Bridging the Gap: OCR and LLM-Assisted Tools for Integrating Commercial Game-Based Language Learning into Formal Instruction

University of Turku
Department of Computing
Master of Science Thesis
June 2026
Jesse Tornberg

TURUN YLIOPISTO
Tietotekniikan laitos

Jesse Tornberg: Bridging the Gap: OCR and LLM-Assisted Tools for Integrating Commercial Game-Based Language Learning into Formal Instruction

Pro gradu -tutkielma, 54 s., 3 liites.
Kesäkuu 2026

Kaupalliset pelit ovat herättäneet kasvavaa kiinnostusta kielten oppimisympäristöinä niiden tarjoamien motivaatiotekijöiden, autenttisen kielenkäytön, ja vuorovaikutuksen mahdollistamisen vuoksi.
Itsenäisen pelaamisen ja kielenopetuksen välillä on kuitenkin puuttuva osa: olemassa olevat avustavat työkalut palvelevat itsenäisiä oppijoita, mutta niistä puuttuu pedagoginen seuranta, opettajan näkyvyys, ja opetussuunnitelman seuranta.
Tämä tutkielma käsittelee tätä puuttuvaa osaa kahden toisiaan täydentävän tutkimuslinjan avulla. Systemaattinen kirjallisuuskatsaus 24 empiirisestä tutkimuksesta tarkastelee, mitä pelejä on käytetty digitaalisen pelipohjaisen kielenoppimisen (DGBLL) tutkimuksessa, millä perusteilla ne valittiin ja mitä rajoituksia niissä on havaittu.
Tutkielman empiirinen osa, joka koostuu puolistrukturoiduista haastatteluista neljän yliopistotason japanin kielen opettajan kanssa. Haastattelut vastaavat siihen, että olisiko tarkoitukseen rakennettu työkalu - joka yhdistää optisen tekstintunnistuksen (OCR) ja suuren kielimallin (LLM) teknologian –pedagogisesti toteuttamiskelpoinen ratkaisu.
Kirjallisuuskatsauksessa todetaan, että pelien valintaa ohjaavat ensisijaisesti motivoivat ja sosiaaliset tekijät pikemminkin kuin pedagoginen suunnittelu, ja että siitä johtuva tukirakenteiden puute ja opetuksen läpinäkymättömyys ovat toistuvia esteitä eri tutkimuksissa.
Haastattelutulokset vahvistavat, että kouluttajat ovat avoimia pelipohjaiselle oppimiselle, mutta tarvitsevat morfologista tukea kokonaisten lauseiden kääntämisen sijaan. Opetussuunnitelman mukaan suodatettu istuntoraportointi ja opettajan ohjaamien kieliopin seurantalistojen luominen ovat myös osa ratkaisua, ennen kuin pelaamista voidaan pitää opiskeluna. Yhdessä nämä havainnot viittaavat siihen, että este kaupallisten pelien integroinnille opetukseen ei ole ensisijaisesti teknisellä, vaan organisoinnin tasolla: pelaamisen läpinäkyvyyden ja opetussuunnitelmaan mukauttamisen mahdollistava infrastruktuuri, tässä tapauksessa erillisenä työkaluna on puuttuva pala itsenäisen pelaamisen ja institutionaalisen tunnustuksen välillä.

Asiasanat: videopelit, pelipohjainen oppiminen, kieltenopiskelu, optinen tekstintunnistus (OCR), suuret kielimallit (LLM), kielenoppimistyökalut

UNIVERSITY OF TURKU
Department of Computing

Jesse Tornberg: Bridging the Gap: OCR and LLM-Assisted Tools for Integrating Commercial Game-Based Language Learning into Formal Instruction

Master of Science Thesis, 54 p., 3 app. p.
June 2026

Commercial off-the-shelf (COTS) games have attracted growing interest as language learning environments due to their motivational affordances, authentic linguistic input, and capacity for meaningful communicative interaction.
However, a structural gap persists between autonomous gameplay and formal language instruction: existing assistive tools serve independent learners but lack any mechanism for instructor oversight or curricular alignment.
This thesis addresses that gap through two complementary strands of inquiry. A systematic literature review of 24 empirical studies examines which games have been used in Digital Game-Based Language Learning (DGBLL) research, on what grounds they were selected, and what limitations they consistently present.
A qualitative empirical component consisting of semi-structured interviews with four university-level Japanese language educators then investigates whether a purpose-built bridge tool — integrating Optical Character Recognition (OCR) and Large Language Model (LLM) technology — would be considered a pedagogically viable solution.
The literature review finds that game selection is driven primarily by motivational and social factors rather than pedagogical design, and that the resulting scaffolding gap and instructional opacity are consistent barriers across studies.
Interview findings confirm that educators are open to game-based learning but require morphological scaffolding over full-sentence translation, curriculum-filtered session reporting, and teacher-directed grammar watchlists before autonomous gameplay can be considered a legitimate academic activity. Together, these findings suggest that the barrier to integrating COTS games into formal instruction is not primarily technical but organizational: the infrastructure for transparency and curricular alignment is the missing link between autonomous play and institutional recognition.

Keywords: digital game-based language learning, DGBLL, commercial off-the-shelf games, COTS, language acquisition, OCR, large language models, language learning tools

Contents

1 Introduction **1**

1.1 Disclosure on the use of Artificial Intelligence 4

1.2 Thesis Structure 4

2 Background **6**

2.1 Digital Games and Gamification 7

2.2 Language Learning 8

2.3 Digital Game-Based Language Learning 10

2.4 Third-Party Tools: OCR and LLM Integration 12

2.5 Literature Review 14

3 COTS Games in DGBLL **20**

3.1 Second Life 22

3.2 Ragnarok Online 25

3.3 World of Warcraft 26

3.4 Simulation, RPG, and Action Titles 28

3.4.1 Life-Simulation and Narrative-Driven Games 29

3.4.2 RPGs, Visual Novels, and Puzzles 30

3.4.3 Action, Music, and Social Interaction 32

3.5 Recurring Limitations 34

4 Experimental Setup **35**

4.1 Semi-Structured Interview 35

4.2 Tool Prototype 36
4.2.1 Text Capture and Processing 37
4.2.2 Morphological Scaffolding and Reader View 37
4.2.3 Flashcard Integration and Data Logging 38
4.3 Interview Process and Data Analysis 39
4.3.1 General Perceptions of DGBLL 40
4.3.2 Visual Probing and Screenshot Analysis 40
4.3.3 Mockup Evaluation and Technical Requirements 41

5 Results and Analysis **42**
5.1 Teacher Attitudes and Interview Results 42
5.1.1 Scaffolding and Automated Support 43
5.1.2 Information Requirements and Data Filtering 43
5.1.3 Course Alignment and Teacher-Directed Watchlist 44
5.1.4 Obscurities in Reporting and Assessment 44
5.2 Game Selection 45
5.3 Vocabulary Acquisition and Outcomes 46
5.4 Barriers to Implementation 47
5.5 Answering the Research Questions 49
5.6 Limitations 50

6 Conclusion **52**
6.1 Contribution 53
6.2 Further Research 54

References **55**

Appendices

A Interview Guide **A-1**

1 Introduction

Learning a new language is a long, demanding process that requires consistent effort, patience and motivation. Unlike subjects or topics that can be learned and nearly mastered through short-term memorization and rote learning, language acquisition unfolds gradually through continuous exposure, practice, refinement and immersion. It engages multiple skills: listening, speaking, reading and writing, as well as the emotional resilience to persist through slow progress and frequent errors. For many learners, this journey is marked by cycles of enthusiasm and discouragement, making motivation one of the most decisive factors for success. As psychologists and linguists have long emphasized, motivation sustains learners through inevitable challenges of acquiring a second language and beyond[1].

In formal classroom environments, motivation is often supported by external structures such as teachers, deadlines and peer interaction. However, with the growing accessibility of mobile applications and online resources, many individuals can now pursue languages through self-directed learning, studying independently and outside of institutional settings. Self-directed learning offers flexibility and autonomy but also introduces motivational challenges. Without external reinforcements or feedback, learners must rely heavily on internal drives and self-regulation to stay engaged over long periods. Studies have shown that maintaining motivation in self-study contexts is often more difficult than in formal education, as learners face distractions, declining novelty and uncertainty about their progress[2]. As learners

take on greater responsibility for their own success, they must also develop skills to plan, monitor and evaluate their learning effectively.

To tackle these challenges, researchers and developers have turned towards gamification, the integration of game elements, such as points, badges, levels and rewards, into traditionally non-game environments[3]. The rationale behind gamification is the same psychological mechanism that makes games engaging: clear goals, feedback, challenges and a sense of progression can also increase learning motivation and engagement. In language learning applications like *Duolingo*, *Memrise*, and *Babbel*, gamified systems encourage users to practice regularly through daily streaks, achievement badges and competitive leaderboards. These designs create small, attainable goals that can foster persistence and enjoyment, potentially counteracting the monotony and self-doubt often found in self-study.

Research from the past decade has offered a more comprehensive understanding of how gamification functions as a motivational design strategy. Hamari, Koivisto, and Sarsa (2014) describe gamification as the process of enhancing services with motivational affordances to create gameful experiences that encourage desired behaviors[4]. This definition emphasizes that gamification is not about turning an activity into a fully-fledged game. Instead, it involves using selected elements, mainly feedback loops and progress indicators, to evoke the engagement typically experienced while playing games. When applied to learning, these elements aim to transform repetitive tasks into experiences that feel more purposeful.

However, while gamification has proven effective at increasing short-term motivation and user retention, its impact on deep pedagogical outcomes and long-term linguistic proficiency remains questionable. Because gamified applications often rely on extrinsic rewards rather than immersive content, there has been a growing interest in Digital Game-Based Language Learning (DGBLL). This approach moves beyond simple game elements by using commercial off-the-shelf (COTS) games as primary

learning environments. Unlike gamified apps, commercial games offer complex narratives and authentic linguistic contexts that provide a more robust foundation for language acquisition.

Despite the motivational promise of COTS games, a structural gap remains between autonomous gameplay and formal language instruction. Existing third-party assistive tools, such as Game2Text, YomiNinja, and Migaku, serve independent learners but lack any mechanism for instructor oversight or curricular alignment. This thesis addresses that gap by examining which games have been used in DGBLL research and why, what limitations they present, and whether language educators would consider a purpose-built bridge tool a viable solution. The proposed tool integrates Optical Character Recognition (OCR) and Large Language Model (LLM) technology to extract in-game language, provide morphological scaffolding, and generate session reports for instructor review.

The study is guided by three research questions:

1. RQ1: Which commercial games have been featured in DGBLL research, what were the specific pedagogical objectives, and how was the selection of these titles justified?

2. RQ2: What limitations do researchers identify regarding the use of commercial games, particularly concerning content difficulty, the lack of instructional feedback, and learner support?

3. RQ3: What are language instructors' attitudes toward using a digital bridge tool to facilitate learning during gameplay, and what specific information would they require from a student's individual gaming session to consider it a pedagogically valid learning activity?

1.1 Disclosure on the use of Artificial Intelligence

Artificial intelligence tools were used during the writing of this thesis. Google Gemini was used in summarizing and identifying initial reference candidates.
Claude Opus and Google Gemini were used to assist in drafting and stylizing portions of the thesis text.
Google AI Studio was used to generate the tool prototype.
The author takes full responsibility for all content, arguments, and conclusions presented in this thesis.

1.2 Thesis Structure

This thesis is structured to bridge the gap between theoretical language acquisition and the practical application of computing tools in gaming environments. The document is organized as follows:

Chapter 2 builds the theoretical foundation in three steps. It first defines digital games and gamification to establish the motivational logic that makes game-based environments engaging. It then examines second language acquisition theory, establishing the criteria against which game-based learning is later evaluated. Finally, it surveys the current landscape of Digital Game-Based Language Learning (DGBLL) and existing third-party assistive technologies.

Chapter 3 introduces the literature review component, focusing on Commercial Off-The-Shelf (COTS) games used in language learning contexts. It categorizes existing research by game genre—ranging from narrative-driven simulations to massively multiplayer online games—to identify the specific linguistic objectives targeted by researchers.

Chapter 4 presents the methodology of the empirical component of the study. It details the semi-structured interview design, the recruitment of language educators

as participants, and the tool prototype used during the data collection process.

Chapter 5 synthesizes the findings from both the literature review and the primary data collection. It opens with the interview results, detailing the practical requirements educators identified for a bridge tool to be viable in formal instruction (RQ3). These findings then establish the lens through which the literature is examined, addressing game selection patterns (RQ1), vocabulary acquisition outcomes, and the structural barriers that limit the effectiveness of COTS games in educational settings (RQ2). The chapter concludes by directly answering all three research questions and acknowledging the study's limitations.

Chapter 6 draws the thesis to a close by summarizing the overall argument. It outlines the study's theoretical and practical contributions and identifies directions for future research.

2 Background

This chapter establishes the background for the study by synthesizing research across four interconnected domains. It proceeds in stages, each building on the last to bridge the gap between autonomous play and formal instruction. The first section examines digital games and gamification as motivational design systems, establishing why game-based environments are capable of sustaining learner engagement through dynamics and mechanics that traditional instruction often lacks. The second situates language learning within a framework of strategies and immersion, identifying the pedagogical criteria—such as interaction and feedback—against which any learning environment must be evaluated. The third brings these bodies of literature together by surveying the field of Digital Game-Based Language Learning (DGBLL), tracing how researchers have attempted to harness the motivational affordances of commercial games for second language acquisition.

Moving toward the practical implementation of these theories, the fourth section examines the current landscape of third-party assistive technologies, specifically how Optical Character Recognition (OCR) and Large Language Models (LLMs) are used by independent learners to overcome the technical and linguistic barriers of complex games. This background concludes by outlining the systematic methodology used for the subsequent literature review. Together, these sections provide the framework necessary to evaluate the potential of a dedicated "bridge tool" for formal instruction.

2.1 Digital Games and Gamification

Video games are commonly understood as interactive digital systems in which players engage with rule-based environments to achieve goals through meaningful action. Juul (2011) defines games as formal systems governed by rules where player effort influences variable outcomes that are assigned different values, with video games distinguished by their computer-mediated execution and audiovisual representation[5]. Salen and Zimmerman (2003) emphasize that games enable "meaningful play," in which player actions produce immediate and interpretable feedback within the system[6]. This distinguishes them from passive media; as Aarseth (1997) notes, games are "ergodic texts" requiring non-trivial effort from the player to progress and generate meaning[7]. Consequently, language often serves as the essential interface for this interaction, as players must comprehend instructions, narratives, and system feedback to succeed.

While video games represent a complete ecosystem of rules and play, the concept of gamification has developed over the past decade as a design strategy to isolate and apply these mechanics to non-game contexts. It is commonly defined as the use of game design elements—such as points, badges, leaderboards, and progress indicators—to enhance user experience and motivation in otherwise routine activities. Rather than creating a full ludic experience, gamification seeks to replicate the sense of progression and satisfaction found in games to influence user behavior.

The effectiveness of this strategy stems from its ability to satisfy fundamental psychological needs. Hamari and Koivisto (2019) conceptualize gamification as "motivational information systems," arguing that its value lies in fulfilling the user's need for autonomy, competence, and relatedness[8]. Similarly, van Roy and Zaman (2017) stress that successful gamification must support these needs through meaningful goals and feedback rather than imposing external control through rigid rewards[9]. This is a crucial distinction, as "pointsification"—the shallow use of game

elements—may produce temporary excitement but often fails to foster long-term engagement[10].

Empirical research has identified clear layers to effective gamified design: dynamics, mechanics, and components[11]. In educational settings, elements such as progress indicators and badges have proven effective for increasing perceived autonomy[12]. In the domain of language learning, platforms like Duolingo demonstrate how these elements can lead to higher enjoyment and consistent vocabulary practice[13]. These motivational benefits are well-supported across various systematic reviews, particularly regarding mobile-assisted learning[14].

However, the transition from engagement to durable learning outcomes remains complex. While experimental evidence shows that gamified platforms can significantly improve academic performance, they can also introduce negative emotional consequences, such as increased anxiety or a decreased preference for teamwork[15]. Furthermore, many studies in the DGBLL field remain short-term and design-focused, lacking the longitudinal data necessary to verify sustained improvement in language proficiency[13], [14]. Ultimately, gamification emerges as a powerful motivational framework, but its long-term educational value depends on a coherent alignment between pedagogical grounding and adaptive design elements that sustain intrinsic motivation.

2.2 Language Learning

Language learning is a complex process that develops steadily through exposure, interaction and practice. It relies on four connected skills: listening, speaking, reading and writing. Each skill supports the others to build communicative competence; effective learning requires not just building vocabulary and grammar but also using language appropriately in different situations and contexts. Research on successful learners shows they approach learning with a plan, setting goals, tracking progress

and adjusting methods based on feedback[16]. Oxford (1990) identified six main categories of strategies: memory, cognitive, compensation, metacognitive, affective and social, arguing that skilled learners combine these strategies flexibly to make learning more independent and effective[17].

Lee (2010) expanded on this by showing that learners' choices of strategies are affected by motivation, beliefs and cultural background. Learners who use a wide range of strategies, including teamwork and emotional engagement, tend to reach higher proficiency[18]. White (2008) highlighted that real-life tasks that blend understanding, production and feedback result in more lasting learning. Feedback helps learners notice gaps between what they intend to do and what they actually do, turning mistakes into chances to learn and increasing awareness of their progress[19].

While motivation helps maintain effort, feedback and immersion provide the practical foundation that turns practice into communicative competence. Exposure to language in meaningful situations aids learners in internalizing patterns and gaining confidence, making this effect hard to achieve in isolated and purely theoretical exercises. Self-directed learners often miss immediate reinforcement and social interaction, which can slow down progress and lower satisfaction[2]. Therefore, structured feedback systems and interactive settings are crucial for keeping learners engaged and supporting long-term growth.

Overall, research shows language learning is an active process driven by feedback that thrives on immersion and purposeful communication. Listening, reading, speaking and writing improve best when learners receive meaningful responses and can use language in context. These aspects help explain why newer digital methods, especially gamification and game-based environments, aim to replicate the motivational and feedback-rich qualities of natural language use.

2.3 Digital Game-Based Language Learning

Digital Game-Based Language Learning (DGBLL) has emerged as a research area that explores how commercial and educational games can support language learning by combining interaction, feedback and meaningful use of language. Early theoretical examinations highlight why games align well with language acquisition, while more recent scoping reviews map the empirical evidence on their effectiveness.

Ang and Zaphiris (2008) argue that computer games naturally support core principles of language learning because they combine imitation, feedback, exploration and social interaction—processes central to major learning theories such as behaviorism and constructivism [20]. They outline two main pathways through which games can facilitate language learning. The first involves player-game interaction, where commercial or purpose-built games present language as part of gameplay through instructions, narratives or linguistic input required to progress. Examples such as *Kana no Senshi*, *PARLING* and the Tactical Language Training System show how immediate feedback, multimodal input and immersive scenarios can support character recognition, vocabulary development and oral communication skills.

The second path involves player-player interaction, particularly within massively multiplayer online games (MMOGs). Ang and Zaphiris highlight that players naturally use language to coordinate tasks, negotiate meaning, role-play, joke and build social relationships—all activities that closely mirror authentic communicative practice [20]. They conclude that games offer motivational and contextual advantages unavailable in most traditional computer-assisted language learning environments, though existing systems remain limited either by narrow pedagogical goals or the pre-authored nature of their content.

Chik (2014) examined how learners actually use commercial off-the-shelf (COTS) games in informal self-directed environments. Chik's study on university students shows that most DGBLL takes place outside formal classrooms, where players choose

games for entertainment but still encounter meaningful linguistic input [21]. Learners frequently select English versions of games, encounter vocabulary through menus, dialogues and quests, and use gameplay as an opportunity to practice comprehension. A further insight in Chik's work is that language learning through games can be socially supported: learners routinely turn to communities to clarify vocabulary and solve comprehension problems, with these communities often acting as informal teachers by recommending suitable games or providing glossaries for game-specific terminology. At the same time, Chik notes that COTS games present clear limitations for language learning, including inconsistent vocabulary exposure, non-pedagogical difficulty progression, and a lack of built-in scaffolding [21].

Kronenberg (2012) addresses the challenge of selecting and integrating COTS games into formal L2 learning environments by proposing a comprehensive framework of nine evaluation criteria [22]. The framework emphasizes that successful implementation requires games to provide intrinsic motivation and flow experiences, clearly defined and spaced goals, accessible game mechanics that prioritize linguistic comprehension over motor skills, and authentic content that supports meaningful language use. Kronenberg argues that COTS games offer distinct advantages over purpose-built educational games, including higher production quality, genuine entertainment value and authentic cultural content that native speakers engage with in everyday contexts. However, the framework also acknowledges that learning occurs primarily around the game rather than within it, requiring careful pedagogical scaffolding through pre-teaching activities, collaborative play, content creation tasks and structured reflection. Kronenberg analyzed three example games—*Buzz* (a customizable quiz game), *Heavy Rain* (an interactive narrative) and *SingStar* (a karaoke game)—to demonstrate how different genres can support various learning objectives when appropriately integrated. The framework highlights that while COTS games can provide immersive, multimodal and highly motivating environments, their edu-

cational value depends on thoughtful curricular alignment, instructor support and activities that extend gameplay into meaningful language production and analysis [22].

A broader picture of the field is offered by Xu et al. (2020), who conducted a wide-ranging review of digital game-based technology in English language learning. Their analysis shows increased research interest across the 2000s and 2010s, with studies consistently reporting motivational benefits and strong learner engagement across vocabulary, grammar and speaking activities [23]. Educational games, mobile games and commercial games were all found to support language learning, though the strongest effects appeared in vocabulary acquisition tasks. However, the review notes several limitations: many studies use small samples, short interventions, or lack rigorous experimental design. As a result, long-term language development outcomes remain underdocumented. Xu et al. emphasize the need for better methodological consistency, clearer theoretical grounding and more research on authentic language use within game environments.

2.4 Third-Party Tools: OCR and LLM Integration

While the majority of existing DGBLL research evaluates games as self-contained units, contemporary language learners increasingly rely on toolkits that function externally to the game software. Currently available tools such as *Migaku*, *YomiNinja* and *Game2Text* serve as evidence of demand for assistive technologies among independent learners who want to utilize games for language study. These tools are designed to overcome technical barriers inherent to commercial games, such as non-selectable text and rapid dialogue, by utilizing Optical Character Recognition (OCR) and Large Language Models (LLMs) to extract, parse and translate the language present on-screen. For instance, real-time OCR allows learners to capture on-screen text and transform it into interactive subtitles linked to digital dictionar-

ies, while transparent overlays allow for immediate dictionary retrieval and "sentence mining" without requiring the learner to interrupt their gameplay.

However, despite the evidence of learner initiative and the widespread use of these tools within independent learning communities, a significant gap exists in how these technologies could be applied within formal education with meaningful teacher input. To date, these tools have not been used in formal research studies, which typically rely on manual modding or pre-selected game content rather than external assistive software. Because these tools were designed primarily for autonomous and informal use, they lack a formal pedagogical framework and suffer from several critical limitations. Most notably, they function as an isolated experience for the student, lacking any mechanism to aggregate or report linguistic data to an instructor. Furthermore, these tools suffer from instructional blindness, as they cannot distinguish between general game text and specific curriculum-aligned vocabulary and grammar targets an instructor might wish to prioritize.

By integrating LLMs, these tools can offer real-time grammatical breakdowns and explanations for game-specific jargon, effectively providing some of the pedagogical scaffolding that Miller and Hegelheimer (2006) identified as a critical missing component in COTS games[24]. Yet, without a proper bridge to the instructor, a gaming session remains opaque and cannot be easily validated in an educational setting. There remains a vital need to evaluate whether formalizing these third-party tools—through features such as teacher-directed grammar watchlists and language exposure reports—is feasible, helpful and acceptable from the perspective of an educator. This thesis therefore examines whether such a tool could transform an independent, isolated activity into a structured academic task.

2.5 Literature Review

Given the multi-disciplinary nature of Digital Game-Based Language Learning (DGBLL), which intersects computer science, pedagogy, and linguistics, it was necessary to adopt a structured methodology for identifying and analyzing the relevant literature. This thesis employs a systematic selection process to identify empirical studies that offer specific insights into the use of Commercial Off-The-Shelf (COTS) games.

The identification of the studies discussed in the following chapters was grounded in two primary strategies. First, a targeted search was conducted across academic databases using a refined boolean search string designed to capture the intersection of gaming technologies and Second Language Acquisition (SLA). The search string utilized a variety of technical descriptors including "MMORPG," "COTS," "MUD," and "Virtual Worlds" to ensure comprehensive coverage across different game architectures.

Second, this research utilized a snowballing identification method derived from two foundational systematic and scoping reviews: Li (2019) [25] and Xu et al. (2020) [23]. These sources provided a validated baseline of high-impact studies, which were then subjected to a rigorous three-stage pruning process:

1. *Title Screening* for relevance to commercial gaming.
2. *Abstract Review* for focus on language outcomes.
3. *Full-text Analysis* to ensure the availability of data regarding game-specific affordances and technical/pedagogical limitations.

By following this pipeline, the thesis ensures that the games analyzed—ranging from the social complexities of *World of Warcraft* to the narrative-driven mechanics of *The Walking Dead*—represent a cross-section of the most significant empirical evidence available in the field today. This systematic approach allows for a more

reliable answer to the research questions regarding the selection of game titles, the identification of practical limitations, and the specific information instructors require to bridge the gap between autonomous play and formal assessment.

The integration of COTS games within DGBLL reveals a systematic tension between game architecture and pedagogical utility. Research consistently prioritizes MMORPGs like *World of Warcraft* and *Guild Wars 2* because their social mechanics necessitate spontaneous, goal-oriented interaction, which aligns with interaction-based theories of second language acquisition. These multiplayer environments facilitate a "psychologically safe" space where avatar anonymity lowers the communicative threshold, allowing learners to progress from simplified language to nuanced pragmatic markers through collaborative necessity. In contrast, single-player narrative or simulation games such as *The Sims* or *The Walking Dead* are selected for their authentic linguistic input and high-fidelity mapping of domestic vocabulary, though they offer fewer opportunities for the real-time negotiation of meaning.

A recurring distinction in the field is the functional difference between mechanics-driven and narrative-driven titles. Mechanics-oriented games support frequent exposure to formulaic language and vocabulary acquisition through repetitive actions, whereas narrative-heavy games provide rich input but fewer measurable learning moments for experimental assessment. Despite high motivational aspects, a primary barrier remains in the scaffolding gap, as COTS titles lack the built-in sequencing and explicit feedback inherent in dedicated educational software. This often leads to a cognitive overload paradox where the high interactivity and motor-skill demands of the game—such as Quick Time Events or complex interface navigation—compete for the mental resources required for language processing, potentially hindering recall. Consequently, the educational value of COTS games is frequently framed as supplementary, requiring external instructional guidance and bridge materials to translate incidental encounters into sustained linguistic development.

The studies covered in the literature review are summarized in the table below and will each be covered in more detail in the following chapter.

Table 1: Table of Reviewed Studies (n=24)

Author& Year	**Game**	**Mode**	**Setting**	**Participants**	**Results**
Miller & Hegelheimer (2006)	The Sims	SP	In-Class	18 EFL	Gains only when supplemental materials were mandatory.
Ranalli (2008)	The Sims	SP	In-Class	9 EFL	Significant improvements in vocabulary with support materials.
Rankin et al. (2009)	EverQuest II	MP	In/Out-Class	18 EFL, 8 Native	Social interactions developed communicative strategies and vocabulary.
deHaan et al. (2010)	PaRappa 2	SP	In-Class	80 EFL	Watchers recalled more vocabulary; interactivity hindered players.
Peterson (2010)	Second Life	MP	In-Class	7 EFL	High learner-centered interaction; managed social/technical hurdles.
Reinders& Wattana (2011)	Ragnarok Online	MP	In-Class	16 EFL	Text-based chat produced more words and longer sentences than voice.

Continued on next page

Note. SP = Single-player; MP = Multiplayer

Table 1 – continued from previous page

Author& Year	Game	Mode	Setting	Participants& Language	Results
Wehner et al. (2011)	Second Life	MP	In-Class	41 Spanish	Users reported lower anxiety and higher motivation than traditional class.
Liang (2012)	Second Life	MP	In-Class	11 EFL	Motivation driven by collaborative storytelling and "language play."
Liou (2012)	Second Life	MP	In-Class	25 EFL	Significantly reduced anxiety; provided real-life context for practice.
Peterson (2012)	Second Life	MP	In-Class	8 EFL	Facilitated collaborative interaction in a low-stress virtual setting.
Rama et al. (2012)	World of Warcraft	MP	Out-of-Class	2 Spanish	Gaming expertise compensated for lower linguistic proficiency.
Zheng et al. (2012)	World of Warcraft	MP	Out-of-Class	4 EFL	Learning occurred via multimodal coordination (avatar, text, voice).
Vandercruysse et al. (2013)	Divine Divinity	SP	In-Class	83 EFL	Competition increased perceived competence but not direct gains.

Continued on next page

Note. SP = Single-player; MP = Multiplayer

Table 1 – continued from previous page

Author& Year	Game	Mode	Setting	Participants& Language	Results
Shahriarpour & Kafi (2014)	L.A. Noire	SP	Out-of-Class	25 EFL	Increased motivation toward vocabulary vs. traditional methods.
Reinders& Wattana (2015)	Ragnarok Online	MP	In-Class	30 EFL	Gameplay sessions led to decreased anxiety and increased WTC.
Chen (2016)	Second Life	MP	In-Class	9 EFL	Avatars helped learners overcome shyness and psychological barriers.
Shirazi et al. (2016)	The Walking Dead	SP	In-Class	40 EFL	Improved pragmatic competence in speech acts (apologies/requests).
Scholz & Schulze (2017)	World of Warcraft	MP	Out-of-Class	14 German	L2 development was nonlinear; shift from acquisition to use.
Azman & Dollsaid (2018)	World of Warcraft	MP	Out-of-Class	5 EFL	Anonymity reduced anxiety and facilitated negotiation of meaning.

Continued on next page

Note. SP = Single-player; MP = Multiplayer

Table 1 – continued from previous page

Author& Year	**Game**	**Mode**	**Setting**	**Participants& Language**	**Results**
Liang (2019)	Second Life	MP	In-Class	17 EFL	Improved complex storytelling and use of digital/visual resources.
Lai & Chen (2021)	The Lost Angel	SP	In-Class	30 EFL	Significant gains for both VR and PC; no platform difference found.
Calvo-Ferrer (2021)	Among Us	MP	In-Class	54 EFL	Active word use led to higher retention than passive encounter.
Ng et al. (2022)	Guild Wars 2	MP	Out-of-Class	4 EFL	Strategy use influenced by storyline and freedom from classroom rules.
Lai & Chen (2025)	Monkey Island	SP	In-Class	79 EFL	Paired gameplay outperformed individual play in retention tests.

Note. SP = Single-player; MP = Multiplayer; EFL = English Foreign Language

3 COTS Games in Digital Game-Based Language Learning

This chapter examines the commercial off-the-shelf (COTS) games that have appeared in previous studies related to computer-assisted language learning (CALL) and more specifically digital game-based language learning (DGBLL). Rather than providing an exhaustive catalogue of individual titles, the chapter focuses on identifying recurring patterns in game selection, genre preferences, learning affordances and the limitations reported across studies. By doing so, it aims to clarify why certain types of games have received more attention in the literature and how these choices shape reported learning outcomes.

A clear trend in existing research is the frequent use of massively multiplayer online role-playing games (MMORPGs), such as *World of Warcraft*, *EverQuest II*, *Ragnarok Online* and *Guild Wars 2*. These games are typically selected because they require players to interact with others in order to progress, creating conditions for spontaneous and goal-oriented language use[26]. Communication in these environments is not explicitly instructional but emerges naturally through collaboration, coordination, and social interaction, which aligns well with interaction-based views of second language learning. Acting through an avatar has been found to lower the communicative threshold; analysis of chat logs suggests that while learners may initially use crude or simplified language, they quickly adapt their behavior, adding

subtlety and pragmatic markers, when they realize that nuanced language yields better collaborative results in-game[26]. Furthermore, titles like *Ragnarok Online* have been utilized for their technical flexibility, allowing content modding to facilitate specific teaching themes, though this requires specialized technical expertise from the instructor[27].

Single-player games appear less frequently in DGBLL research. When they are used, they often belong to specific genres such as adventure or simulation games and are chosen for a specific linguistic purpose; these include titles like *The Sims*, *L.A. Noire*, *The Walking Dead* and *Return to Monkey Island*. These games provide rich and often authentic linguistic input through dialogue, narrative and environmental storytelling, but they offer fewer opportunities for real-time language production or negotiation with other players. For example, *The Sims* is selected because it simulates everyday situations, providing a high-fidelity mapping between digital objects and domestic vocabulary[24], [28]. As a result, they are less commonly chosen for studies that rely on short-term interventions or measurable interaction-based outcomes.

A second important distinction concerns the focus of gameplay itself. Many of the games most frequently examined in DGBLL research are mechanics-driven rather than narrative-driven. MMORPGs and other system-oriented games emphasize repeated actions, task completion, and progression systems, which support frequent exposure to vocabulary and formulaic language. This makes them particularly suitable for research on motivation and vocabulary acquisition, which are the two most commonly investigated outcomes in the field[23]. In contrast, narrative-heavy games prioritize story and dialogue over repetitive actions. While they provide substantial language input, they offer fewer distinct learning moments that can be easily measured, making it difficult to design experiments or assess progress.

3.1 Second Life

Peterson (2010) utilized an exploratory case study to examine the interaction patterns of intermediate EFL learners during seventy-minute chat sessions within the virtual world Second Life. The platform was selected for its immersive, task-based nature, which encouraged a high degree of learner autonomy and participation. The study setting focused on student-to-student interaction, where the majority of communication occurred in the target language. Findings indicated that the virtual environment fostered a low-stress atmosphere that students preferred over traditional classrooms, leading to increased politeness and social cohesion. However, while learners enjoyed the experience, few could identify specific instances of new vocabulary acquisition. Implications suggest that Second Life may be more effective for developing conversational fluency than linguistic accuracy, as the data revealed frequent uncorrected language errors. Limitations included the short time-frame of the study and the small group size, as well as initial technical hurdles that hampered early gameplay[29].

Wehner, Gump, and Downey (2011) conducted a comparative study to investigate how the use of Second Life affects learner motivation, anxiety, and attitudes toward language learning compared to a traditional classroom approach. The researchers chose this avatar-based environment because it provides a "psychologically safe" space where the fear of making mistakes is reduced. The study setting involved a university-level Spanish course where students utilized avatars to engage in informal, communicative language use. The results showed that students using Second Life reported significantly higher motivation and lower anxiety than those in the control group, while also developing more positive attitudes toward Spanish culture. The study implies that virtual worlds can effectively increase a learner's willingness to communicate (WTC) by providing a sense of anonymity. Key limitations include a small, non-random sample size and a reliance on self-reported motivational

measures rather than direct assessments of vocabulary or grammar development[30].

Peterson (2012) expanded on previous research by focusing specifically on the collaborative interactions between EFL students in Second Life. The game environment was selected for its ability to facilitate task-based learning and peer-scaffolding. The study setting involved learners engaging in collaborative tasks that required them to negotiate meaning and provide peer feedback on lexis and form. The findings revealed that participants actively maintained a positive social atmosphere through adaptive language use, with learner attitudes improving as they became more familiar with the interface. The study implies that the reduced social presence of avatars creates a less stressful environment for collaborative learning than face-to-face settings. However, the research was limited by a lack of engagement in the platform's content-creation features, a small number of participants, and a relatively short observational period[31].

Liou (2012) explored the integration of Second Life into a college-level CALL course in Taiwan through structured pedagogical tasks. The study utilized the platform to provide students with opportunities for contextualized language use and virtual cultural exploration. Findings suggested that students perceived significant benefits in vocabulary learning and cultural understanding, particularly during exploratory "in-world" activities. An important implication of this study is that the effectiveness of the virtual world depends more on instructional guidance and task design than on the technology itself. Significant limitations included severe technical issues, such as unstable internet access, which led to student frustration. Additionally, the study noted that voice-based interaction was underutilized and that substantial training time was required for students to become proficient with the game mechanics[32].

Liang (2012) investigated "language play" during collaborative role-playing activities within a narrative-driven Second Life environment. The platform was chosen

to observe how advanced learners utilize linguistic creativity, bilingual resources, and pop culture references during gameplay. The study setting involved competitive and collaborative tasks that pushed learners toward conversational fluency and identity performance. The results demonstrate that such gaming environments can foster high levels of pragmatic and semantic creativity. However, a major limitation was that lower-proficiency learners struggled with the interface demands, suggesting that the environment may be best suited for advanced students. Other limitations included the small sample size and a focus exclusively on text-based chat, which excluded the nuances of spoken interaction[33].

Chen (2016) examined Second Life as a task-based environment for adult EFL learners, utilizing authentic, real-world-inspired tasks such as virtual field trips and collaborative object building. The study selected this environment for its multimodal affordances, which allowed learners to engage in simulated real-world contexts. The setting fostered increased speaking confidence and productive vocabulary growth while reducing communicative anxiety. The implications highlight Second Life as a powerful supportive tool, though the researcher notes it is resource-intensive and requires significant instructor scaffolding. Limitations included the constraint of limited nonverbal cues, uneven participation among students, and technical instabilities. Furthermore, the reliance on self-reported data and a small sample size necessitates caution in generalizing the results[34].

Liang (2019) analyzed how L2 learners utilized the multimodal resources of Second Life for avatar-based storytelling tasks. The environment was used because it allowed students to combine verbal language with visual and digital resources, such as in-game objects and environment manipulation, to express humor and metaphor. The study setting involved students transitioning their narratives from face-to-face formats to digital video productions. The findings imply that Second Life is particularly effective for "multimodal meaning-making" rather than just linguistic accuracy.

However, limitations included the fact that some learners struggled to exploit these digital affordances without explicit instruction. Additionally, the study found that technical constraints and reduced embodiment in the digital format sometimes interfered with clear pronunciation and expressive potential[35].

3.2 Ragnarok Online

Reinders and Wattana (2011, 2015) examined the use of Ragnarok Online as part of university EFL courses in Thailand, motivated by persistent issues of learner reticence, communication anxiety, and low willingness to communicate in traditional classrooms. The game was selected due to its popularity among Thai students, strong social and collaborative mechanics, and the researchers' ability to host and modify the game on a private server. Across both studies, the game was extensively adapted through custom-designed quests aligned with course content, controlled dialogue with NPCs, and structured collaborative tasks requiring sustained interaction in English[36][27].

The 2011 pilot study demonstrated significant increases in L2 interaction, willingness to communicate, and positive affect over multiple gameplay sessions. Text-based chat produced higher quantities of output and greater grammatical accuracy, while voice-based chat encouraged more spontaneous interaction, clarification requests, and risk-taking. Native language use decreased over time, and students who were typically shy in face-to-face classes became more active participants. The 2015 follow-up study reinforced these findings through longitudinal interviews, showing that most participants experienced reduced anxiety, increased confidence, and greater motivation to communicate in English. Learners attributed these changes to the supportive, low-pressure environment created by avatar anonymity, collaborative goals, and the absence of overt teacher evaluation, leading to a virtuous cycle of increased participation and self-perceived competence[36].

Despite these positive outcomes, the studies also reveal key limitations relevant to COTS game-based learning. The educational success of Ragnarok Online depended heavily on extensive modification, including custom quests, controlled vocabulary, pedagogical sequencing, and structured pre- and post-game activities. Small sample sizes, reliance on self-reported data, and uneven learner engagement limit generalizability. One learner consistently reported negative experiences due to cognitive overload, unfamiliarity with gameplay, and difficulty managing real-time communication. The authors acknowledge that without such modifications and instructional scaffolding, commercial games are unlikely to align with curricular goals or support learners with lower proficiency. As such, these studies illustrate both the strong communicative affordances of MMORPGs and the practical challenge of using COTS games for language learning without significant external support[27].

3.3 World of Warcraft

Rama et al. (2012) conducted a qualitative case study to examine the affordances of the massively multiplayer online game (MMOG) World of Warcraft (WoW) for second language (L2) learning and socialization. The researchers chose the Spanish language version of WoW specifically because it offers an immersive, graphically rich 3D environment where learners can engage in informal, contextualized interaction with native speakers. The study setting involved two college-age Spanish learners—one experienced gamer and one novice—who were permitted to play the game at their own discretion rather than in a controlled classroom environment. Data were collected through participant observation, interviews, in-game chat logs, and student journals. The study implies that online games can be effective tools for language teaching if they leverage the medium's unique affordances for both experienced and inexperienced players. While the summary does not explicitly list limitations, it notes that communicative competence in WoW differs significantly

from face-to-face interaction, as players must manage simultaneous game actions and secondary chat[37].

Zheng, Newgarden, and Young (2012) utilized a multimodal analysis to investigate English language learning (ELL) through a 47-minute gameplay episode in World of Warcraft. The game was selected for its "ecological niche," which engages learners in a diverse tapestry of communicative activities—such as "killing" and "caring"—that are rarely found in traditional classroom settings. The study setting focused on small-group gameplay where players coordinated actions to achieve shared goals. A key implication of the research is the need to reconceive language learning as "skilled linguistic action" grounded in situated, technology-mediated L2 networks. The findings suggest that the social capital earned through in-game networking can be used to access dynamic language feedback and intercultural learning. A limitation inherent in the study's design is its focus on a single, short gameplay session, though it provides rich visualizations of interactional flow[38].

Scholz and Schulze (2017) explored second-language development (SLD) through the lens of complex adaptive systems (CAS) by observing university students playing the German version of World of Warcraft over four months. The game was used because its "text-heavy" nature facilitates the "near transfer" of linguistic constructions from the game environment to non-gaming contexts, such as group conversations. The study setting was extramural, allowing for authentic gameplay at the players' leisure, which addressed a common limitation in previous research where restricted classroom sessions hindered the observation of long-term learning trajectories. The researchers employed "retrodictive qualitative modeling" to capture the nonlinear variability of individual learners. The study implies that SLD in computer-assisted language learning (CALL) is a complex, non-reductionist process that must be analyzed over time. A noted limitation in the field this study sought to address, is the difficulty of substantiating the transfer of knowledge to external life activities[39].

Azman and Dollsaid (2018) investigated the use of MMOGs as "serious games" for English as a Foreign Language (EFL) teaching. The researchers focused on these games because they induce learners to become active generators of information and provide authentic, role-playing opportunities for communication among online multilingual speakers. The study setting was a case study investigating the communication behaviors and negotiation of meaning—such as clarification and comprehension checks—between novice and expert players in a non-threatening virtual environment. The implications of the study affirm the viability of online games as a teaching tool for "digital natives" in the 4.0 era. A limitation identified by the authors is that the high level of on-demand question-and-feedback exchanges seen in the game is often not viable in a traditional classroom situation[40].

3.4 Simulation, RPG, and Action Titles

This section examines studies that have utilized single-player and genre-specific COTS games outside the MMORPG category, organized into three thematic groupings. The first subsection addresses life-simulation and narrative-driven games, focusing on titles that leverage authentic environmental storytelling and dialogue to support vocabulary acquisition. The second subsection covers role-playing games, visual novels, and puzzle-based titles, which foreground narrative complexity and player decision-making as drivers of lexical engagement. The third subsection turns to action, music, and social interaction games, including titles where interactivity, competitive mechanics, and real-time communication shape — and sometimes complicate — language learning outcomes.

3.4.1 Life-Simulation and Narrative-Driven Games

Miller and Hegelheimer (2006) investigated the use of the popular life-simulation game *The Sims* to enhance vocabulary learning among adult ESL learners. The game was selected because its authentic representation of daily life activities provides a rich, context-driven environment where linguistic items are naturally mapped to visual actions. The study setting involved a five-week unit where learners used the game under varying conditions of supplemental support: mandatory materials, voluntary access, or no support. The findings imply that while simulation games are highly motivational, they are most effective as pedagogical tools when paired with structured materials that bridge the gap between gameplay and curricular goals. A key limitation noted was the difficulty in controlling for incidental vocabulary exposure, as the open-ended nature of the simulation meant players encountered words at different rates[24].

Ranalli (2008) expanded upon the research into *The Sims* by applying Chapelle's (2001) criteria for CALL task appropriateness to determine if a mass-market simulation could be rendered pedagogically beneficial. *The Sims* was chosen specifically because its "realistic animation" and "impressive interactivity" provide a cognitively engaging virtual world that mirrors real-world communication needs. The study setting took place in a university-level ESL classroom where students played the game in pairs using custom-designed supplementary materials. The study demonstrated statistically significant improvements in vocabulary knowledge and high levels of student engagement. However, the researcher noted that the lack of built-in pedagogical sequencing in COTS games remains a limitation, requiring instructors to provide significant external scaffolding to ensure learning outcomes are met[28].

Shahriarpour and Kafi (2014) explored the effect of the neo-noir detective game *L.A. Noire* on the English vocabulary motivation of Iranian EFL learners. The researchers chose this specific single-player game because its "popular" status among

teenagers and its heavy reliance on interrogation and investigation mechanics provide a compelling reason for learners to understand complex vocabulary to progress. The study setting utilized a combination of interviews and observations to track how intermediate learners engaged with the game's script-heavy content. The study implies that single-player games with high production values can serve as a powerful pedagogical strategy to overcome the "boring" nature of rote memorization. However, the study was limited by its small scope and the need for teachers to provide significant introductory support so that the game's difficulty did not discourage the learners[41].

Shirazi, Ahmadi, and Mehrdad (2016) examined the acquisition of pragmatic speech acts—specifically apologies and requests—through the single-player graphic adventure game *The Walking Dead*. This game was chosen because its core mechanic revolves around making difficult moral choices through dialogue trees, which provides a unique "facilitative tool" for developing interlanguage pragmatic competence in a way that traditional textbooks cannot. The study setting involved 40 Iranian intermediate EFL learners who were tested on their ability to recognize and produce appropriate speech acts after playing. The implications suggest that the emotional immersion and consequence-based narrative of single-player adventures force players to pay closer attention to the nuances of social interaction. Limitations included the relatively small sample size and the fact that the game's fixed dialogue options, while helpful for modeling, do not allow for the same level of free-form production found in multiplayer environments[42].

3.4.2 RPGs, Visual Novels, and Puzzles

Vandercruysse et al. (2013) investigated the impact of competition and player perception on learning outcomes using the fantasy role-playing game (RPG) *Divine Divinity*. This single-player title was selected because its rich narrative and complex

task structures allow for the isolation of specific game mechanics—namely "competition"—to see how they influence vocabulary acquisition and student motivation. The study setting involved 83 students working in a computer-based language learning environment where the researchers could manipulate the presence of a competitive leaderboard. By using a single-player game as a baseline, the researchers were able to probe whether the addition of social-competitive elements altered the students' perception of the environment as either a "gaming" or "learning" space. The study implies that while competition can boost motivation, its effectiveness is highly dependent on how students perceive the instructional intent. A primary limitation was the difficulty in disentangling the intrinsic appeal of the RPG's story from the external competitive mechanics being tested[43].

Lai and Chen (2021) conducted a comparative study on vocabulary learning using a sci-fi visual novel game, *The Lost Angel*, across both VR and PC platforms. This single-player visual novel was picked specifically because it necessitates reading and understanding dialogue choices to alter the plot and prevent the player character's "death." This design aligns with the "involvement load hypothesis," where the single-player nature ensures the learner must personally grapple with word meanings (need and search) to advance. The study setting involved university students divided into VR and PC groups to measure differences in immersion and retention. The findings imply that the high "salience" of words in a focused single-player narrative contributes significantly to long-term memory. Limitations noted include the potential for "VR sickness" to distract from the linguistic tasks and the limited interactivity of the visual novel genre compared to more open-ended games[44].

Lai and Chen (2025) investigated L2 vocabulary acquisition through the commercial off-the-shelf (COTS) adventure game *Return to Monkey Island*. The researchers selected this single-player title because of its sophisticated humor, text-heavy puzzles, and the "salience" of its key vocabulary. A central focus of the study was

comparing individual gameplay against paired gameplay to see if the social interaction of playing a single-player game together enhanced learning. The study setting involved 79 university students and utilized a mixed-methods approach, including translation tests and productive vocabulary levels tests. The results imply that word salience is a more reliable predictor of learning than frequency of exposure in narrative-driven games. A key limitation was that while paired gameplay encouraged discussion, some learners found it difficult to balance social coordination with the cognitive demands of the game's complex puzzles[45].

3.4.3 Action, Music, and Social Interaction

Rankin et al. (2009) explored the role of language socialization in the MMORPG *EverQuest II*, specifically focusing on the social interactions between native and non-native speakers. The game was selected due to its complex social structure and the necessity of coordinated dialogue to complete "quests," which creates a natural environment for second language acquisition (SLA). The researchers developed a specialized tool, ClockWerk, to visualize and analyze communication patterns during gameplay. The study implies that the high frequency of social interaction in MMORPGs facilitates a "third place" for learning that is distinct from traditional classrooms. A limitation of the study was the complexity of the game's interface, which sometimes overwhelmed novice players, suggesting that social benefits may be delayed until a certain level of game proficiency is reached[26].

deHaan, Reed, and Kuwada (2010) examined the effect of interactivity on vocabulary recall using the music video game *PaRappa the Rapper 2*. Unlike narrative-driven games, this title was chosen to test the "cognitive load" associated with simultaneous physical interaction (rhythm-matching) and linguistic processing. In a unique study setting, the researchers compared "players" against "watchers" who viewed the gameplay. Interestingly, the results showed that watchers recalled signif-

icantly more vocabulary than players, implying that the high interactivity of certain games can actually hinder learning by creating "extraneous cognitive load." This study provides a crucial counter-point to the assumption that more interactivity is always better for SLA. Limitations included the relatively short 20-minute gameplay session and the specific, repetitive nature of the music-game genre[46].

Calvo-Ferrer and Belda-Medina (2021) investigated incidental and intentional vocabulary learning through the multiplayer social deduction game *Among Us*. The game was selected for its recent global popularity and its heavy reliance on rapid, high-stakes communication during "emergency meetings," which forces players to use the target language to defend themselves or accuse others. The study setting involved university students participating in online sessions, comparing those who received intentional vocabulary instruction with those who learned incidentally through play. The findings imply that the competitive and urgent nature of the game increases the "salience" of specific terms, leading to better retention. However, technical issues and the varying proficiency levels of players were cited as limitations that could lead to uneven participation[47].

Ng, Azizie, and Chew (2022) studied the vocabulary learning strategies (VLS) employed by experienced ESL players in the MMORPG *Guild Wars 2*. This game was picked because its expansive storylines and collaborative community enrich the learning experience beyond simple gameplay. The researchers used a qualitative approach, conducting semi-structured interviews to identify metacognitive, cognitive, and memory-based strategies used during play. The study implies that the "freedom to learn" away from classroom rules allows players to develop highly personalized and effective VLS. A primary limitation was the reliance on experienced gamers, which leaves open the question of how novice players might navigate the same complex linguistic environment without prior gaming literacy[48].

3.5 Recurring Limitations

The studies reviewed across sections 3.1 through 3.4 consistently point toward a shared structural problem that cuts across genre, platform, and pedagogical context. Generally, COTS games were not designed with language learning in mind and therefore lack pedagogical sequencing, adaptive difficulty and explicit learner support[24]. While these games facilitate authentic exposure, learning outcomes often depend on external scaffolding. Furthermore, mechanical barriers can impede non-gamers; specifically, Quick Time Events (QTEs) in narrative titles can shift the learner's focus from language processing to motor-skill acquisition[47]. This is compounded by the risk of cognitive overload, where high levels of interactivity force the brain to prioritize game mechanics over the retention of new vocabulary[46]. Consequently, COTS games are often framed as supplementary tools for learners who already possess sufficient grammatical competence to follow the narrative. In such contexts, reported gains are primarily limited to vocabulary development through incidental encounters, while complex skills like syntactic development or controlled production are rarely observed.

At the same time, researchers note that educators have limited control over the linguistic content of COTS games. Vocabulary frequency and thematic relevance are determined by game design rather than pedagogical intent, making it difficult to align gameplay with specific learning objectives. To address this lack of control and analyze the resulting data, researchers have employed specialized tools like ClockWerk, which visually depicts communication patterns in text-based logs to determine how language complexity evolves[26]. These findings reinforce the view that the educational value of these games is highly variable and dependent on learner proficiency and external support, confirming that COTS games function more effectively as complementary language environments than as standalone instructional tools.

4 Experimental Setup

While the preceding chapters have established the pedagogical affordances of Digital Game-Based Language Learning (DGBLL) and the technical potential of OCR and LLM integration, the implementation of these findings within formal university curricula depends largely on the practitioners who manage those spaces. The literature demonstrates that games can facilitate significant linguistic gains; however, transitioning these gains into a formal setting requires an understanding of instructor needs.

This chapter presents the empirical part of this study, transitioning from theoretical foundations to a practical investigation into the professional stances of university-level language instructors. By situating proposed technical solutions within actual teaching practice, this study seeks to determine if the scaffolding provided by modern AI tools can address the barriers—such as high cognitive load and difficulty in tracking progress—that have previously hindered the adoption of complex digital games in formal education.

4.1 Semi-Structured Interview

To explore the intersection of pedagogy and digital entertainment, this thesis employs semi-structured interviews as its primary qualitative research method. Unlike structured interviews that rely on rigid scripts, the semi-structured format utilizes a flexible interview guide. This approach ensures that while core research objec-

tives are addressed consistently, there is sufficient conversational space to pursue unexpected insights.

The choice of semi-structured interviews is particularly suited for this study as it moves beyond binary data to uncover the underlying reasons for instructor attitudes. As noted by Newcomer et al. (2015), the strength of the semi-structured format lies in its ability to use closed-ended queries as gateways to open-ended probing, allowing the researcher to ask "why" or "how" following a specific response [49]. This provides the depth necessary to understand perceived barriers such as time constraints or the lack of specialized monitoring tools. The full interview guide is provided in Appendix A.

The cohort for this study consisted of four professional Japanese language educators. Participants were recruited through purposive sampling, specifically targeting instructors known to the researcher through prior academic engagement.

All participants were briefed on the research goals and the intended use of the data. Verbal informed consent was obtained from each participant before the start of the audio-recorded sessions.

4.2 Tool Prototype

The transition from theory to practice is materialized in the design of a tool mockup and prototype. This prototype explores a modular, non-invasive architecture that overlays existing games to address the pedagogical gap found in COTS titles. The prototype is a web-based application designed to bridge autonomous play and formal study. The prototype is accessible as a web application[1] [50].

[1] `https://game-learning-umber.vercel.app/`

4.2.1 Text Capture and Processing

The application facilitates text acquisition through two primary methods. Users can initiate a full-screen translation for immediate environmental context or utilize a localized selection tool to monitor specific regions, such as dialogue boxes.

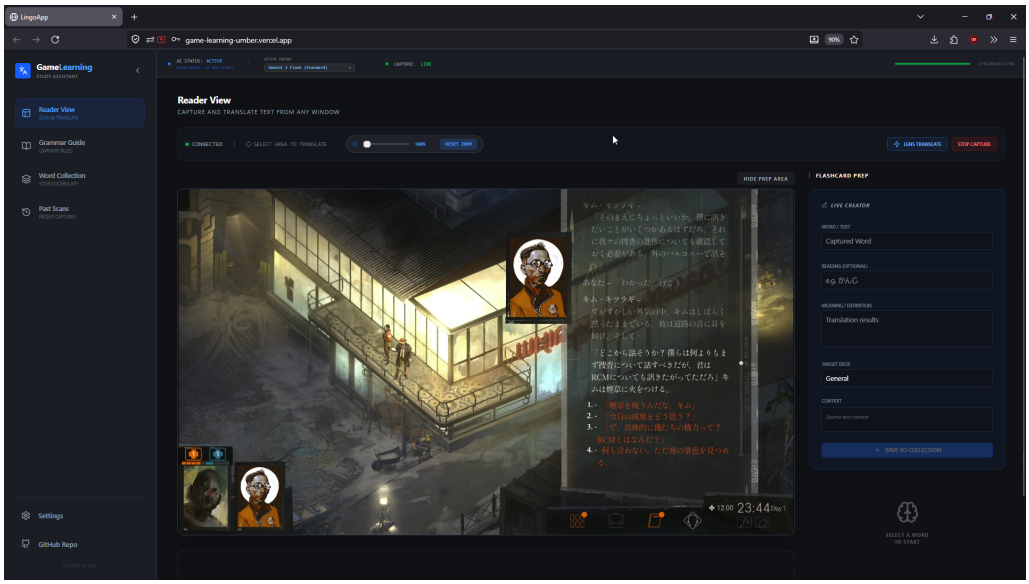

Figure 1: The primary web-based interface and capture selection view.

The system includes a confirmation modal that allows the user to verify the visual clarity of the captured text before it is processed. This step ensures that the Optical Character Recognition (OCR) provides high-quality input for the subsequent analysis. While the current prototype is optimized for the Gemini 3 Flash engine, the architecture is designed to be model-agnostic, allowing different Large Language Models (LLMs) to handle the linguistic breakdown based on the user's needs. The tool also features a grammar watchlist that can be configured to highlight desired patterns relevant to the learner via the LLM agent.

4.2.2 Morphological Scaffolding and Reader View

Once the text is processed, the results are displayed in a specialized Reader View. Rather than providing a simple full-sentence translation, the tool provides a deep

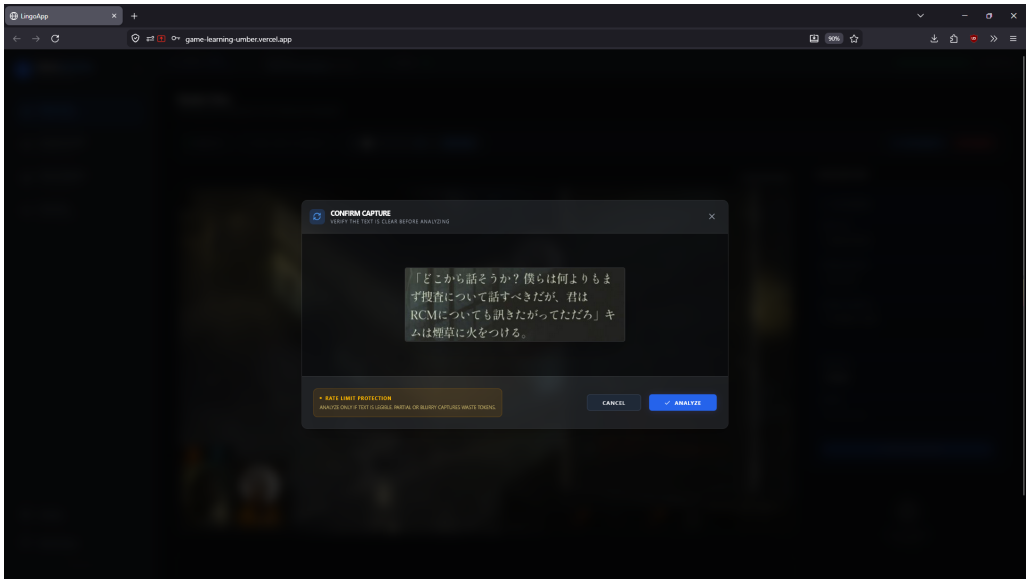

Figure 2: The confirmation modal for verifying OCR accuracy.

morphological analysis. The interface color-codes individual words by part of speech and highlights complex grammatical structures. When a user hovers over a highlighted term, a popup provides the specific definition and reading. This design allows students to parse authentic, complex sentences from games like *Disco Elysium* that would otherwise be beyond their independent reading level due to the character recognition barrier.

4.2.3 Flashcard Integration and Data Logging

To support long-term retention, the mockup includes an internal flashcard preparation sidebar. This feature allows learners to save queried words into a study collection. The interface automatically populates the word, its reading, and the specific sentence context from the game to provide a narrative anchor for the new vocabulary. While the flashcard system is currently internal to the mockup to provide clarity on the tool's utility, it demonstrates how autonomous gameplay can be converted into a structured review process. Furthermore, these captures are logged in a history view, providing the transparency required by instructors to validate a

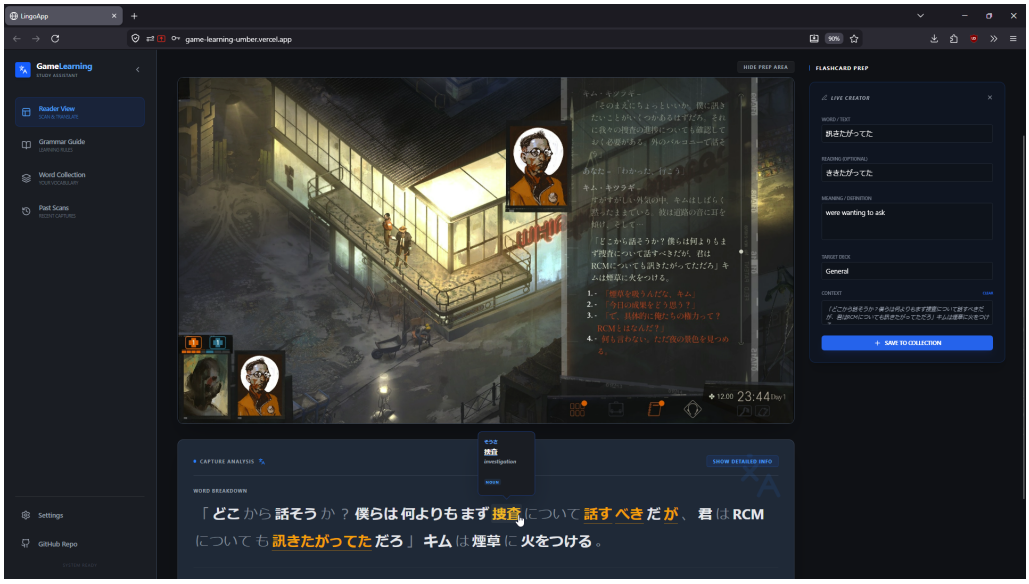

Figure 3: The morphological analysis view and internal flashcard preparation sidebar.

student's gaming session as a legitimate academic activity.

4.3 Interview Process and Data Analysis

The interview covered four professional Japanese language educators from university and vocational higher education settings. Their teaching experience ranges from ten to nearly twenty years. While their personal familiarity with digital gaming varied from low to moderate, all participants currently utilize digital tools like online flashcards or online dictionaries in their professional practice.

The interviews followed a standardized guide (see Appendix A) designed to move from general attitudes to technical requirements across three phases: (1) general perceptions of DGBLL, (2) analysis of the visual aids consisting of screenshots from the games *Disco Elysium* and *Skyrim*, and (3) evaluation of the tool mockup and prototype. The data were analyzed through a process of individual case summarization and comparative thematic synthesis. Each interview transcript was summarized to preserve the specific context of the instructor's background. These summaries were

then analyzed through close reading to identify common themes and recurring requirements, such as the need for morphological scaffolding over full translation and the desire for labor-saving automation.

4.3.1 General Perceptions of DGBLL

Educators consistently emphasized that commercial games serve as a vital bridge between academic study and practical application by exposing students to natural, colloquial speech and diverse personae that are often missing from standardized textbooks. These participants observed that narrative-driven games require high levels of cognitive engagement, forcing students to actively process cause-and-effect and make decisions based on their comprehension rather than passively consuming text. While some vocabulary encountered in these games may be highly specific, such as terminology related to police investigations, instructors argued that these terms are not useless because the game provides a natural context that makes memorization more engaging than traditional rote learning.

4.3.2 Visual Probing and Screenshot Analysis

Regarding visual barriers, a consensus emerged that the lack of reading aids like furigana serves as the primary obstacle for students, often creating a disconnect where a student's verbal proficiency far outpaces their ability to read complex kanji. Educators generally categorized the language level in these narrative-heavy environments as intermediate to advanced. They noted that the manual effort required to look up unknown characters is often so taxing that it causes students to lose focus on the story, effectively blocking their progress despite their communicative potential.

4.3.3 Mockup Evaluation and Technical Requirements

When evaluating the proposed tool, educators strongly opposed full-sentence translations, fearing that such features would cause students to stop thinking and rely entirely on the software. Instead, they advocated for features that provide morphological breakdowns and contextual examples to help students understand how grammatical elements function. There was significant interest in automated reports that log which words or grammar points a student struggles with, as this data would allow instructors to bridge the gap between a student's independent play and classroom instruction. Furthermore, participants highlighted that such a tool could significantly reduce the extensive time and effort teachers currently spend manually preparing vocabulary sheets and explanations for authentic gaming materials.

5 Results and Analysis

This chapter presents the findings of the study by opening with the primary empirical contribution: semi-structured interviews with four Japanese language educators, whose practical requirements for a bridge tool establish the lens through which the broader literature is then examined. The interview findings are followed by a synthesis of the systematic literature review, addressing how commercial games are selected for language learning (RQ1) and the technical and pedagogical barriers that limit their effectiveness (RQ2). Rather than treating these two strands as separate bodies of evidence, the chapter reads them in dialogue: the scaffolding gap, instructional opacity, and lack of curricular alignment that educators identify as obstacles in Section 5.1 recur as consistent absences across the empirical literature reviewed in Sections 5.2 through 5.4. Section 5.5 draws both strands together to show how the findings across all three research questions form a coherent pattern.

5.1 Teacher Attitudes and Interview Results

Semi-structured interviews with four Japanese language educators (Participants 1–4) reveal the practical requirements that must be met before game-based learning can be considered viable in institutional settings. The educators expressed consistent openness toward narrative-driven games, despite their own lack of gaming experience, viewing them as high-fidelity "digital novels" that offer exposure to authentic, colloquial language often absent from standard textbook-centered courses. However,

the transition from leisure play to valid educational activity is currently hindered by a high threshold of entry. To address this, the participants all identified a need for a tool that reduces cognitive load without inducing learner passivity through direct translations, emphasizing requirements for data filtering and teacher-led curation of learning materials.

5.1.1 Scaffolding and Automated Support

A central theme across all interviews was the tension between providing helpful scaffolding and avoiding over-reliance on technology. A language-specific barrier was also noted: in the case of Japanese, the learning of kanji without reading aids or accessible translation and extraction from the game was seen as the primary deterrent for learners. While the participants viewed the proposed tool mock-up as a feasible solution to this barrier, they expressed a cautious stance toward fully translating entire sentences.

Consequently, the primary pedagogical requirement identified is for morphological and grammatical breakdowns rather than finished translations. The participants suggested that the tool should isolate specific particles, conjugations, and grammar patterns as the learner encounters them. This approach ensures that the student remains an active participant in the construction of meaning, maintaining active engagement comparable to a traditional reading exercise.

5.1.2 Information Requirements and Data Filtering

To validate gaming as a legitimate learning session, educators indicated they require evidence of active engagement rather than raw volume of exposure. The participants prioritized reports that would map the student's individual hurdles, such as tracking which grammar points or words the student chose to query during gameplay. This data provides the instructor with a review of the student's current level and their

pain points.

Crucially, the participants required a way to filter out unnecessary noise and information, such as basic particles or grammar and vocabulary that should already be known by the students. Participants also suggested that the tool must allow teachers to curate the output, ensuring that the generated reports focus exclusively on topics relevant to their curriculum. By removing known language, the tool allows the instructor to monitor progress without being overwhelmed by irrelevant data points.

5.1.3 Course Alignment and Teacher-Directed Watchlist

The practical application of the tool was envisioned primarily as a resource for monitored self-study or elective assignments. A significant requirement emerged for a teacher-directed watchlist feature. In this configuration, the instructor inputs a specific set of grammar patterns currently being taught in the course, along with their own explanations and examples.

This feature addresses the scaffolding gap identified in the literature by creating a direct link between a student's leisure activity and the classroom syllabus. Participants 1 and 4 observed that such a system could transform the game into a "lead-in" for formal instruction, where the teacher can use in-game encounters as prompts for classroom discussion.

5.1.4 Obscurities in Reporting and Assessment

Despite the general consensus on tool utility, several areas remain unresolved regarding proper implementation. The most prominent uncertainty concerned the frequency and format of reporting. The workload of reviewing reports for a large cohort would become prohibitive depending on how often reports were generated and what they contained. There was no clear agreement on whether these reports

should be delivered per-session, per-stage, or only after encountering a sufficient quantity of new or important material or specific patterns of struggle.

Additionally, the educators remained uncertain about how to objectively verify deep comprehension. While tracking dialogue choices and what words needed translation provides a proxy for engagement, the participants noted that confirming whether a student understood the broader nuance or subtext of a narrative would still require human auditing. This suggests that while a bridge tool can facilitate exposure, reduce cognitive load, and promote language acquisition, it remains a supplementary device that necessitates a final layer of instructional verification.

5.2 Game Selection

The educator requirements identified in Section 5.1 — curricular alignment, scaffolded support, and instructor visibility — raise a prior question: what games has the research community actually used, and on what grounds were they chosen? The studies examined reveal that the selection of games for Second Language Acquisition (SLA) is rarely incidental; it is a choice between the technical architecture of the software and the needs of the learner. Traditionally, educators and researchers have focused on games that offer high situational authenticity. This explains the prominence of simulation titles like *The Sims* (Miller & Hegelheimer, 2006; Ranalli, 2008), where the virtual world serves as a direct mirror of everyday life[24] [28]. In these environments, the selection rationale for a COTS game in language learning is grounded in the need-to-know principle: a player cannot progress unless they comprehend domestic vocabulary and social directives.

However, the analysis suggests that the reason for choosing a particular game is often tied to familiarity and popularity rather than purely pedagogical design. In studies involving *World of Warcraft* (Scholz, 2017; Rama et al., 2012), the primary reason for selection was often that students were already familiar with the interface

and mechanics[39] [37]. This familiarity serves as a vital facilitator. By utilizing a game the student already understands, the cognitive resources typically spent on learning game controls are redirected toward processing the target language. Popularity also ensures a persistent community of native speakers, which is critical for collaborative discourse. In MMORPGs, this is further supported by having an avatar. Acting through a digital representative lowers the "affective filter," meaning students feel a lower sense of risk and anxiety when communicating in the target language. Research shows that while students may start with crude and simplified language in text chats, the collaborative nature of the game environment quickly incentivizes them to adopt more subtle and complex linguistic behaviors to achieve better in-game outcomes[26].

Furthermore, the choice of genre is dictated by the desired learning outcome, whether the goal is vocabulary breadth (simulations) or fluency (MMORPGs). This is most visible in studies involving technical customization. For example, *Ragnarok Online* was selected specifically because its architecture allowed researchers to modify the content to facilitate specific teaching themes (Reinders & Wattana, 2015) [27]. This highlights a crucial point: if teachers are able to act as content creators for the game world, it significantly helps the students, though this requires specialized technical expertise that may remain as a barrier for many educators.

5.3 Vocabulary Acquisition and Outcomes

The game selection patterns identified in Section 5.2 help explain what kinds of learning outcomes the literature has been able to measure. Because commercial games provide a high density of repetitive, contextualized lexical input, vocabulary acquisition stands out as the most consistent and tractable measurement of linguistic progress across the examined studies, with acquisition occurring as a natural byproduct of gameplay. The study by deHaan et al. (2010) explicitly frames this

process through the stages of noticing and recall, utilizing immediate and delayed vocabulary tests to quantify how much language is retained[46] .

The focus on vocabulary often involves a distinction between "breadth"—the total number of words recognized—and "depth," which involves the ability to use those words in specific contexts. This is evidenced in the work of Shirazi et al. (2016), where the narrative of *The Walking Dead* was used to measure how students acquire the specific vocabulary and pragmatic structures required for speech acts such as apologies and requests [42]. However, the data suggest that while vocabulary is easily measured, the sustainability of these gains is highly dependent on the "mining loop." Modern independent learners often close this gap using third-party tools like *Migaku* or *Game2Text* to export game-specific vocabulary into Spaced Repetition Systems (SRS). Notably, while these tools are prevalent in learner communities, they remain absent from formal research, which still relies largely on manual testing and pre-selected word lists.

5.4 Barriers to Implementation

The vocabulary gains documented in Section 5.3 are real, but they come with a set of structural conditions that the educators interviewed in Section 5.1 would immediately recognize: the learning that occurs is incidental, unsupported, and invisible to anyone outside the gameplay session. Despite high engagement levels, significant obstacles prevent COTS games from being seamless educational tools. A primary barrier is the assumption of player proficiency; people are not necessarily familiar with games and gaming, and the requirement to learn a game can distract from the language itself. While these mechanics provide the necessary immersion that sustains player motivation, they serve as mechanical hurdles for learners that can shift their focus from processing language to managing the interface.

This mechanical distraction is a core component of the trade-off between interac-

tion and cognition identified by deHaan et al. (2010). In their study of music-based video games, they found that the high cognitive load required to physically interact with the game actually hindered vocabulary recall. Paradoxically, students who played the game remembered significantly fewer words than those who simply watched the gameplay. This suggests that the mental effort needed to monitor game state and coordinate motor skills can compete with the cognitive resources required for language processing and acquisition. Consequently, when mechanics are too demanding, the interactivity becomes a limitation that diminishes pedagogical value[46].

Another recurring limitation is the scaffolding gap. Because COTS games are designed for entertainment, they lack the explicit feedback and grammar explanations found in dedicated educational software. Miller and Hegelheimer (2006) found that without supplemental bridge materials provided by the instructor, incidental learning remains disorganized[24]. Additionally, the social environment of MMORPGs often encourages an "internal language" or jargon—as seen in titles like *EverQuest II* or *Guild Wars 2*—that does not always translate well outside the game world, creating a gap between game-specific fluency and general language proficiency goals.

The complexity of these interactions is often measured through specialized computational tools. For example, the ClockWerk tool analyzes text-based conversation logs to map communication patterns[26]. These analyses show that students change their communication behavior quickly in collaborative settings; while language may be cruder or simplified at first, students eventually add subtlety to achieve better results. However, novices consistently use simpler language than experienced players, suggesting that linguistic development in a game is inherently tied to game mastery.

Finally, the sustainability of these gains is questioned by the varying study designs and short durations. While studies like Shirazi et al. (2016) demonstrate positive progress in acquiring speech acts like apologies and requests[42]. The longest

follow-up period in the selected studies remains short, often around four months[39]. This leaves the long-term retention of language acquired in these high-interactivity environments as an area requiring further longitudinal research.

5.5 Answering the Research Questions

Regarding the selection of commercial games for language learning (RQ1), the evidence shows that titles are chosen primarily based on motivational and contextual factors—such as player familiarity, situational authenticity, and social architecture—rather than inherent pedagogical design. Educators and researchers often prioritize games that students already understand to ensure that cognitive resources are directed toward processing the target language rather than mastering game mechanics. Furthermore, the choice of genre is typically dictated by the desired linguistic outcome, with simulation games favored for vocabulary breadth and MMORPGs for communicative fluency. This non-pedagogical selection logic is not incidental; it serves as the foundational reason for the structural barriers identified in the subsequent research questions.

The recurring limitations identified in these commercial games (RQ2) converge on a shared tension between entertainment and instruction, manifested through cognitive overload and a significant scaffolding gap. The high mental effort required to manage complex game states and motor skills can compete with the cognitive resources needed for language processing, which can paradoxically hinder recall in highly interactive environments. Additionally, because these games are not designed for education, they lack the explicit feedback and grammar explanations necessary for organized progress. This instructional invisibility makes it difficult for teachers to monitor student growth or ensure that gameplay aligns with specific learning objectives, often resulting in gains that are limited to incidental vocabulary acquisition rather than deep syntactic development.

Finally, the investigation into instructor attitudes and technical requirements (RQ3) confirms that while educators view an AI-assisted bridge tool as a viable solution, its implementation must meet specific pedagogical standards to be accepted in formal settings. Educators strongly advocated for morphological and grammatical breakdowns over full-sentence translations to maintain learner agency and prevent passivity, particularly in the context of the Japanese script barrier. They also identified a critical need for curriculum-filtered reporting and teacher-directed watchlists, which would allow them to prioritize specific language points and bridge the gap between autonomous play and classroom syllabi. Ultimately, the findings suggest that while such a tool is a necessary condition for integrating games into formal instruction, it remains a supplementary device that requires human verification to confirm deep narrative comprehension and subtextual understanding.

5.6 Limitations

Several limitations directly affect how broadly the conclusions of this study can be applied. First, all four interview participants teach Japanese in similar institutional settings, meaning their responses were shaped by challenges specific to Japanese script—particularly kanji recognition and the absence of reading aids. As a result, the tool requirements identified, such as morphological breakdown and furigana support, reflect this linguistic context specifically. Educators working with alphabetic or tonal languages face different barriers, and the same tool design may not map onto their needs in the same way. Second, the bridge tool was presented as a partial prototype rather than a live system, meaning participant responses reflect anticipated rather than experienced utility, which introduces a risk of overestimating usability and underestimating friction in real classroom conditions. The actual impact on student engagement or proficiency remains untested, and it is possible that live usage would surface requirements or objections not captured here. Finally, restricting

the literature review to English-language publications likely excluded relevant work from other parts of the world, as well as community-level documentation of tools like Migaku and Game2Text. This means the review may under represent both the technical sophistication of existing tools and the cultural contexts in which game-based learning is most actively practiced.

6 Conclusion

The core problem this thesis addresses is the following: students are already using commercial games to learn languages, right now, without institutional support or recognition. They are navigating complex linguistic environments alone, building vocabulary, encountering authentic grammar, and developing real communicative intuition—none of which rarely counts for anything in a formal academic context. At the same time, educators often lack visibility into what their students are doing or how it might connect to classroom goals. This matters because the gap is not a technical one—the tools to bridge it largely exist. It is an organizational and pedagogical one. Until gaming sessions can be made transparent, curricular alignment remains impossible and the learning goes unacknowledged. Closing this gap does not mean turning games into classrooms. It means recognizing that meaningful language learning is already happening outside them, and building the infrastructure to take it seriously.

To investigate this infrastructure, this research utilized a two-stage methodology. First, a systematic literature review of 24 empirical studies was conducted to identify the pedagogical limitations of current DGBLL practices. Second, an empirical study involving semi-structured interviews with four university-level Japanese language educators was performed. These interviews utilized a bridge tool prototype—integrating OCR and LLM technology—to evaluate how automated scaffolding and data reporting could meet the requirements of formal instruction.

The findings indicate that while commercial games provide rich, authentic input, they impose a high cognitive load that often necessitates external support. The literature review highlighted a persistent scaffolding gap, where students struggle to transition from passive consumption to active acquisition. The empirical data further revealed that for educators to bridge this gap, they require more than just translated text; they need granular morphological analysis and session reports that are filtered through the specific lens of their curriculum. This study concludes that the integration of commercial games into formal education is contingent upon making the learner's process of encountering, querying, and reviewing language in game transparent and actionable for the instructor.

6.1 Contribution

The primary contribution of this thesis is twofold. First, it provides a synthesized analysis of game selection patterns and practical limitations across 24 empirical DGBLL studies, offering a structured evidence base that is currently fragmented across the literature. Second, it identifies and preliminarily validates a design space for a class of tools—here termed bridge tools—that have existed in independent learner communities for some time, such as Game2Text, but have not previously been examined within a formal pedagogical framework. The interviews demonstrate that the infrastructural gap between autonomous play and institutional recognition is not merely technical but organizational: educators require transparency, curatorial control, and curricular alignment before they can validate gaming as a legitimate academic activity.

6.2 Further Research

Three directions for further research follow directly from the findings. The most immediate arises from educator concern about workload: participants were open to reviewing student session reports but uncertain about frequency and volume. Future research should establish what an optimal reporting threshold looks like in practice—how much data an instructor can meaningfully engage with per student, per week, without it becoming a professional burden. This cannot be determined without a live implementation study. Second, the tension between morphological scaffolding and full-sentence translation—which educators strongly favored resolving toward the former—remains empirically untested. It is not yet known whether AI-generated grammatical breakdowns actually produce better long-term retention than translations would. A controlled classroom study comparing these two modes of support would directly address this gap. Third, participants acknowledged that session reports can confirm a student queried certain words but cannot confirm genuine comprehension of narrative subtext. Future tool development should explore whether automated "hurdle detection"—flagging patterns of repeated queries around the same grammar point, for instance—might serve as an indirect indicator of comprehension difficulties, reducing the need for manual instructor auditing.

References

[1] Z. Dörnyei, "The psychology of the language learner: Individual differences in second language acquisition", *The Psychology of the Language Learner: Individual Differences in Second Language Acquisition*, pp. 1–270, Apr. 2014. DOI: 10.4324/9781410613349.

[2] H. Reinders and C. White, "20 years of autonomy and technology: How far have we come and where to next?", *Language, Learning and Technology*, vol. 20, pp. 143–154, Jun. 2016. DOI: 10.64152/10125/44466.

[3] S. Deterding, D. Dixon, R. Khaled, and L. Nacke, "From game design elements to gamefulness: Defining gamification", *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments, MindTrek 2011*, vol. 11, pp. 9–15, Sep. 2011. DOI: 10.1145/2181037.2181040.

[4] J. Hamari, J. Koivisto, and H. Sarsa, "Does gamification work? – a literature review of empirical studies on gamification", in *2014 47th Hawaii International Conference on System Sciences*, 2014, pp. 3025–3034. DOI: 10.1109/HICSS.2014.377.

[5] J. Juul, *Half-real: Video Games Between Real Rules and Fictional Worlds*. MIT press, 2011.

[6] K. S. Tekinbas and E. Zimmerman, *Rules of Play: Game design Fundamentals*. MIT press, 2003.

[7] E. J. Aarseth, *Cybertext: Perspectives on ergodic literature*. JHU Press, 1997.

[8] J. Koivisto and J. Hamari, "The rise of motivational information systems: A review of gamification research", *International Journal of Information Management*, vol. 45, pp. 191–210, 2019, ISSN: 0268-4012. DOI: https://doi.org/10.1016/j.ijinfomgt.2018.10.013.

[9] R. van Roy and B. Zaman, "Why gamification fails in education and how to make it successful: Introducing nine gamification heuristics based on self-determination theory", in *Serious Games and Edutainment Applications : Volume II*, M. Ma and A. Oikonomou, Eds. Cham: Springer International Publishing, 2017, pp. 485–509, ISBN: 978-3-319-51645-5. DOI: 10.1007/978-3-319-51645-5_22. [Online]. Available: https://doi.org/10.1007/978-3-319-51645-5_22.

[10] Z. Luo, "Gamification for educational purposes: What are the factors contributing to varied effectiveness?", *Education and Information Technologies*, vol. 27, pp. 1–25, Jan. 2022. DOI: 10.1007/s10639-021-10642-9.

[11] J. Tao, S. Ferrão, F. Vitoria, and R. Silva, "Gamification for technology-enhanced language learning (TELL) – success factors of gamified language learning platform design", Jun. 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:226411495.

[12] J. R. Chapman and P. J. Rich, "Does educational gamification improve students' motivation? If so, which game elements work best?", *Journal of Education for Business*, vol. 93, no. 7, pp. 315–322, 2018. DOI: 10.1080/08832323.2018.1490687. [Online]. Available: https://doi.org/10.1080/08832323.2018.1490687.

[13] M. Shortt, S. Tilak, I. Kuznetcova, B. Martens, and B. Akinkuolie, "Gamification in mobile-assisted language learning: A systematic review of duolingo

literature from public release of 2012 to early 2020", *Computer Assisted Language Learning*, vol. 36, May 2021. DOI: 10.1080/09588221.2021.1933540.

[14] K. Ishaq, N. Mat Zin, F. Rosdi, M. Jehanghir, S. Ishaq, and A. Abid, "Mobile-assisted and gamification-based language learning: A systematic literature review", *PeerJ Computer Science 7:e496*, 2021. [Online]. Available: https://doi.org/10.7717/peerj-cs.496.

[15] R. Araya, E. Arias Ortiz, N. L. Bottan, and J. P. Cristia, *Does gamification in education work?: Experimental evidence from Chile*, 2019. DOI: 10.18235/0001777. [Online]. Available: https://doi.org/10.18235/0001777.

[16] J. M. O'malley and A. U. Chamot, *Learning Strategies in Second Language Acquisition*. Cambridge university press, 1990.

[17] R. L. Oxford, "Language learning styles and strategies", *Georgetown University Round Table on Languages and Linguistics (GURT) 1990: Linguistics, Language Teaching and Language Acquisition: The Interdependece of Theory, Practice and Research*, p. 438, 1990.

[18] C. K. Lee, "An overview of language learning strategies.", *Annual Review of Education, Communication & Language Sciences*, vol. 7, 2010.

[19] C. White, "Language learning strategies in independent language learning: An overview.", *Language Learning Strategies in Independent Settings*, pp. 3–24, 2008.

[20] C. S. Ang and P. Zaphiris, "Computer games and language learning", in Jan. 2008, pp. 449–462, ISBN: 978-159904-988-5. DOI: 10.4018/9781599048659.ch032.

[21] A. Chik, "Digital gaming and language learning: Autonomy and community", *Language Learning & Technology*, vol. 18, no. 2, pp. 85–100, 2014. [Online]. Available: http://llt.msu.edu/issues/june2014/chik.pdf.

[22] F. A. Kronenberg, "Selection criteria for commercial off-the-shelf (cots) video games for language learning", *IALLT Journal of Language Learning Technologies*, vol. 42, no. 2, pp. 52–78, 2012.

[23] Z. Xu, Z. Chen, L. Eutsler, Z. Geng, and A. Kogut, "A scoping review of digital game-based technology on english language learning", *Educational Technology Research and Development*, vol. 68, no. 3, pp. 877–904, 2020.

[24] M. Miller and V. Hegelheimer, "The Sims meet ESL incorporating authentic computer simulation games into the language classroom", *Interact. Techn. Smart Edu.*, vol. 3, pp. 311–328, Nov. 2006. DOI: 10.1108/17415650680000070.

[25] J. Li, "A systematic review of video games for second language acquisition", in *Handbook of Research on Integrating Digital Technology With Literacy Pedagogies*, P. M. Sullivan, J. L. Sullivan, J. L. Lantz, and B. A. Sullivan, Eds. Hershey, PA: IGI Global, 2020, pp. 472–499. DOI: 10.4018/978-1-7998-0246-4.ch021.

[26] Y. Rankin, D. Morrison Wells, M. McNeal, B. Gooch, and M. Shute, "Time will tell: In-game social interactions that facilitate second language acquisition", *FDG 2009 - 4th International Conference on the Foundations of Digital Games, Proceedings*, pp. 161–168, Apr. 2009. DOI: 10.1145/1536513.1536546.

[27] H. Reinders and S. Wattana, "Affect and willingness to communicate in digital game-based learning", *ReCALL*, vol. 27, no. 1, pp. 38–57, 2015.

[28] J. Ranalli, "Learning english with the sims: Exploiting authentic computer simulation games for L2 learning", *Computer Assisted Language Learning*, vol. 21, pp. 441–455, Dec. 2008. DOI: 10.1080/09588220802447859.

[29] M. Peterson, "Learner participation patterns and strategy use in Second Life: An exploratory case study", *ReCALL*, vol. 22, no. 3, pp. 273–292, 2010.

[30] A. K. Wehner, A. W. Gump, and S. Downey, "The effects of Second Life on the motivation of undergraduate students learning a foreign language", *Computer Assisted Language Learning*, vol. 24, no. 3, pp. 277–289, 2011.

[31] M. Peterson, "EFL learner collaborative interaction in Second Life", *ReCALL*, vol. 24, no. 1, pp. 20–39, 2012.

[32] H.-C. Liou, "The roles of Second Life in a college computer-assisted language learning (call) course in Taiwan, ROC", *Computer Assisted Language Learning*, vol. 25, no. 4, pp. 365–382, 2012.

[33] M.-Y. Liang, "Foreign ludicity in online role-playing games", *Computer Assisted Language Learning*, vol. 25, no. 5, pp. 455–473, 2012.

[34] J. C. Chen, "The crossroads of english language learners, task-based instruction, and 3d multi-user virtual learning in Second Life", *Computers & Education*, vol. 102, pp. 152–171, 2016.

[35] M.-Y. Liang, "Beyond elocution: Multimodal narrative discourse analysis of L2 storytelling", *ReCALL*, vol. 31, no. 1, pp. 56–74, 2019.

[36] H. Reinders and S. Wattana, "Learn english or die: The effects of digital games on interaction and willingness to communicate in a foreign language", *Digital Culture and Education*, vol. 3, pp. 4–28, Apr. 2011.

[37] P. S. Rama, R. W. Black, E. Van Es, and M. Warschauer, "Affordances for second language learning in World of Warcraft", *ReCALL*, vol. 24, no. 3, pp. 322–338, 2012.

[38] D. Zheng, K. Newgarden, and M. F. Young, "Multimodal analysis of language learning in World of Warcraft play: Languaging as values-realizing", *ReCALL*, vol. 24, no. 3, pp. 339–360, 2012.

[39] K. W. Scholz and M. Schulze, "Digital-gaming trajectories and second language development", 2017.

[40] H. Azman and N. F. Dollsaid, "Applying massively multiplayer online games (MMOGs) in EFL teaching.", *Arab World English Journal*, vol. 9, no. 4, pp. 3–18, 2018.

[41] N. Shahriarpour et al., "On the effect of playing digital games on Iranian intermediate EFL learners' motivation toward learning English vocabularies", *Procedia-Social and Behavioral Sciences*, vol. 98, pp. 1738–1743, 2014.

[42] M. Shirazi, S. D. Ahmadi, and A. G. Mehrdad, "The effect of using video games on efl learners' acquisition of speech acts of apology and request", *Theory and Practice in Language Studies*, vol. 6, no. 5, pp. 1019–1026, 2016.

[43] S. Vandercruysse, M. Vandewaetere, F. Cornillie, and G. Clarebout, "Competition and students' perceptions in a game-based language learning environment", *Educational Technology Research and Development*, vol. 61, no. 6, pp. 927–950, 2013.

[44] K.-W. Lai and H. Chen, "A comparative study on the effects of a VR and PC visual novel game on vocabulary learning", *Computer Assisted Language Learning*, vol. 36, pp. 1–34, Jun. 2021. DOI: 10.1080/09588221.2021.1928226.

[45] K.-W. K. Lai and H.-J. H. Chen, "L2 vocabulary learning through a COTS adventure game: Individual vs. paired gameplay and the influence of word exposure frequency, word level, and word salience", *Computers & Education*, p. 105 343, 2025.

[46] J. deHaan, W. Reed, and K. Katsuko, "The effect of interactivity with a music video game on second language vocabulary recall", *Language Learning and Technology*, vol. 14, Jun. 2010. DOI: 10.64152/10125/44215.

[47] J. R. Calvo-Ferrer and J. Belda-Medina, "The effect of multiplayer video games on incidental and intentional L2 vocabulary learning: The case of Among Us",

Multimodal Technologies and Interaction, vol. 5, no. 12, 2021, ISSN: 2414-4088. DOI: 10.3390/mti5120080. [Online]. Available: https://www.mdpi.com/2414-4088/5/12/80.

[48] L. L. Ng, R. Azizie, and S. Y. Chew, "Factors influencing ESL players' use of vocabulary learning strategies in massively multiplayer online role-playing games (MMORPG)", *The Asia-Pacific Education Researcher*, vol. 31, May 2022. DOI: 10.1007/s40299-021-00578-6.

[49] K. Newcomer, H. Hatry, and J. Wholey, *Handbook of Practical Program Evaluation: Fourth edition*. Oct. 2015, pp. 1–864, ISBN: 9781118893609. DOI: 10.1002/9781119171386.

[50] J. Tornberg, *Gamelearning tool prototype*, https://game-learning-umber.vercel.app/, Published: 2026-04-18, 2026.

Appendix A Interview Guide

1. Background, Demographics, Experience with Games

- How familiar are you with digital games in general? On a scale of 1–5, how would you rate your own personal familiarity with video games?
- How long have you been teaching languages at the university level?
- To what extent do digital tools or mobile applications currently feature in your teaching or your students' self-study? (e.g. dictionaries, translation tools)
- Have you ever used or considered using digital games as a supplementary resource for students? (Not gamified apps such as Duolingo)

2. Attitudes toward Digital Game-Based Language Learning (DGBLL)

- What is your general impression of students playing commercial games in their free time as a method for language exposure?
- What do you perceive as the primary benefits of using narrative-driven games for language acquisition?
- What are the most significant deterrents or deal-breakers that prevent you from incorporating games into your formal curriculum?

3. Visual Analysis and Probing

- *[Show Screenshot 1: Authentic In-Game Dialogue/Quest Logs]*
  Looking at the density and complexity of this text, what specific linguistic challenges do you think your students would encounter here?

- *[Show Screenshot 2: Proposed Tool Mock-Up/Overlay]*
  If a student had access to an overlay providing real-time contextual explanations for this text, how would that influence your view of the game's educational feasibility?

- Does this type of automated scaffolding address the concerns you previously mentioned regarding cognitive load or difficulty?

4. Game Session Information and Utility

- If a student spent several hours playing a game in the target language, what specific information from that session would you need to see to consider it a valid learning activity?

- How useful would it be to receive a vocabulary list or summary of grammar points a student encountered during gameplay?

- How would you potentially use this information to offer help and guidance to a student?

- What format should this information take to be actionable without requiring you to play or understand the game itself?

5. Feasibility and Institutional Integration

- What are the main limitations you foresee in terms of assessment when using commercial games?
- How much time could you realistically dedicate to reviewing student gameplay data or dashboards on a weekly basis?
- What kind of learning environment could game-based learning fit into? (e.g. one-on-one vs. classroom settings)

6. Concluding Thoughts

- Where do you see the role of AI-assisted language tools heading?
- Would you be interested in trying out such an application in the future?