

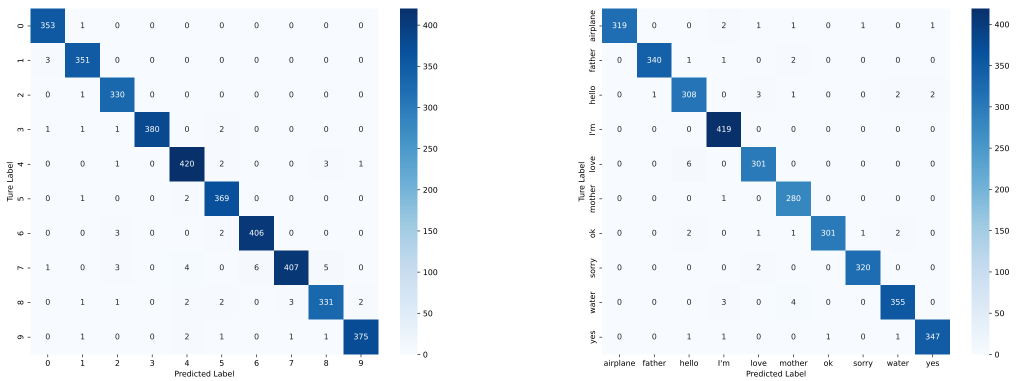

Figure 10.1: (left) Confusion matrix for digit classification model (10 classes). (right) Confusion matrix for word classification model (10 classes).

10.3 LLM feedback generation

Inputs and scope The system relied on a structured dictionary of canonical finger states. Two dictionaries were defined: one covering digits 0–9 and another covering ten ASL vocabulary words. Each entry specified the expected position of the thumb, index, middle, ring, and pinky fingers (e.g., "index: extended upward", "middle: curled inward"). At runtime, the hand-tracking recogniser produced a symbolic description of the learner's current gesture in the same schema. Both the recognised (current) gesture and the target gesture were then passed to the LLM. Importantly, no images or videos were provided to the model; the prompts contained only structured text.

Prompt structure The LLM call used a two-part prompt. The fixed *system prompt* instructed the model to respond in strict JSON format:

> You are a helpful assistant that returns responses in JSON format.

The *user prompt* contained three sections: (i) the current and target gestures, (ii) the full canonical dictionary (digits or words), and (iii) explicit output requirements, including a JSON schema template. At runtime, placeholders {current} and {target} were replaced with the actual gestures, and {data} was filled with the relevant dictionary. The embedded template was as follows:

1. **Input Data**: current: {current}, target: {target}, Use this gesture dictionary for reference: {data}
2. **Pair Generation Rules**: – Treat each pair as bidirectional (current vs target)
3. **Output Requirements**: For each pair produce JSON data: {

```
[”X, Y”]: {
  ”comparison”: {
    ”current_gesture”: {”name”: ”X”, ”finger_positions”: {...}},
    ”target_gesture”: {”name”: ”Y”, ”finger_positions”: {...}}
  },
  ”differences”: {
    ”thumb”: ”Both curled (correct)”,
    ”index”: ”Both extended (correct)”,
    ”middle”: ”Current: curled — Target: extended”,
    ”ring”: ”Both curled (correct)”,
    ”pinky”: ”Both curled (correct)”
  },
  ”correction_tips”: [
    ”Actionable step 1”,
    ”Actionable step 2”
  ]
}
}
```

Outputs For every (*current*, *target*) pair, the LLM returned a structured JSON object that contained: (a) a *comparison* section with both gestures and their finger states, (b) a per-finger *difference report*, and (c) two short, learner-facing *correction tips*. These tips were limited to 12 words each and framed as concrete motor actions (e.g., “curl middle inward behind index” or “pinch thumb to index to form an ‘O’ shape").

The same prompt template was applied to both the digit set (0–9) and the vocabulary word set. For reproducibility, the results were stored separately in two JSON files: one containing all digit pairwise comparisons and tips, and another containing the equivalent results for the ten vocabulary words. This separation ensured that the AI-generated feedback was consistently structured, while also enabling a direct and controlled comparison with the non-AI baseline condition in the experimental design.

The symbolic dictionary abstracts away wrist orientation, movement, and timing, so the generated feedback was limited to static finger configurations. This design choice ensured clarity and reproducibility, while leaving more complex phonological features (e.g., motion or location) for future work.

10.4 Interview Process

This study employed direct note-taking and typed responses. This approach was chosen for two reasons: (1) **comfort and privacy of the participants**, as several participants expressed reluctance to being recorded; and (2) **efficiency**, as the researcher captured key responses in real time, eliminating the need for subsequent transcription.

The researcher conducted the interview and documented the responses verbatim. Immediately after each session, the notes were cross-verified by both the researcher and the participant to ensure precision.

Table 10.1: Post-Interview Questionnaire for AI-Assisted vs. Non-AI Groups

Section	Question (AI-specific wording in *italics*, Non-AI wording in bold)	Rationale / Theoretical Basis
1. Learning Experience		
Q1	How clear and useful were the *AI-generated feedback instructions* / **predefined learning materials** during your training?	Instructional clarity and usability, aligning with Kirkpatrick's Four-Level Model [218] (Levels 1 and 2).
Q2	Did the *AI feedback* / **materials** help you identify and correct mistakes on your own? Why or why not?	Measures learner autonomy and error correction, reflecting guidelines for effective AI feedback [179].
Q3	Were there any challenges in following the *AI's guidance* / **predefined instructions**?	Identifies usability barriers, related to construct validity in evaluation [219].
Q4	Which signs did you find most difficult, and how did you work on improving them?	Captures difficulty perception and self-regulated learning strategies.
Q5	If you had to describe the *AI-assisted learning experience* / **predefined learning path** in one sentence, what would it be?	Summative perception, supporting evaluation of learner satisfaction (Kirkpatrick Level 1).
2. General Engagement		
Q6	How motivated did you feel during the learning sessions?	Motivation assessed through Self-Determination Theory [196].
Q7	Did you find this learning method engaging compared to other learning styles you've tried?	Engagement in relation to gamification/adaptive learning benefits [167].
3. Assessment & Game Reflection		
Q8	How did you feel about your final assessment compared to your initial assessment?	Based on Assessment as Learning framework [220], measuring perceived progress.
Q9	How did the whack-a-mole game influence your recall and speed?	Draws from game-based learning research on skill reinforcement [168].
4. Hypothetical Comparison & Open Feedback		
Q10	How do you think you would have performed if you had **not used AI assistance** / *used AI assistance instead*?	Comparative judgment theory [221], assessing counterfactual perceptions.
Q11	Any suggestions for improving the *AI-assisted learning experience* / **predefined learning experience**?	User-centered design principles, eliciting actionable feedback [222].