UNIVERSITY OF CALIFORNIA
SANTA CRUZ

EVIDENCE-BASED GAMIFICATION AND MNEMONICS FOR LOGOGRAPHIC WRITING SYSTEMS.

A dissertation submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

COMPUTATIONAL MEDIA

by

Oleksandra G. Keehl

March 2023

The Dissertation of Oleksandra G. Keehl is approved:

Professor Edward Melcer, Chair

Professor Sri Kurniawan

Ryan Redmond, Ph.D.

Peter Biehl
Vice Provost and Dean of Graduate Studies

Table of Contents

List of Figures vi

List of Tables x

Abstract xi

Dedication xiii

Acknowledgments xiv

1 Introduction 1
- 1.1 Core Concepts . . . 3
- 1.2 Research Questions . . . 5
- 1.3 Relevance and Contributions . . . 5
- 1.4 Methodology, Goals and Principles . . . 7
- 1.5 Dissertation Overview . . . 8

2 Background 10
- 2.1 Challenges and methods of learning logographic characters . . . 11
- 2.2 Motivation . . . 15
  - 2.2.1 Motivation in Education . . . 16
  - 2.2.2 Motivation in non-Educational Digital Games and Gamification . . . 19
  - 2.2.3 From Motivation to Addiction . . . 24
- 2.3 Digital Educational Games . . . 26
  - 2.3.1 Unique challenges . . . 27
  - 2.3.2 Recent Commercial examples . . . 29
  - 2.3.3 Education Research via Game-Making . . . 35
- 2.4 Mobile Games . . . 38
  - 2.4.1 Unique Restrictions and Affordances . . . 38
  - 2.4.2 Psychology behind mobile game appeal . . . 40
  - 2.4.3 Free-to-Play (F2P) games . . . 40
  - 2.4.4 Minigames . . . 41

2.4.5 Features linked to success . . . 45

3 Mobile Game Case Studies **49**
3.1 Case Study 1. The Walking Dead: Our World . . . 49
3.1.1 Game Overview . . . 49
3.1.2 Clan Mechanic . . . 50
3.1.3 Periodic Quests . . . 54
3.1.4 Educational Game Application . . . 55
3.2 Case Study 2. Cat Game – The Cats Collector! . . . 56
3.2.1 Game overview . . . 56
3.2.2 Unlockable Content . . . 58
3.2.3 Supplementary Activities . . . 59
3.2.4 Educational Game Application . . . 61

4 Radical Tunes: A Mnemonic Musical Game for Learning to Write Kanji **64**
4.1 Game Overview . . . 64
4.2 Study 1 . . . 66
4.2.1 Methods . . . 66
4.2.2 Results . . . 71
4.3 Study 2 . . . 73
4.3.1 Rationale . . . 73
4.3.2 Methods . . . 74
4.3.3 Results . . . 79
4.4 Discussion . . . 83
4.4.1 Music and Immersion . . . 83
4.4.2 The Meaning Mystery . . . 83
4.4.3 Scalability . . . 84
4.4.4 Implications for Design . . . 84
4.5 Limitations and Future Works . . . 85
4.6 Conclusion . . . 86

5 Zen Hanzi (怎汉字): A Component-Focused Game for Learning Hanzi **90**
5.1 Game Overview . . . 90
5.1.1 Rationale . . . 93
5.1.2 *Zen Hanzi* Design Principles . . . 95
5.1.3 Designing for Content Scalability . . . 96
5.1.4 Unconventional Hanzi Component Design . . . 98
5.1.5 What *Zen Hanzi* is and isn't . . . 100
5.2 Methodology . . . 100
5.2.1 Quizlet . . . 101
5.2.2 Hypotheses . . . 102
5.2.3 Participants . . . 103

5.2.4 Procedure . . . 104
5.2.5 Educational Content . . . 107
5.2.6 Measurements . . . 107
5.3 Results . . . 108
5.4 Discussion . . . 109
5.4.1 Component-focused vs. Whole-hanzi Design . . . 109
5.4.2 Novelty . . . 110
5.4.3 Competence . . . 110
5.5 Limitations and future work . . . 111
5.5.1 Design Limitations . . . 111
5.5.2 Evaluation Limitations . . . 112
5.6 Conclusion . . . 112

6 Discussion **114**
6.1 Variable Rate Reinforcement, Gacha, and the Loot Box Controversy . . . 114
6.2 Zeigarnik (a.k.a. the Unfinished Business) effect . . . 121
6.3 Ethics . . . 124

7 Conclusion **125**
7.1 Revisiting the Research Questions . . . 126
7.2 Looking to the Future . . . 128

Bibliography **129**

List of Figures

1.1 Screenshots of *Skritter*, a game by Chineasy Limited, with visual mnemonics for various hanzi components. As these two illustrations show, some hanzi components lend themselves to visual mnemonics better than others. 2

1.2 Screenshots of apps on Google Playstore listed as games for learning Japanese and Chinese. Left to right: Learn Japanese Yami [189], Chinese Dungeon: Learn C-Word [183], Kanji Battle [11]. 4

2.1 The anatomy of kanji. A kanji can consist of one or more radicals, which in turn consist of one or more strokes. While radicals can carry a meaning of their own (which may or may not relate to the meaning of the kanji), strokes are purely graphical elements. 12

2.2 WaniKani's 'poop' radical is a great example of shock-value driven mnemonics. 15

2.3 The Octalysis framework for classification of user/player engagement mechanics, developed by YuKai Chou. 21

2.4 Duolingo Android app screenshots left to right: the achievements screen, the league screen, and avatar customization screen. 30

2.5 Screenshots of Chinese Spy. From top to bottom: in-game narrative, bomb-disarming exercise, decoding secret message. 32

2.6 Learn Japanese to Survive: Kanji Wars screenshots.Screenshots of Learn Japanese to Survive. Upper left: the game's impressive character design; Upper right: the player encounters an enemy in an RPG setting; Lower left: choosing "Answer" during the battle results in the familiar mechanic of "select correct answer to hit."; Lower right: one of the lessons offered to the player at the beginning of each game day. 34

2.7 Screenshot of Refraction, an educational game for learning fractions. . . 36

2.8 The Cat Game screenshots [69]. (a) The game's main screen where the player's cat collection takes place. (b) The Poppy Cats minigame, where the player must quickly merge same-colored cats into biggest possible combos and pop them to earn points until the time bar below runs out. (c) The Blocky Cats minigame, where the player places given parts onto the field, aiming to create complete lines, which disappear similar to Tetris. There's no timer, and the game is over when the player can't fit a given shape onto the screen. 42
2.9 Castle Cats minigame screenshots [184] (These minigames are duplicated in the same company's dog version of the game Dungeon Dogs [185].) Top: Murrmory (or Donna's memory in the dog version) is a simple memory matching game. Bottom: Matching Mania is a variation of the match-3 Bejeweled clone. 43

3.1 TWD: Our World screenshots. Left to right: the game's overworld view; FPS battle mode; Power-up screen. 50
3.2 TWD: Our World screenshots. (a) Weekly clan challenge board (b) Internal clan chatroom (c) Global clan leaderboard (d) Seasonal challenge tiers (e) Player profile. 51
3.3 TWD: Our World screenshots. Left to right: Daily individual mission list; Weekly individual mission list; Overworld with interactable items on reset countdown. 54
3.4 Screenshots of Cat Game: a. home screen; b. crafting screen; c. supplementary activities selection screen with a level-locked item; d. avatar and showroom customization screen. 57
3.5 Screenshots of Cat Game: a. daily room decoration contest; b. contest voting screen; c. Blocky Cats minigame; d. Poppy Cats minigame. . . . 60
3.6 Left: screenshot of Radical Tunes, a musical game for learning to write kanji [111]. Right: screenshot of Zen Hanzi, a puzzle-like hanzi-learning game [109]. 62

4.1 Melodies for the radicals "tree" (above) and "water" (below). I tried to make the pitch of the melody follow the direction of the strokes. A horizontal stroke is represented by a steady note. A long downward stroke is accompanied by a significant drop in pitch, while a shorter upward stroke is matched by a moderate rise in pitch. 66
4.2 Radicals and kanji used in the first study. Left to right: person, mouth, tree, rest, episode, annoyed. 67
4.3 *Radical Tunes* screenshots. (a) Meaning pre- and post-test. (b) Stroke order pre- and post-test. (c) Stroke learning screen. (d) Timed and scored game portion. 68

4.4 Screenshots of Radical Tunes. Left: the meaning test screen. Middle: the learning screen. The dotted line indicates where the next stroke should be drawn, with the biggest dot indicating the beginning of the stroke. Right: the writing test screen. The second line of text displays how many strokes the target kanji has and how many the participants already drew. The "listen" button in the treatment condition would play the kanji's melody. The button is absent in the control version of the app. . 76
4.5 The scores for meaning retention (calculated by subtraction of post-test meaning scores from the longitudinal test scores) plotted against the time that passed between the two tests. The linear fit function shows no negative correlation between the length of time and the changes in scores. 78
4.6 The 15 characters selected for this study. All of the kanji consist of some combination of the six common radicals (highlighted). 78
4.7 Two of the post-test images for the kanji "foreign." For scoring purposes, the color indicates the order in which the strokes were drawn, while the large initial dot indicates the direction of the stroke. The participant on the left drew the strokes in the correct order, but the first stroke of the moon radical (purple) is drawn in the wrong direction. The participant on the right drew all the strokes in the correct direction (with the exception of the missing last stroke), but switched the order of the first two strokes in the "ten" radical. The participant on the left received 9/9 on order and 8/9 on direction. The participant on the right received 6/9 for order and 8/9 for direction. 80
4.8 Normalized scores for stroke knowledge of individual characters. Top to bottom: stroke direction post-test scores, stroke order post-test score, stroke direction longitudinal test score, stroke order longitudinal test score. The scores with statistically significant differences are marked with an asterisk. 82

5.1 *Zen Hanzi* screenshot. The player must drag the components from three rows at the bottom into one of the four structures above to assemble the hanzi indicated by the English translation and pinyin pronunciation at the top of the screen. 91
5.2 Some examples of look-alike components. Some differ by a single extra stroke, some by comparative length of two strokes, some by whether a stroke crosses another. 92
5.3 In the learn mode, the player is shown the hanzi before they attempt to assemble it. 92

5.4 The player receives feedback on their attempt, including whether they used the correct structure, how many components are correct, and whether any of the components were in the correct position. The game helps the player remember which options they tried by coloring incorrect components and structures red, and correct ones green. Once the correct structure is identified, the game also marks which components were incorrectly tried in each slot. 94
5.5 An example of one hanzi encoding 96
5.6 Component variations with differences highlighted in red. The 'water' component (A) is written with a vertical line at the top in some of the hanzi (B). It's still considered the same component according to databases I examined, but it would be incorrect to omit that line or add it where it isn't needed. Similarly, the 'great' component (C) is occasionally written in a slanted form (D). In *Zen Hanzi*, I include both versions of each component to draw the player's attention to the possible mistakes they could make. 97
5.7 Two entries from my component database. The principal component is on the left. The name is composed of numbers in Fordham's list, Chinese names for specific strokes, horizontal or vertical division of the component, and a suffix 'r,' indicating the component is reversed left to right. On the right are the names of other components within the database, that can be considered look-alikes for this component. 98
5.8 The conventional way of hanzi structure classification would identify this hanzi as a vertically-split two-part hanzi. However, each of those hanzi parts consist of two basic components, each carrying its own potential for a mistake. It's also worth noting that division into two parts in this case results in the transformation of the lower-left component into a different variant of itself, which may confuse the student about the proper shape within the original hanzi. Because of this, *Zen Hanzi* breaks this hanzi into four parts. 99
5.9 Quizlet screenshots. On the top is a typical two-sided *Quizlet* flashcard from a Chinese 1 course. On the lower left is *Quizlet*'s timed matching game, on the right is their Asteroid game, where entering the corresponding option destroys the asteroids speeding towards Earth. 101
5.10 *Qualtrics* pre- and post-test screenshots. On the left is the hanzi meaning test, on the right is the individual hanzi components test. 103

6.1 Gachapon: modern day toy capsule vending machines outside a game arcade in the Namba entertainment district of Osaka, Japan. Some are featuring a set of Harry Potter-themed necklaces [46]. 115
6.2 Screenshot of a typical quest list in World of Warcraft. 121

List of Tables

4.1 IEQ mean scores, standard deviations, significant differences between the two conditions, and the effect size, which is in the medium to large range for significant differences. 72
4.2 Descriptive statistics of learning outcomes (meaning tests, stroke order tests, and in-game scores). 73
4.3 Descriptive statistics of meaning tests. 79
4.4 Descriptive statistics of combined stroke knowledge scores. 80
4.5 Descriptive statistics of post-test stroke knowledge scores. 88
4.6 Descriptive statistics of longitudinal post-test stroke knowledge scores. . 89

5.1 Questions Used in Experience Analysis 105
5.2 Hanzi Test Scores and significance values from ANCOVA with pre-test scores as covariates 106
5.3 Hanzi Test Scores and significance values from ANCOVA with pre-test scores as covariates 106
5.4 Experience Evaluation results and significance values from ANCOVA with pre-test scores as covariates 106
5.5 Experience Evaluation results and significance values from ANCOVA with pre-test scores as covariates 107

Abstract

Evidence-based gamification and mnemonics for logographic writing systems.

by

Oleksandra G. Keehl

People have many reasons to want to learn a foreign language: some want to enjoy foreign media in its original form, some need it for business or travel, personal enrichment, and so on. Roughly 1/6 of Earth's population's learn languages with logographic writing systems (LWS) such as Chinese, Japanese and Korean as their native tongues. These peoples harbor an enormous cultural, economic and academic wealth. As a result, these languages are a popular choice for a variety of learners ranging from serious businessmen to K-pop[1] and anime fans. Whatever the motivation, learning additional languages carries a slew of cognitive benefits and should be encouraged.

Maintaining the motivation and persistence needed to learn a language is a difficult task, and the problem is amplified for learners with alphabetic roots tackling an LWS language. One needs to memorize around 2000 characters to be considered literate. This is routinely done through rote memorization, and the prospect discourages many would-be learners. When it comes to LWS as a foreign language, there is a divide between conversational fluency and literacy unlike that in any alphabetic language. My research aims to help bridge this gap through development and evaluation of LWS

[1]K-, C- and J- prefixes refer to Korean, Chinese and Japanese origin of the media in question. K-pop refers to Korean pop music, C-drama stands for Chinese TV drama, J-rock is Japanese rock music.

learning games. In this, I make a point to rely on insights from LWS teaching researchers, best practices of educational game design, and advice from the videogame industry experts.

My approach brings innovation in two areas:

I contribute to LWS instruction by introducing two LWS-learning game designs that focus on production tasks, while the vast majority of games out today target recall only.

I contribute to educational game design field by exploring the use of music as a mnemonic for learning, and conceptualizing the potential future use of notorious commercial game engagement and retention mechanics in an educational context.

I dedicate this dissertation to my grandfather, Roman Urbanski

and to LWS learners everywhere

Acknowledgments

I wish to extend my gratitude to my committee members for their time and guidance; to my advisers, Dr. Adam Smith and Dr. Eddie Melcer, who advised, advocated, encouraged and helped secure funding; to my labmates from the Design Reasoning Lab (DRL) and the Alternative Learning Technologies and Games Lab (ALT Games Lab) for being such an inclusive, inspiring and encouraging community; to the Computational Media department for covering conference attendance; to the UCSC Japanese language department for allowing me to play-test Radical Tunes with their students; to the UCSC Disability Resource Center for helping me succeed when I struggled the most; to the UCSC military retiree association for awarding me the Bruce Lane memorial scholarship three years in a row; to all UCSC undergrad and Masters' students who directly or indirectly assisted me in my work; to collaborators who collaborated; to friends and family who supported and believed in me; to pets who were there through sleepless nights; to my 700+ FocuseMate mates who were there to keep me accountable during the pandemic. I'm probably forgetting someone. I apologize. I promise I am not taking any of you for granted for a moment. A special thank you to Nicholas P.W. Lovell who not only kept me fed and housed throughout most of my PhD, but also assisted with countless technical issues I encountered in my course and research work.

Chapter 1

Introduction

Millions of people worldwide are taking up languages with logographic writing systems (LWS), such as Chinese or Japanese, as their second languages [97, 101]. However, one of the aspects that presents a unique challenge to the students coming from alphabetic backgrounds is learning to read and write the logographic characters [164]. The United States Foreign Service Institute classifies Chinese, Japanese and Korean among only four of the "super hard" category IV languages, describing them as exceptionally difficult for native English speakers [194]. They suggest 88 weeks or 2200 class hours for working proficiency in these languages. For comparison, category III languages (such as Farsi and Russian) require half as much time. When it comes to LWS as a foreign language, there is a divide between conversational fluency and functional literacy unlike that in any alphabetic language [163]. In recent years and with some success, many aspiring polyglots are turning to mobile apps to help them overcome this hurdle [27, 133]. The advantages of using such tools are clear: digital media allows for instant feedback, gamification, inclusion of engaging graphics, audio, video, and even finger-writing practice anytime and anywhere [27, 110, 111, 153, 29].

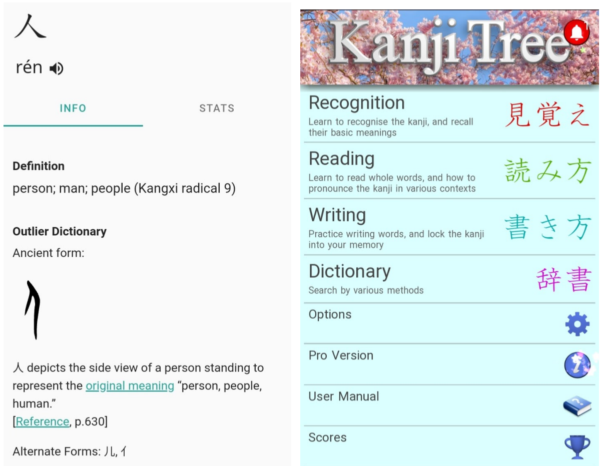


Figure 1.1: Screenshots of *Skritter* [176] (left) and *Japanese Kanji Tree* [179] (right), two popular feature-rich apps for learning hanzi and kanji respectively.

Surprisingly, despite the potential and wide range of game-based learning approaches, hanzi learning apps (not to be confused with Chinese language learning apps) fall largely into two categories of design [127]. First, purely educational apps with Spaced Repetition System-enabled (SRS) flashcards, stroke order animations, dictionary examples and so on (see Figure 1.1 for examples). SRS stems from the phenomenon that training sessions spaced with irregular intervals can lead to enhanced learning, in comparison to consecutive sessions, or sessions with very short intervals [178]. Second, games-based approaches with almost exclusively exogenous designs [170, 141] that consist primarily of multiple-choice questions tacked onto another game genre such as a fighting game (see Figure 1.2 for examples). I believe that the second category could be improved upon if one were to take into account lessons learned from Chinese language teaching experts and educational game researchers. For instance, studies show that component awareness has a positive impact on a student's ability to memorize characters [60], however, to my knowledge, none of the commercially available games (with a few exceptions covering only introductory characters) make use of that fact. Similarly, educational game research suggests that endogenous integration of educational content into a game results in better learner engagement [77, 200, 153, 154], however, in most hanzi learning games today, the learning content is almost entirely separable from the gameplay (e.g., it could be substituted for math questions without changing the gameplay or story in any way).

1.1 Core Concepts

This section serves to provide brief definitions of frequently used terms. These will be further discussed in depth throughout the dissertation.

Logographic writing systems (LWS).

A logograph (or logogram) is a written character that can represent a whole word, or at least a morpheme. A morpheme is the smallest meaningful part in a language. A logographic writing system, then, is one that is based on use of logographs. Chinese and Japanese are two such languages.

Endogenous and exogenous design

When it comes to educational games, these terms refer to the manner in which

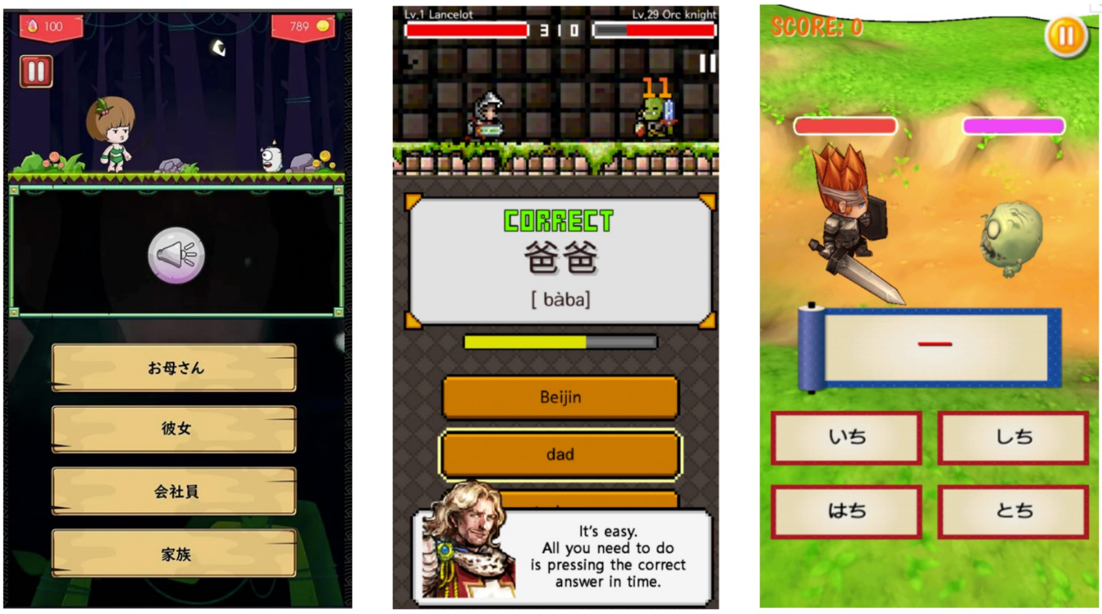


Figure 1.2: Screenshots of apps on Google Playstore listed as games for learning Japanese and Chinese. Left to right: Learn Japanese Yami [189], Chinese Dungeon: Learn C-Word [183], Kanji Battle [11].

the educational content is integrated with the game. With exogenous design, such content could be swapped for another subject without affecting the game story or mechanics in any way. Endogenous design implies the educational content is inseparable from the game world and/or gameplay.

Black and white hat gamification

Just like black and white hat hacking, black here represent gamification strategies that are malicious, harmful or exploitative toward the end user, while white hat strategies are, at worst, neutral.

1.2 Research Questions

In the projects described in this dissertation, I address the following research questions:

- RQ1 **task definition**: Identifying the disconnect between current research on effective LWS learning approaches, educational game design, and the commercially available LWS learning technologies.
- RQ2 **operationalization**: Design, development and evaluation of several LWS learning apps employing the missing elements identified in RQ1.
- RQ3 **scalability**: There are thousands of characters in LWS with complex subsystems of structure and components within those characters. How can the authoring burden be minimized to allow for inclusion of more characters in LWS learning software?
- RQ4 **Future work** : Application of black hat motivational techniques to educational games in general and LWS in particular.

1.3 Relevance and Contributions

My approach brings innovation in two areas: I contribute to LWS acquisition by introducing two LWS-learning game designs that focus on production tasks, while the vast majority of games out today target recall only. I contribute to the educational game design field by exploration of melody-based mnemonics, and conceptualizing the use of notorious commercial game engagement/retention mechanics in an educational context.

Digital gamification for knowledge acquisition is highly relevant in today's world, where access to devices is nearly ubiquitous, and opportunities for entertainment and instant gratifications are nearly limitless. Commercially, this concept has largely been applied either through highly exogenous gamification elements, or through highly development-resource-heavy narrow-focused games aiming to cover a limited number of lessons.

My research focuses on finding ways to aid a learner on their path of gaining and retaining knowledge by making their learning experience both enjoyable and efficient, and at the same time not creating an untenable burden on the game developer teams. I do this by focusing on endogenous integration of learning material, devising scalable content expansion schemes, and exploring mnemonic aids that can be integrated into game mechanics.

As previously mentioned, acquiring literacy is a long-standing challenge for students with an alphabetic background attempting to learn languages with LWS, such as Chinese and Japanese. My research focuses on addressing this gap by developing and evaluating scalable character-learning games with novel mnemonics and game mechanics.

My contributions to SLA include an exploration of mnemonic efficacy of melody for stroke knowledge acquisition, and a scalable, component-focused game design shown to drastically improve learner's hanzi production ability, at least in the short term.

1.4 Methodology, Goals and Principles

My methodology is primarily research through design. I've designed and developed several games and evaluated them through comparison studies, comparing either different versions of the games, or the closest commercially available products.

My goal is to explore alternatives to the currently available apps and games for learning hanzi and kanji. In this I make a point to rely on insights from LWS teaching researchers, best practices of educational game design, and advice from the videogame industry experts.

Based on my research on mobile minigame-like games and educational games, I selected the following principles to follow in my design:

1. Intrinsic pleasure is a must [131]. This can be aided by appealing graphics and sounds, and plentiful feedback to player's actions with visual and audio effects. Think juicy design [103] and Candy Crush Saga [114].

2. The game must have short rounds in order to fit into micro-leisure intervals of contemporary life [122].

3. The learning content must be endogenously integrated into the game [77]. That is to say, no Math Blaster-like "add two numbers to shoot at asteroid" [12]. The game must be inseparable from the learning content.

4. The game must avoid orthogonal mechanics [193]. Orthogonal mechanics are those that detract from the primary purpose of a serious game, in this case - learning.

An example would be requiring the player to become good at aiming and shooting in order to access the learning content. Every action should serve to reinforce or increase knowledge.

5. The game must be scalable to accommodate hundreds of kanji without hand-authoring separate levels for each group of characters.

1.5 Dissertation Overview

The background section will include the current state of LWS learning practices, followed by a discussion on motivation and its role in creating an educational game. Next, I offer a closer look at challenges and advances in educational game technology, followed by some insights on lessons learned from mobile games, and how they could be used in educational games.

Chapter 3 is an in-depth exploration of two popular non-educational multiplayer mobile games with a feature-by-feature discussion of how the engagement mechanisms of those games could be applied to an educational game.

Chapter 4 details my exploration of using melody as a mnemonic device in a musical kanji-learning game.

Chapter 5 describes a hanzi-learning game I developed using a component focused approach.

In chapter 6 I share my vision for use of black hat motivational devices in educational games.

In chapter 7 I summarize my work and offer possible directions for future research.

While portions of the work presented here were collaborative, I'll be using the singular first person "I" for the purpose of this dissertation.

Chapter 2

Background

This section begins with an overview of the unique challenges in mastering LWS faced by learners with background in alphabetic languages. Rather than an exhaustive explanation of the learning process and grammatological peculiarities of LWS, this subsection is meant to highlight the problem I intend to address with my research, as well as introduce the relevant linguistic and language acquisition vocabulary.

Next follows an examination of how motivation is talked about in the context of college education; non-educational games targeting older audiences; and the gamification field.

This is followed by a section on effective educational game design, highlighting the principles I intend to follow in my own research and development of the educational minigames.

Since my approach to tackling the LWS learning challenges centers on harnessing the engagement, accessibility and versatility affordances of mobile games, in the final subsection I dive into psychology behind mobile game appeal. This subsection too is meant to introduce relevant vocabulary, as well as highlight specific features employed in mobile games that have potential learning game applications but have not yet been employed in educational software to my knowledge.

2.1 Challenges and methods of learning logographic characters

Most of the modern-day logographic scripts came from the Chinese hanzi, which have been adopted into Japanese and Korean written scripts. In my research I primarily focus on these characters in the context of Chinese and Japanese foreign language acquisition (FLA). Due to common origin, similarities in structure, and, in fact, a significant overlap (about 60% of the 2000+ most commonly used Chinese and Japanese characters were the same or similar as of 2012 [204]) between Japanese kanji and Chinese hanzi, it is fairly safe to say that challenges and methods for learning to write one group of these characters apply to the other as well. This is why I draw from literature on learning to write in both Japanese and Chinese languages when motivating my research. Notably, learning the pronunciation and meaning of the characters is quite different between the languages and is not explicitly addressed in my research.

The LWS, such as Chinese and Japanese, are notoriously difficult to learn. The traditional and most widely used way of learning these characters is rote memorization, requiring students to write the characters over and over [125]. As a result, acquiring

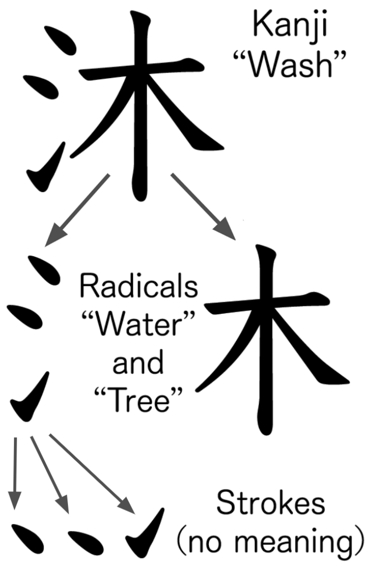


Figure 2.1: The anatomy of kanji. A kanji can consist of one or more radicals, which in turn consist of one or more strokes. While radicals can carry a meaning of their own (which may or may not relate to the meaning of the kanji), strokes are purely graphical elements.

hundreds of characters through rote learning is a daunting task and can discourage a casual learner from pursuing a language. Games have become a frequently utilized tool to aid in FLA, addressing topics such as vocabulary and meaning [186, 41], tones [80], conversational phrases [35], and culture [21]. However, there has been relatively little focus on how FLA games can be employed to teach writing or other production skills for logographic characters. For instance, while there are many educational games and applications out there intended to help people memorize Japanese kanji, most focus solely on the reading and meaning portions and don’t teach stroke order (such as the popular service WaniKani[1]). Furthermore, the majority of commercial writing apps primarily amount to digital flashcards. While convenient and useful to a motivated learner, they offer little extrinsic motivation to a casual student, which is a significant hurdle for many kanji students [163].

The importance of learning the correct stroke order is debated by some, since some of the important practical application of that knowledge, such as looking up a character in a dictionary or being able to produce characters by hand, have lost much of their value with the advent of romaji/pinyin (the romanized phonetic scripts for Japanese and Chinese respectively) digital input methods, which allow users to type the character's pronunciation and then select the appropriate character from the list of those that match that pronunciation [203]. However, stroke knowledge has been linked with successful writing skill development, especially at the early stages of mastering an LWS [130]. Incorrect stroke order has also been linked with errors in hanzi production [124]. The stroke order knowledge is also instrumental for reading and writing handwritten scripts.

A survey of strategies of successful students of Japanese kanji revealed that they employ a number of techniques, one of the prominent ones being component analysis, that is breaking up the characters into their components (see Fig. 2.1), or graphemes, for further analysis and memorization. Dr. Heath Rose [164], a researcher who extensively studied the struggles of mastering writing of Japanese kanji for students from alphabetic background, suggests that developing graphemic awareness at the beginning

[1] https://www.wanikani.com

of a student's kanji learning quest can aid them in developing their writing skills.

In recent years, mobile apps for learning kanji and hanzi have gained popularity. Apple's App Store and Google Play Store list hundreds of apps and games developed for that purpose[2], some with millions of downloads. A study on use of multimedia (including but not limited to apps) for learning kanji, showed that such methods were effective when used strategically (e.g. taking advantage of Spaced Repetition System[3] (SRS) included in many flashcard apps), and were also more enjoyable than the pen and paper flashcards and writing practice [125]. The vast majority of the apps and games focus on flashcard-style recall task, with few offering writing practice and none, to my knowledge, offering mnemonics for remembering the stroke order and direction. There is also little to be found in terms of graphemic awareness development. One exception to the latter is WaniKani, which makes an emphasis on graphemic awareness by assigning unique and often comical meanings to each radical (see fig. 2.2). The service then uses these meanings to come up with short mnemonic-laden stories for each kanji based on its radical composition. However, many of the mnemonics are stretched and abstract, adding a layer of information for a student to memorise which has little to do with the kanji. Rose warns against using such methods, as sometimes a student might remember the story, but not the kanji itself [163].

My research focuses on aiding students in developing graphemic awareness through utilizing novel mnemonic techniques, and providing motivation through use of

[2]estimates obtained by searching for "kanji learning" and "hanzi learning" in the stores' web interface and counting the rows of results

[3]SRS stems from the phenomenon that training sessions spaced with irregular intervals can lead to enhance learning, in comparison to consecutive sessions, or sessions with very short intervals[178].

Figure 2.2: WaniKani's 'poop' radical is a great example of shock-value driven mnemonics.

motivational devices found in modern F2P games. My target population are novice and intermediate learners of Chinese and Japanese languages, who are already familiar with such basics as phonetic scripts of the languages and have learned at least a handful of the simplest and most frequently used characters. I target this audience because the vast majority of kanji/hanzi learning games and apps target absolute beginners, and because the beginner-appropriate material is fairly limited and has already been covered in many creative ways. At the same time, novice and intermediate learners have hundreds of characters to master and could benefit from a scalable system that could provide motivation and mnemonics to aid in this momentous task.

2.2 Motivation

Motivation is a complex matter and has long been discussed in works of philosophy and psychology. Before I dive into the intricacies of how educational games could be used to motivate students to learn, it is worthwhile to examine how the subject of motivation has been addressed in the two relevant fields of interest: education (particularly, adult education) and games (particularly, mobile F2P games). Motivation in educational games is addressed within the Recent Commercial Examples part of the Digital Educational Games background section.

I will also briefly discuss motivation's dark cousin when it comes to games: addiction.

2.2.1 Motivation in Education

There exist many excellent, highly cited literature reviews on motivation in education. In this section I briefly touch on several of them and share the authors' conclusions.

In an essay titled "What Theories of Motivation Say About Why Learners Learn," the authors do an excellent job of summarizing the practical implications of contemporary theories of motivation to education of college students [39]. They define motivation as the force that initiates and sustains behavior. For education in particular, it is what causes students to value learning and promotes active engagement in classroom activities and homework assignments. According to the Self-Determination Theory (SDT) [39], there are two basic types of motivation, intrinsic and extrinsic. Intrinsic motivation refers to actions taken because they are enjoyable or interesting in and of themselves. Extrinsic motivation refers to actions taken in order to achieve or avoid a certain outcome. While intrinsic motivation is believed to lead to high quality of learning and creativity, caution must be exercised when resorting to extrinsically motivating students [168].

In their discussion of motivation as a function of needs and expectations, McMillan and Forsyth are talking of extrinsic motivation [137]. In this context, needs means that a student perceives a skill to be learned as desirable or even necessary. Expectations here refers to the student's belief about their ability to attain that skill. In essence, if the student finds a lesson useless (i.e. "I'm an engineer, why would I ever need biology"), they are unlikely to apply much effort into learning it. On the other hand, even if they find a skill useful, if they don't believe they can master it (i.e. "My brain just isn't wired for math. I could never get an A in this course."), they are liable to give up early, negating their chance at success.

The authors further identify several kinds of value a person might ascribe to an action (such as a student choosing to participate in a learning activity). Among them are the intrinsic value, described as the measure of how enjoyable or interesting the task is in itself; and the utility value, a measure to which attaining one's long-term goals depends on the success in the given task. These two concepts translate very well to the SDT's intrinsic and extrinsic motivation.

Other factors related to student motivation have been identified over the years. Competence (or the pleasure one derives from feeling competent) is considered one of the core psychological drives. Dweck and Leggett demonstrated that students focused on learning goals, i.e. increasing their competence, are more likely to select challenging tasks than the students focused on performance goals, i.e. good grades or other extrinsic rewards [54]. With this in mind, it may be prudent to search for ways to shape the reward and progression structure of the game in a way that maximizes a student's feeling of newly developed competence, rather than relying heavily on badges and points. It is also of note that student motivation tends to be higher when students succeed with reasonable rather than with high effort, as it leads them to see themselves as capable (spending sleepless nights studying to pass a test is not a reasonable effort).

It was found that meaningful, reachable goals of moderate difficulty tend to be most motivating [38]. The challenge then lies in designing tasks which would be perceived by students as moderately difficult, and helping students understand how the material is relevant to them, and therefore meaningful. When it comes to learning a foreign language, students can have different motivations. Some might want essential survival phrases like "where is the bathroom" for their upcoming vacation. Others might want to learn enough every day words and phrases to have simple conversations with their foreigner in-laws. Some need to learn specific terminology and the formal ways of address for their work (for example, an ER nurse may wish to learn phrases like "where does it hurt" in several languages). Yet others may wish to learn the language in order to consume foreign media in its original form. Addressing such a variety of goals in a single course is impossible, as would it be in any linear educational structure. Games, however, have the advantage of the option of not being linear. It may be worth exploring designing content in more or less self-contained modules and thus allowing students to pick and choose the most personally relevant sections.

To sum it up, the recent consensus on keeping students engaged and motivated relies on students choosing meaningful and moderately challenging tasks that are relevant to their long-range goals. The intrinsic enjoyment a task provides is another contributing factor. As game makers, it is up to us to design moderately challenging learning tasks that can be adapted to a range of goals, and to make those tasks intrinsically enjoyable. A good way to explore how to accomplish this is to examine the strategies non-educational game makers employ for capturing their audiences. Another perspective worth exploring is that of gamification experts, as getting people to persevere in using a product or a service is often one of their primary goals.

2.2.2 Motivation in non-Educational Digital Games and Gamification

To provide a more authentic representation of the subject of motivation in these two fields in the way it is seen by its natives (game designers and gamification experts), I chose to predominantly focus on the relevant non-academic writing as sources for this section, such as industry professional blogs, magazine articles and the like. These tend to have more detail on practical implementation and application of gamification, and of motivational elements in games. The main goal of this section is to establish a vocabulary for the subsequent case studies.

So what is gamification? In the decade since gamification became a mainstream concept, different definitions for it have emerged, some more flattering than others.

In their 2012 paper titled "Defining Gamification [...]" Huatori and Hamari described it as "a process of enhancing a service with affordances for gameful experiences in order to support [the] user's overall value creation."[92]. This is not surprising, as gamification emerged primarily as a marketing or business boosting tool, aiming to motivate employees and to incentivise customers to spend more money.

A more contemporary view of the concept is reflected in a definition offered by Yu-kai Chou, a well-known specialist on gamification: “Gamification is the craft of deriving all the fun and engaging elements found in games and applying them to real-world or productive activities”[25].

Game makers and gamification gurus ultimately have different goals and different approaches to designing their user experience, however, there is a significant overlap between the methods employed by the two fields.

Yu-kai Chou’s Octalysis framework for gamification offers a convenient structure for discussing player/user engagement methods (Fig. 2.3). Of particular interest to me is his division of the methods into white and black hat gamification. He describes **white hat techniques** as follows:

> *“If something is engaging because it lets you express your creativity, makes you feel successful through skill mastery, and gives you a higher sense of meaning, it makes users feel very good and powerful.”*

Among these, he lists “Epic meaning and calling,” which refers to making the user feel as though they are a contributing part of a greater whole; “Development and accomplishment,” referring to mechanisms for making users feel as though they are honing their skills and achieving new heights, the elusive flow[4] falls into this category; and the self-descriptive “Empowerment of creativity and feedback.”

[4]Flow is a concept coined by Mihaly Csikszentmihalyi in the 1970’s [33]. Some of the Flow characteristics include extreme focus on the task, sense of control, merging of action and awareness, loss of self-awareness, distortion of the experience of time and the experience of the task being the only justification needed for continuing it.

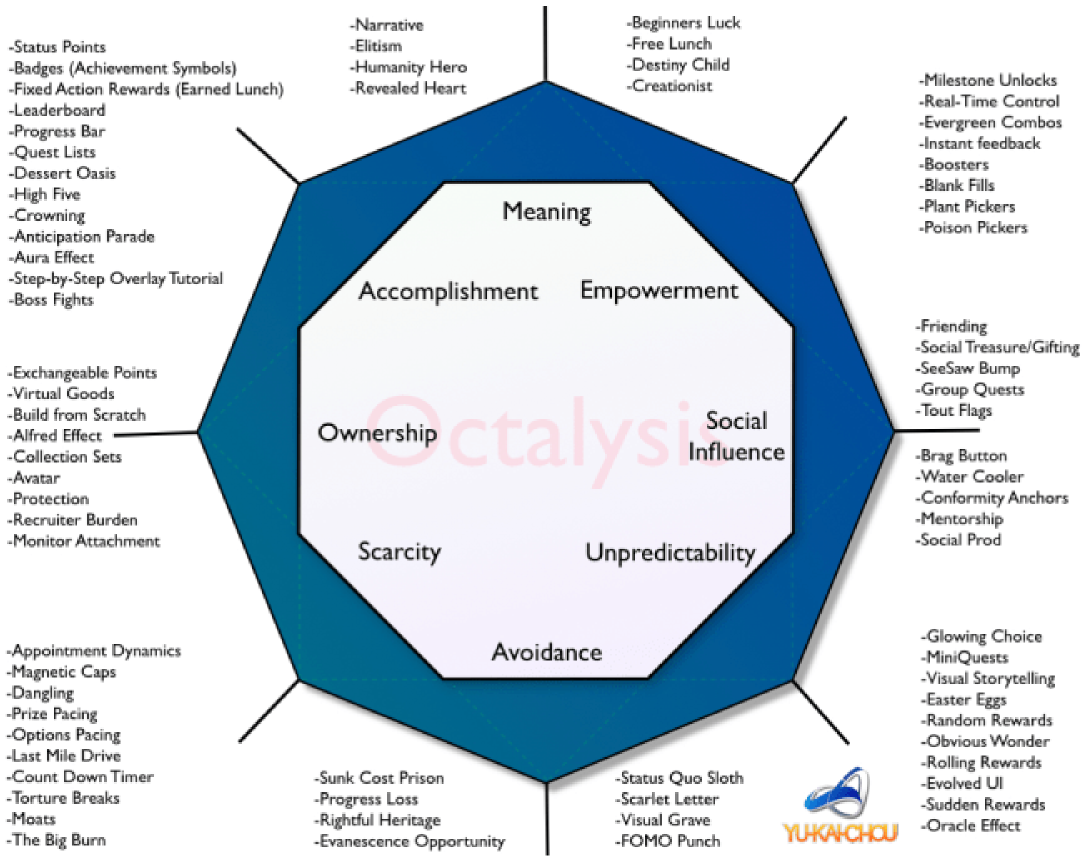


Figure 2.3: The Octalysis framework for classification of user/player engagement mechanics, developed by YuKai Chou.

Of **black hat techniques** Chou says:

> *"If you are always doing something because you don't know what will happen next, you are constantly in fear of losing something, or because there are things you can't have, even though you would still be extremely motivated to take the actions, it can often leave a bad taste in your mouth."*

Among the black hat techniques he lists "Scarcity and impatience," in which players are made to endure artificially imposed time intervals before they can, for instance, open a treasure chest, or have to grind to obtain a rare item; "Unpredictability and curiosity," which includes gambling-like gacha mechanics in addition to the more or less harmless desire to know what happens next in a story; "Loss and avoidance," which makes use of Fear of Missing Out[5] (FOMO) (for instance, a chance to obtain limited time content) in addition to fear of losing investment of time or money (such as breaking a daily streak and not being able to collect the grand prize). Chou also describes two categories which can swing both ways. One is "Social influence and relatedness," which can play into the feeling of belonging to something bigger, but can also make use of peer pressure. The second is "Ownership and possession," which plays on human drive to collect things and accumulate wealth, but can be used to manipulate players through FOMO and fear of loss.

While some of these technique categories have been used frequently in educational games and apps (think badges, points, daily streaks and leader boards), sometimes to great success, others have not (I have yet to see gacha in an educational game). This may at least partially be due to the notoriety of some of the techniques, particularly in the black hat group: the techniques purportedly designed to make players part with their money and time better spent sleeping or studying. Allow me to discuss the nefarious nature of some of these techniques and why I believe their evil powers could be harnessed for the good of the students.

In his Gamasutra article on F2P game design, N. Lovell, an author and game industry consultant (among other things), identified three distinct components of a

[5]The term FOMO was coined in a 2004 op-ed in relation to college students fear of missing out on social activities [135]. The term has since been applied more broadly to nearly anything one might fear to miss out on.

game which require separate strategies [131]:

- **The Core Loop** is the basic activity the player performs over and over. This is where the intrinsic pleasure is a must. Examples include a single round of Candy Crush Saga [114], or a single match in League of Legends [71]. In terms of an educational game, core loop aligns well with the learning and reviewing activity - something that a language student would have to do repeatedly and that's liable to become boring and repetitive in the long run. Since this is where the learning would take place, special care must be given to designing both an academically sound and intrinsically pleasant experience. Lovell states that while important, the core loop is usually insufficient to maintain long-term interest on its own.

- **The Retention Game** is the structure that entices the player to keep coming back to the game. Retention mechanics include scores, progression, narrative, achievements and competition. This is the category that almost perfectly overlaps with gamification, as retention tends to be one of its main goals. While many of the retentive elements have been employed in educational games and software over the years, I argue that there are some that have not yet been explored in this context, and some that have not been employed to their full potential.

- **The Superfan Game** is also known as the End Game, where the most passionate players and big spenders (a.k.a. whales) flourish. This can be facilitated with player vs. player battles, clans and competitions. According to Lovell, this category often develops on its own as the result of skillfully executed retention game, and one shouldn't focus too much on it until the first two categories are covered. Following Lovell's advice, I will primarily be focusing on the core loop and retention components in my research.

The concepts described by Chou and Lovell appear in many other game design and gamification blogs and articles, though at times with other names [113].

2.2.3 From Motivation to Addiction

Addiction to technology has long been a concern for parents and academics alike, be it TV, social media, the Internet in general, smartphones or video games. In the context of this work, I focus on the latter. After all, video games have unique ways to command a player's attention for excessive amounts of time.

There are two parties in any addiction story - the addict and the object of addiction. Psychology works on game addiction identify the following traits that make a person vulnerable to what might be considered unhealthy gaming behavior: loneliness, leisure boredom (a likely consequence of conflicting perceptions of having too much time available with too little to do [86]) and low self-control("the ability to override or change one's inner responses, as well as to interrupt undesired behavioral tendencies and refrain from acting on them" [188]). In today's hectic urban lifestyles, idle time is often fragmented, such as in long commutes and waiting in line for services. During these times, people may have a strong desire to alleviate boredom by engaging in entertaining and stimulating activities [122].

Glued to Games suggests shifting the focus from fun - an ill defined concept, to need satisfaction using the Player Experience of Need Satisfaction (PENS) model [173]. The authors highlight three specific intrinsic human needs: competence, autonomy, and relatedness. According to the authors, video games are compelling because of the immediacy, consistency, and density with which they can satisfy these needs. Unlike planning for a bowling activity with friends, video games are more easily and immediately available, the payoff for effort and mastery in games is reliable (unlike, say, promotion at work), and rewarding experiences are fairly frequent. Aside from need satisfaction, there are some game mechanics that have a higher potential to become exploitative toward the player and can target both the time and the wallet.

- **Zeigarnik Effects** (a.k.a unfinished business) - The book Glued To Games dedicated a chapter to "The Addictive Undertow of Games." The authors call out Zeigarnik Effects as one of the potentially addictive traits of Massive Multiplayer Online Role Playing Games (MMO RPG's). The players are usually pursuing several quests at the same time, some of those being chain quests. While players may try to tell themselves: "I'll stop as soon as I finish this one thing," there's always something else that will be on the verge of completion that will tempt them to continue playing (Fig. 6.2). This technique is commonly employed by F2P games, with multiple different timers counting down to rewards (some examples are Cat Game – The Cats Collector! [69], and Companion for Dying Light [190]).

- **Gacha** (a.k.a. Loot boxes) - This term is a shortened variant of Japanese gachapon (or gashapon), a name of special vending machines that gives the buyer a chance to get a randomly selected toy from a special set. The goal of purchasing these toys is simple - to collect all characters from your favourite anime, manga, cartoon or another special series of toys. In digital games, the term applies to any sort of reward mechanic that requires time and/or money, without guaranteeing a specific outcome. These could be loot boxes, roulettes, monster breeding, etc.

For a more thorough discussion of these mechanics, see the Discussion section at the end of this document.

I argue that these techniques could be used in educational games to promote extended play and learning. Chou warned that these kinds of black hat motivation devices can leave a player with a bad taste. However, I posit that the bitterness comes from the realization of the time and money wasted with little to show for it. In an educational game (with quality learning content), time spent in-game should correlate with knowledge gained, thus counterbalancing the stereotypical harmful effects the addictive game mechanics may cause.

2.3 Digital Educational Games

Learning through play is ubiquitous across time, cultures, and even species [115]. Play in general can help kids develop social and motor skills, and alphabet cubes and Legos assisted generations of kids in learning the ABC's [82] and spatial reasoning [74]. The advent of electronic computers brought with it an opportunity to create more targeted learning experiences for a multitude of subjects and difficulty levels.

A notable early example being PLATO, Programmed Logic for Automatic Teaching Operations [15]. It was distributed through thousands of graphic terminals worldwide and boasted courses from elementary to university level on a variety of subjects, from language to math, to chemistry and music. It included quizzes with free-text answers (implemented via keyword search) and custom feedback to wrong answers. Digital educational games were soon to follow. One of the first was a 1964 text-based resource management game for grade schoolers, called The Sumerian Game [93]. No list of educational games can be complete without the The Oregon Trail [138], or Where in the World is Carmen Sandiego? [16]. Countless other titles, successful and not, followed. This section discusses unique challenges associated with the development of educational games and an overview of existing educational games, focusing on those targeting an older audience and language acquisition, in theme with this dissertation's subject.

2.3.1 Unique challenges

This section discusses three challenges associated with educational game development:

- How to make the game fun and engaging without distracting the player from the game's main purpose - learning.
- How to effectively deliver learning material without making the game boring.
- Financing and monetization.

It takes time, effort and know-how to design and develop a compelling game.

It also takes time, effort and know-how to design and develop an effective educational curriculum. It shouldn't be a wonder that doing both at once, without sacrificing the quality of one or the other, is not a trivial task [65]. This implies two unique challenges for the educational games domain. First, regarding motivation, how should educational content of the game be made fun? Games that don't have a good answer to this question are often described as chocolate-covered broccoli [112]. Second, regarding effectiveness, how should the fun aspects of the game be implemented so that they don't undermine the intended educational impact of the game? Games that don't have a good answer can end up teaching student players less than can be learned by a non-playing, passive observer (possibly ascribed to cognitive load) [41].

In regard to the former question, Habgood et al. described an experiment of using, what they called, intrinsic and extrinsic integration of educational material into a game called Zombie Division [77]. The gameplay consisted of adventuring through a 3D maze and fighting skeletons along the way. In the intrinsically integrated version, skeletons had numbers on their chests, and players had to use appropriate divider-attacks to defeat them. In the extrinsically integrated version, all the same numerical problems were presented in an abstract form between the dungeon sessions. Their experiments showed that kids in the intrinsic group showed higher results on the post-test, and that given a choice, kids preferred the intrinsic version to the extrinsic one.

Kathleen Tuite addresses the latter question in her paper "GWAPs: Games with a problem" [193]. GWAP usually stands for a Game With A Purpose, of which educational games are just one kind. Tuite defined a concept of orthogonal game mechanics, which refers to mechanics within a GWAP that distract or otherwise hinder the player's effort to fulfill the game's purpose. As an example, she describes Onto Galaxy [116]. The purpose of the game is to populate an ontology. The mechanics involve navigating a ship through a space filled with moving objects, identifying those that have the needed ontological properties and collecting them. In order to meaningfully and effectively contribute to the ontology, the player has to get good at the navigation part of the game, which is a skill that doesn't contribute to the game's purpose.

Tuite asserts that for a GWAP to be effective, the mechanics of the game should contribute to a player's effectiveness in fulfilling the game's primary goal, instead of attempting to disguise the task at hand or even detract from it by putting extraneous barriers in the player's path.

Financing and monetizing educational games has also been an ongoing challenge and is often blamed for the poor quality of many educational games. A logical target audience - the schools - rarely has budgets big and flexible enough to sponsor production of a high quality game [32]. While I don't have a solution for this challenge, I offer some ideas for scalability of LWS learning software, that should help lower the development effort and costs.

2.3.2 Recent Commercial examples

The ultimate goal of my work is to assist in creating educational games for adults with the synthesis performed through the lens of a language learning game. With that in mind, this section examines some of the prominent language learning games for

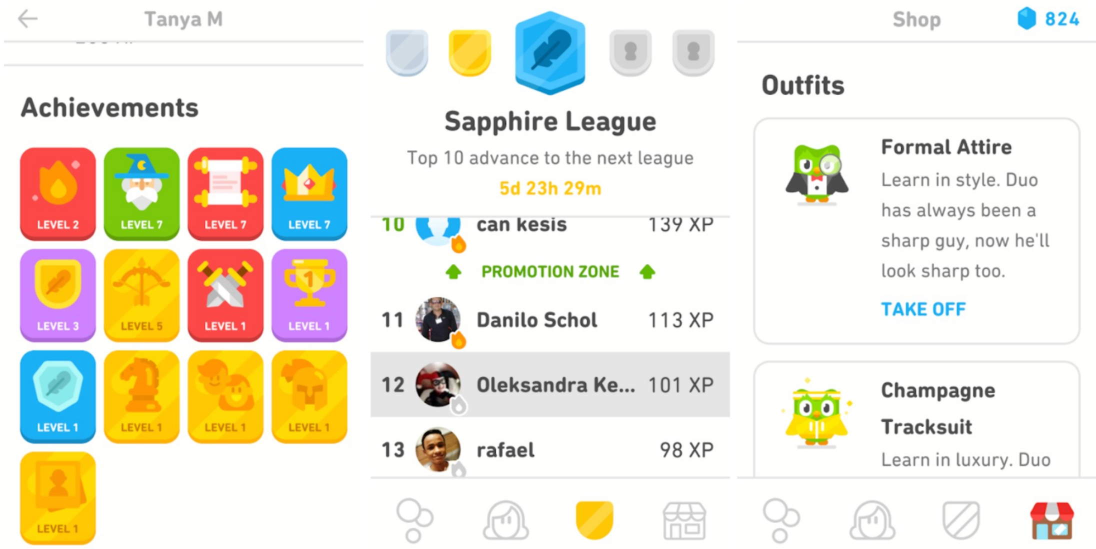


Figure 2.4: Duolingo Android app screenshots left to right: the achievements screen, the league screen, and avatar customization screen.

adults.

Duolingo is one of the most popular programs for second language acquisition, boasting over 300 million active users [52]. It has all the hallmarks of gamification: there are refillable hearts allowing one to continue despite mistakes, there's an in-game currency, there are unlockable customizations, there are level-up-able badges, daily-login streak-tracking, leaderboards and weekly league competitions (Fig. 2.4). One can readily compare their progress with their friends or random strangers who happen to be assigned to the same league. The material is broken into bite-sized lessons, which in turn are grouped into neat sections separated by proficiency tests - a clear way to give students a sense of their progress and accomplishment. The only thing lacking from making it a good game is… the gameplay. If Duolingo claimed to be a game, it would be yet another case of chocolate-covered broccoli, however, it does not, and thus is instead an example of gamification done right. Learning-material-wise, Duolingo offers both grammar and vocabulary, and, depending on the language, from short introductory courses to extensive ones, well into intermediate or even advanced levels.

The majority of language apps advertised as games on Google Play Store fall strictly into the chocolate-covered broccoli category, or what the language learning community likes to call "glorified flashcards" (Fig. 1.2). With a thin veneer of game-like visuals and a few of the most basic of gamification features (like customization and badges) to facilitate motivation, but lacking an engaging story (like Final Fantasy: Brave Exvius [95]) or entertaining mechanics (like Candy Crush Saga), these fall far short of what one may consider a game.

One interesting mobile example that went a few steps beyond "press correct answer to hit" is Chinese Spy [10]. The game's Google Play Store page describes it as embarking on a highly classified mission, which will take the player across China, detonating bombs, decoding encrypted messages, shooting targets and stopping assassins along the way, and culminate in saving both the Chinese president and the world.

For motivation, Chinese Spy employs what the creators of Zombie Division called intrinsic integration of learning material, making acquisition of vocabulary relevant to the game world, albeit at a superficial level (Fig. 2.5). The game leans into the Meaning axis of Chou's gamification Octalysis, enticing the player with saving the world.

The world of PC gaming offers another interesting example: Learn Japanese

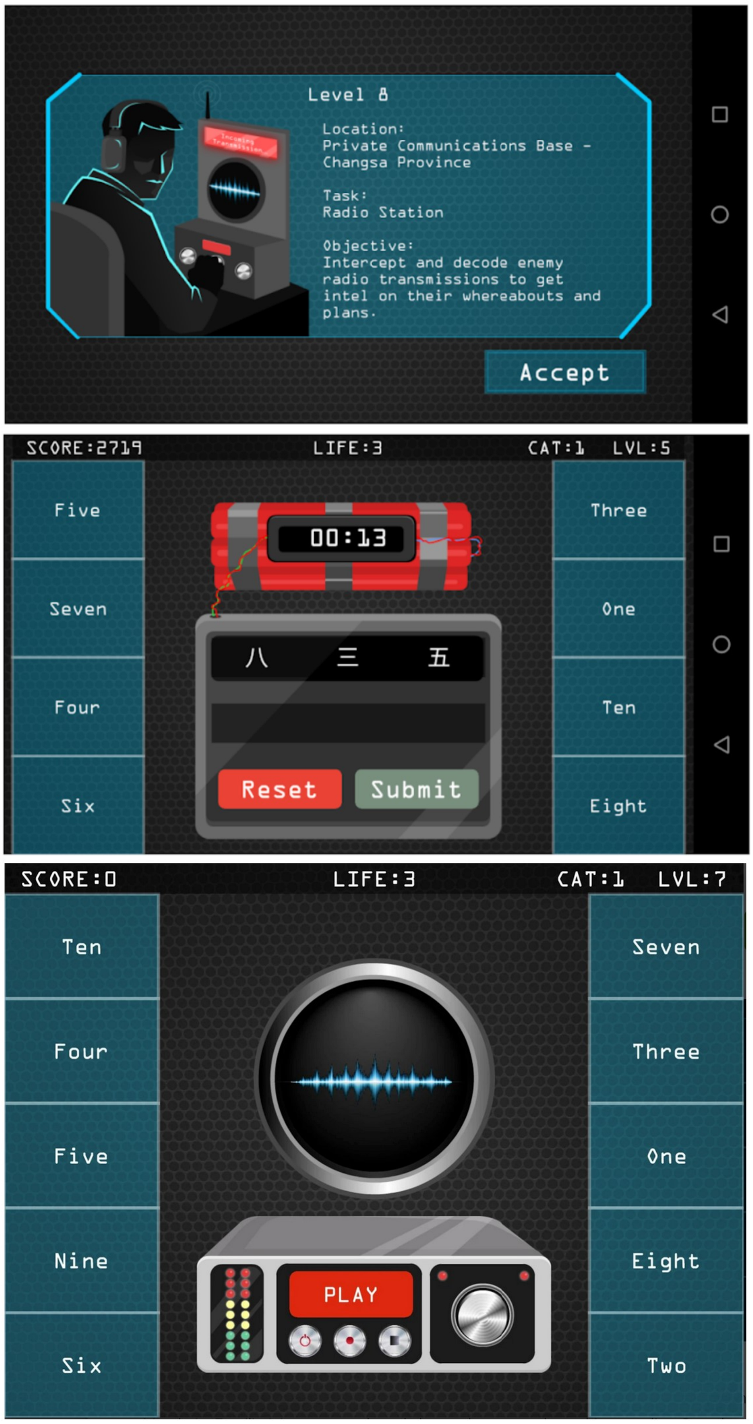


Figure 2.5: Screenshots of Chinese Spy. From top to bottom: in-game narrative, bomb-disarming exercise, decoding secret message.

to Survive (LJS) series [50]. The series consists of three titles which can be played independently: Hiragana Battle, Katakana War and Kanji Combat. The first two fully address the two Japanese phonetic writing systems, while the last one gives a sampling of about 100 basic kanji (one needs to know over 2000 to be considered literate and well over 10,000 exist in more advanced reading material). Each game is an RPG adventure thematically connected with Japan. It boasts professionally created assets and an engaging story (Fig. 2.6). Dozens of players on Steam shared their stories of enjoying the gameplay and successfully mastering the material.

Even though there are undisguised lessons between play sessions and the battle in the game is essentially a multiple-choice quiz, the players report that animation and sound effects make the process sufficiently entertaining. LJS doesn't attempt to spice-up the learning process with badges and gold stars, nor does it hide its primary purpose of teaching the player the basics of Japanese writing systems. Essentially, as recommended by the creators of Zombie Division, LJS developers successfully executed intrinsic integration of learning material into an otherwise entertaining game, and were thus able to keep their players engaged and motivated to learn. One drawback of a beautiful RPG as a learning material delivery method is its lack of scalability. While LJS is sufficient to give newbie Japanese language enthusiasts a taste for what's in store, adding new educational material would at the very least necessitate additional narrative and likely additional art and world development.

In conclusion, the current market is dominated by minimally gamified flashcards with rare exceptions of flashcards immersed in an elaborate story. Intrinsic in-

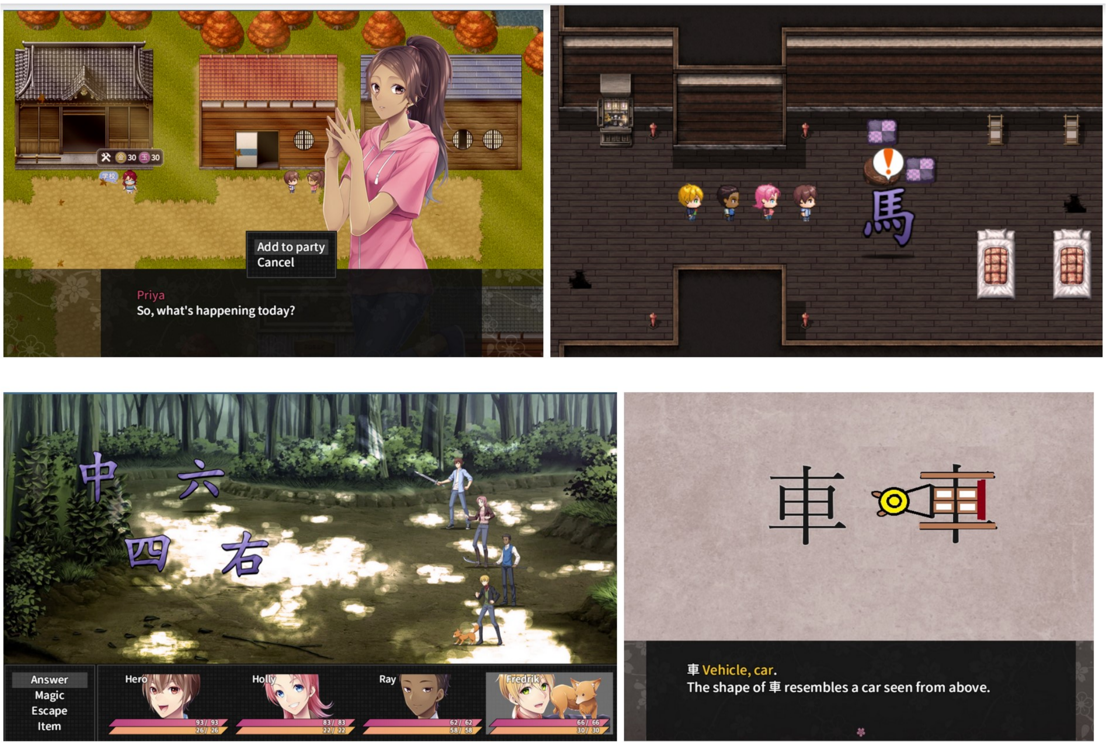


Figure 2.6: Learn Japanese to Survive: Kanji Wars screenshots.Screenshots of Learn Japanese to Survive. Upper left: the game's impressive character design; Upper right: the player encounters an enemy in an RPG setting; Lower left: choosing "Answer" during the battle results in the familiar mechanic of "select correct answer to hit."; Lower right: one of the lessons offered to the player at the beginning of each game day.

tegration of learning material is so far on a superficial level and orthogonal mechanics (such as the ability to quickly find a correct answer on a cluttered screen) are a common occurrence. As an academic passionate about games and language learning I see a lot of room for improvement.

2.3.3 Education Research via Game-Making

One necessary step of developing an engaging and educationally effective game (or better yet, developing a methodology for developing engaging and educationally effective games) is an evaluation of whether each feature produces the desired effect. Such research via game-making has been undertaken in the past and offers some valuable insights. If a game is made to explore a new technology in order to see said technology in action and test its feasibility and scalability, the game doesn't even have to aim for engagement or effectiveness [134].

Butler et al. conducted research on automated game progression generation in an educational puzzle game Refraction [17] (Fig. 2.7). Their paper focused on evaluating the engagement metrics of the generated level progression in relation to the original, human-authored version of the game. They deployed the game to several popular Flash game websites and randomly assigned either the original or generated version to each player, collecting play data from more than 2000 players. In order to assess engagement in these in-the-wild conditions, they examined the total length of play time and the number of levels completed. They used non-parametric tests to evaluate the not normally distributed data and draw conclusions. Although this evaluation does

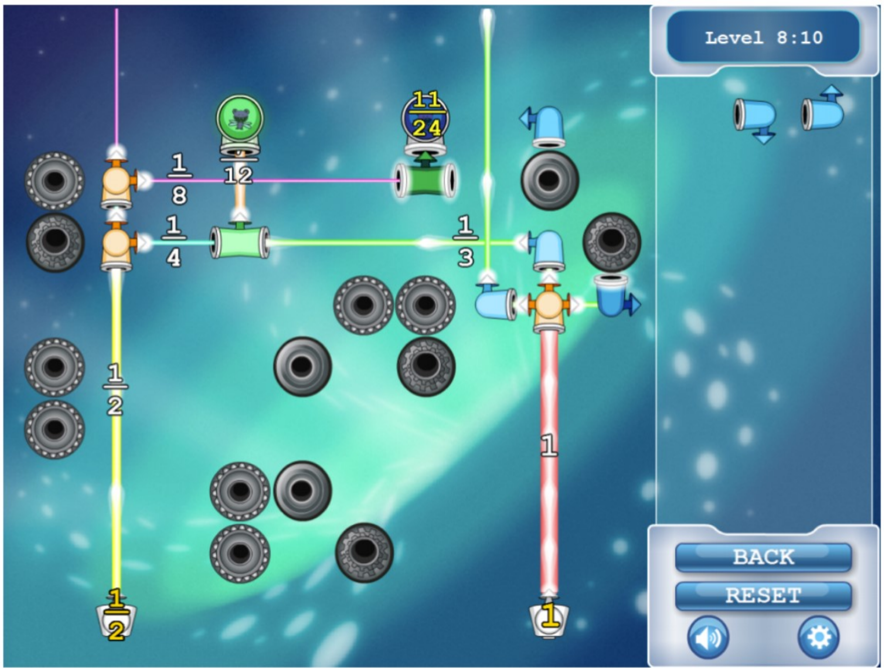


Figure 2.7: Screenshot of Refraction, an educational game for learning fractions.

not tell us whether the game was educationally effective, it tells us that the automatic progression design approach being evaluated indeed produced a valid progression of educational content, which was the goal of this specific project.

Denning et al. designed and evaluated an educational card game for raising computer security awareness [43]. They used multiple play-tests and show-and-tell sessions both with representatives of their target audience and other interested parties using the feedback to iterate on their design. Once the design was finalized, copies of the game were sent to 150 educators, some of whom provided feedback on their experience sharing the game with students both in and outside of the classroom. Fourteen of the responding teachers used the game in class. Many of them reported that the game improved student proficiency and interest in the subject; that they would use the game in class again to teach the topic; and two mentioned that they wouldn't have otherwise covered the material presented in the game. In addition to controlled user studies, getting such feedback from target audience in the wild is crucial to evaluating the game's educational potential and feasibility of use in an intended setting.

Melcer et al. explored the use of tangibles and collaboration in a programming game [140]. The researchers made two versions of the game, one with a point and click mouse interface and another with physical blocks that had to be connected together to execute a program. They ran a 2x2 study, where participants were randomly assigned into either the mouse or tangible group, and further into either individual or collaborative play. The experiment involved self-efficacy and engagement questionnaires, as well as objective scoring and timing of participants' performance on training and challenge levels of the game. With statistical analysis, the authors were able to show that tangibles resulted in higher engagement and enjoyment, as well as better in-game performance as compared to the mouse group.

These are only three examples of educational research via game making amidst many. It is a proven method to gain valuable insights into various aspects of games, such as evaluation of a game's educational and engagement potential, which is of particular interest to my own research.

2.4 Mobile Games

2.4.1 Unique Restrictions and Affordances

The subject of intricacies of mobile design has been discussed since the days of Personal Digital Assistants and early mobile phones [31, 51]. While some of the early concerns, such as processing power and highly limited network bandwidth, have since become less of an issue, many of the other factors remain.

A search for documents matching “challenges designing for mobile platform” returns a multitude of articles on the subject. The majority of these come in the form of industry blogs, official design guidelines from commercial giants like Apple, and some academic works that laid the foundation of what can now be considered common knowledge.

The summary of the common themes is presented below.

Designing for a mobile platform differs from designing for PC or consoles. The main limiting factors are the size of the screen and the number of available controls. The battery life is also something to be considered.

The size of the smartphone screen severely limits the number of UI elements that can be placed on the screen without obscuring the game or running into the “Fat Finger” problem[6], which warranted attention from mobile UI blogs [150], to official guidelines from major companies [44], to academic work on the subject [14].

[6] The term “Fat Finger Error” originated in financial trading, when an erroneous transaction occurred due to inaccurate button press. The term has been adapted to modern days and is widely used in discussions of mobile design.

The variety in screen dimension ratios of popular devices also contributes to the problem, as the developer needs to build flexibility into their UI if they wish to target a wider range of devices.

While on a PC, the designers have the whole keyboard and a mouse at their disposal, and gaming consoles boast on average over a dozen buttons; on a phone, a developer has to make do with touch, swipe, and hold [94]. These days though, accelerator, the gyroscope and cameras all allow for additional innovative input forms.

On the other hand, the any-place-any-time versatility of mobile phones, the nearly ubiquitous adoption by the general population, and the ease of distributing apps and games via Google's PlayStore and Apple's AppStore make smartphones an appealing target for developers. The latter is a double-edged sword. The vast amount of games available, many of which are free to play at least in some capacity, makes for a fierce competition over the player's attention. The retention data from Google Play shows that a game has at most several minutes for the new player to decide whether to delete the game they just installed [19]. Because of this, mobile developers need to be highly conscious of load times, learning curves and tutorial lengths [13] (some games address the latter concern by level-locking most of the features, so as to not overwhelm the new player with options and explanations). Another design challenge that flows out of the any-place-any-time property of mobile games, is that the intended play sessions should fit into the short intervals of time a player might have. The 2019 Game Analytics report on Mobile Gaming Benchmark showed that the median play session length is about 4-5 minutes [68]. With the unique restrictions and affordances of smartphones as gaming devices, both academics and industry professionals alike vie to understand how to make best use of this new platform. Some of the recent efforts in that regard are reviewed in the next section.

2.4.2 Psychology behind mobile game appeal

Mobile devices possess unique affordances when compared to other gaming platforms, including but not limited to any-time-any-place accessibility and touch controls. While making use of these unique affordances is at the heart of my research, it is also worthwhile to include game design principles that apply more broadly. This section draws on literature in both game design and gamification. While game makers and gamification gurus ultimately have different goals and different approaches to designing their user experience, there is a significant overlap between the methods employed by the two fields. The main goal of this section is to establish a vocabulary for the subsequent sections.

2.4.3 Free-to-Play (F2P) games

I chose to use F2P games (particularly those with idle game elements) as a primary model to follow in my research for several reasons:

1. F2P games are a popular mobile medium.

2. The structure (discussed below) is well suited for a learning game for non-sequential material—unlike math, in which many concepts build on one another, in the vocabulary acquisition stage of language learning, rigid order of material presentation

isn't necessary. For example, it might be prudent to learn the names of domestic animals before learning exotic ones, but you don't have to know how to write "cat" and "dog" in order to learn to write "elephant".

3. F2P games employ many retention strategies that I'm curious to explore in the context of an educational game. Are they capable of promoting prolonged engagement, and does that result in increased learning?

2.4.4 Minigames

2.4.4.1 What's and Why's of Minigames

In videogame lingo, the term "minigame" can refer to several different things. In this document, I use the term to describe fairly simple games/activities that could conceivably stand on their own, but are included in a larger game [161].

Traditional F2P games that flooded Facebook in years past limited the player's playtime either through energy/health mechanic, or through increasingly long wait-times for in-game tasks to complete [5]. For examples, think of running out of hearts in Candy Crush Saga [114] or waiting four hours for a peach to grow in Farmville 2 [205]. The player has to either put the game away for a time, pester their friends for assists, or pay hard-earned cash to speed up the in-game clock.

However, there's a novel trend among mobile F2P games that make use of the Zeigarnik effect, and it is inclusion of minigames [69, 184, 185]. These games impose waiting periods too, but they offer alternative activities for players to do while they

Figure 2.8: The Cat Game screenshots [69]. (a) The game's main screen where the player's cat collection takes place. (b) The Poppy Cats minigame, where the player must quickly merge same-colored cats into biggest possible combos and pop them to earn points until the time bar below runs out. (c) The Blocky Cats minigame, where the player places given parts onto the field, aiming to create complete lines, which disappear similar to Tetris. There's no timer, and the game is over when the player can't fit a given shape onto the screen.

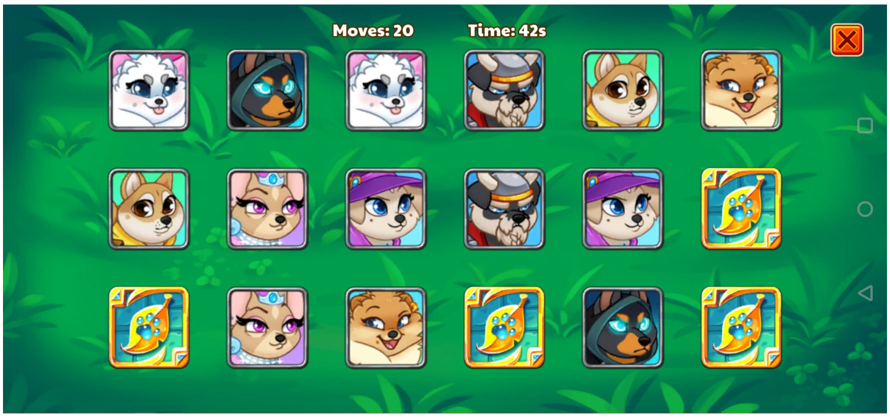


Figure 2.9: Castle Cats minigame screenshots [184] (These minigames are duplicated in the same company's dog version of the game Dungeon Dogs [185].) Top: Murrmory (or Donna's memory in the dog version) is a simple memory matching game. Bottom: Matching Mania is a variation of the match-3 Bejeweled clone.

wait for timers to complete (Fig. 2.8 and Fig.2.9). These simple colorful minigames provide limited amounts of in-game resources to encourage play for a short time, but not quite enough to justify considerable time investment. That is to say, if you're bored waiting for a train, you might as well spend the few minutes earning coins. Alternately, if you open the game and see you have five minutes left for a timer to complete, you might help pass the time with a minigame round or two. Finally, if you're a little short on resources for a big craft, you might play a few minigame rounds to cover the deficit, rather than waiting for resources to slowly accumulate on their own.

This is the model I'm interested in emulating: games that encourage waiting but also offer an activity to pass the time while you wait. In an educational game, timers could pace the student's progress through learning material to prevent the student from taking on too many new characters at once. The minigames could focus on reviewing previously learned characters. By continuously reviewing old material, players could speed up the timers on new characters. Notably, while reviewing is a crucial part of the learning process, the benefits diminish as the length of review/study session extends past a certain point[181]. This could be reflected in gradually reducing the rewards the player earns with consecutive practice sessions to encourage breaks.

2.4.4.2 Educational Minigame Design

F2P games that include minigames usually have at least two of them and they targets different skill sets. For example, quick hand-eye coordination like Poppy Cats vs. logic puzzles like Blocky Cats (Fig. 2.8 (b) and (c)); timed Murrmory vs. turn-based

Matching Mania in Castle Cats (Fig. 2.9). This gives players a choice of an activity that might fit their mood/circumstances/preference. Following this example, I chose to develop two minigames as the core of my LWS learning game.

Minigames are where the bulk of the learning/reviewing will take place in my experiments. It's also where I explore mnemonics and intrinsic pleasure / motivation. Zeigarnik and Gacha are auxiliary devices supposed to make the students want to spend more time playing the minigames. In order to study the effects of these auxiliary devices on students' time in-game, I must have some minigames for students to play.

2.4.5 Features linked to success

The current research on features linked to success in mobile games is somewhat contradictory. There have been efforts to use statistics and machine learning to analyze the top performing (downloads and revenue) games from the Google Play Store and Apple's AppStore to identify features predictive of a game's success. The results weren't consistent across the papers, which may indicate (as some of the authors suggested) that this method may not be well suited to solve the problem at hand. For example, one study concluded that the best predictor of the number of downloads was the marketing budget, which had nothing to do with game features. Another thing to keep in mind is that studying the attributes of successful games doesn't directly tell us about what made them successful – it tells us about how much variation can exist in a design while not blocking that success. This method is subject to survivorship bias [195].

Alomari et al. considered 29 game features and analyzed 50 Apple App Store mobile games from the top 500 grossing games in the US [6]. Their conclusion was as follows: "the major factors that drive mobile game app success include the extent of advertisement investments and approaches; the presence of violent and sexual content; the opportunity to gamble; the opportunity to compete, and the illusions of fantasy. [...] irrespective of the genre, the game has to have an addictive element, as opposed to 'just for fun'." The authors didn't elaborate on what they considered an addictive element.

Karthic et al. considered 24 game features found in 60 games - top and bottom 10 from each of the three categories: top downloaded free games, top downloaded paid games and top grossing games [107]. They attempted to identify the reliability with which individual features or pair-wise combinations of features could predict a game's success. Some of the features that made the top 10 success prediction list both in individual and pairwise analysis were as follows: Points (numeric reflection of a player's progress), Challenges (puzzles or other challenging elements in a game), Virtual Goods (game assets with perceived or real money value), Leader Boards, and Content Unlocking. Limitations of the study included a low sample size and treatment of game features as binary values without taking into account the quality of implementation.

Moreira et al. selected 37 game features and evaluated top 100 games from Google Play store's top download and grossing charts [143]. Their analysis showed that Achievements, Mobage integration (a Japanese gaming social network) and In App Purchases (IAP) were positively correlated with the game's position in the top downloads chart; while Invite Friends, Item Upgrade and Status Upgrade (increasing the characters stats, such as strength or health) were negatively correlated. On the grossing side, Random Elements, Event Offers and Versus Mode were positively correlated, while Customization and Soft Currency Gambling were negatively correlated.

In a follow-up to Moreira's study, Filho et al. analyzed 60 games from the top 100 games in Google PlayStore, and 40 games from the 400-500 range [63]. They found that unlike the six features correlated with a game's rank on the download chart in Moreira's study, no features were statistically tied to downloads. After consulting with game industry specialists, the authors concluded that the download rate was largely tied to external marketing campaigns, which overshadowed the possible effects the features might otherwise have. As for the revenues, instead of five correlated features they found 12. Positively correlated ones include Timed Boosts, Soft Currency, Leaderboards, Time Skips, Hard Currency Gambling, Request Friend's Help, Consumables, Facebook integration, Levels. Negative correlations include Hard Currency, IAP Potential and Gambling. In this study, only permanent items, such as level packs or advertisement removal were considered as IAP's.

Using a different approach, Chou et al. conducted player surveys to identify game characteristics contributing to gamer flow [24]. These included the following statements: "I can feel my power and control over the game world," "I can design my game charter to express myself," "The game is easy to pause and escape when interruption[sic]," "The game can protect me from losing precious accumulated record of accomplishment," "The game's help function is handy." While most of these present a sound advice for good game design in general, character customization is the only game mechanic on this list.

In a 2019 CHI-Play Work-in-Progress category paper, Legner et al. similarly examined the top performing games, but opted for psychological analysis of common features, with the goal to explain the motivational pull of each mechanic through existing psychology theories [121].

In their book Mobile Game Design Essentials, Dr. Scolastici and Nolte detail controls available to a mobile game designer and give extensive UI advice, but say nothing about game mechanics [172]. There are industry blogs on the subject, however they often focus on monetization and usually only contain very general advice, such as "Tell a story," or "Make it easy, but addictive," or "Create a stunning design," [9] and "Bring in Social [media]." A Forbes article titled "This Is The Formula For Creating An Insanely Successful Mobile Game" suggests "Do something no one else has thought of. Take an old, seemingly complicated genre and completely re-imagine it for a touch interface"[66].

Similarly to Legner's work, in the next chapter I present analysis of some motivational features in moderately popular games and discuss how they could be applied to an educational game. This analysis will sidestep the issues of survivorship bias by not counting what features were present in successful games but instead focus on how those features worked in their specific contexts.

Chapter 3

Mobile Game Case Studies

3.1 Case Study 1. The Walking Dead: Our World

3.1.1 Game Overview

TWD: Our World [70] is an official augmented reality GPS-based mobile game for the popular AMC TV series The Walking Dead. Since its launch in July 2018, the game has been downloaded over 5 million times on Google Play store alone. The premise of the game is simple: the map of the local area (courtesy of Google Maps) is populated with zombies, raiders, supply crates, and survivors in distress. The player must walk around to interact with the objects on the map (Fig. 3.1).

The game uses cards as a power-up mechanic. The player collects character, weapon and perk cards as rewards for completing missions. When a certain number of a specific card is collected, the player can spend some in-game gold to level-up the power of that card.

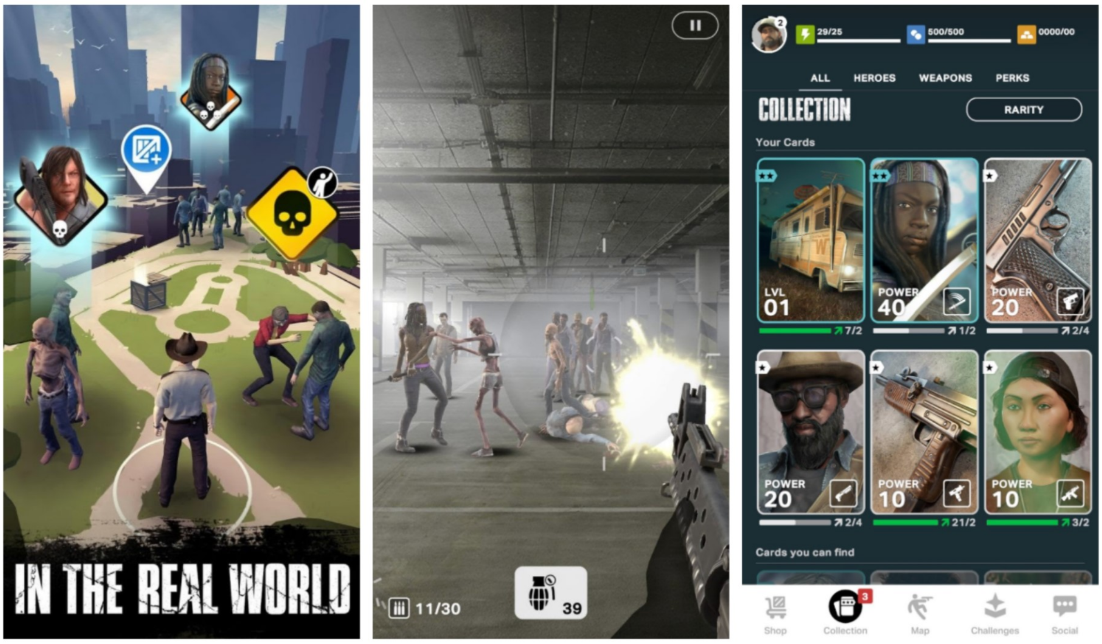
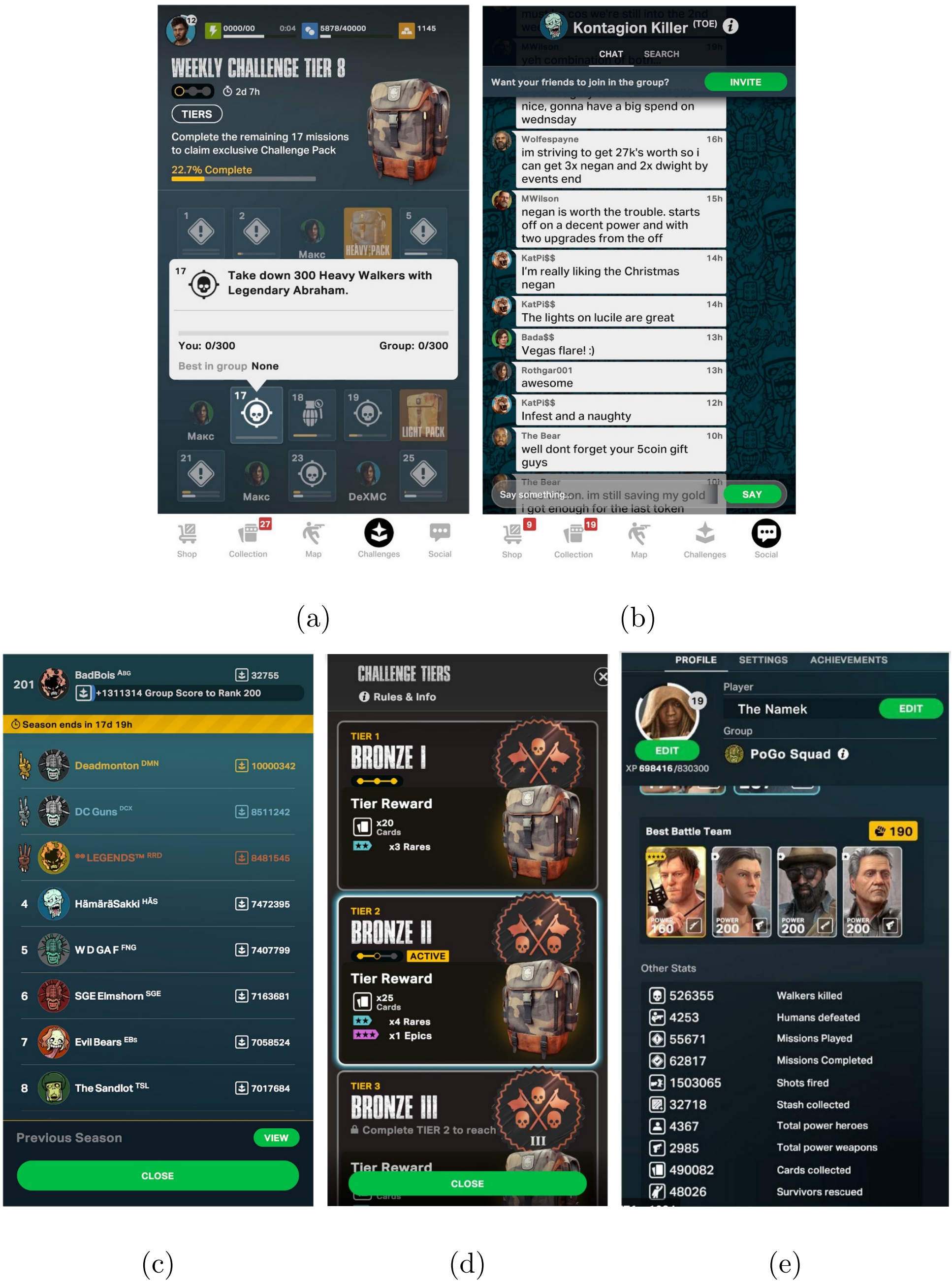


Figure 3.1: TWD: Our World screenshots. Left to right: the game's overworld view; FPS battle mode; Power-up screen.

Encountering enemies results in a First Person Shooter (FPS) style gameplay, where the player, assisted by a chosen character and weapon, clicks on the approaching zombies to shoot them.

3.1.2 Clan Mechanic

The game is both collaborative and competitive. Both features are implemented through the clan mechanic. Players are encouraged to organize into groups of up to 25 players guided by a single clan leader and any number of co-leaders and officers. Clan members have a private chat to talk and coordinate with their clan-mates (Fig. 3.2b) as they work together to complete challenge boards (Fig. 3.2a).



(a) (b) (c) (d) (e)

Figure 3.2: TWD: Our World screenshots. (a) Weekly clan challenge board (b) Internal clan chatroom (c) Global clan leaderboard (d) Seasonal challenge tiers (e) Player profile.

The gameplay is temporally divided into seasons (as in seasons of a TV show, not the time of year seasons). Each season brings with it a set of tiers (Fig. 3.2d), and each tier consists of three challenge boards. The seasons are further subdivided into weeks. If a clan is part-way through completing a tier at the end of the week, their progress is reset to the beginning of that tier when the next week begins.

Since this is a post-zombie-apocalyptic game, the tasks include things like "Kill 100 rotten zombies with a shotgun," "Rescue 50 survivors with Carol," "Build 15 shelters," etc. With each completed board, the tasks become more specific and the numbers increase dramatically. At any point, each player can see exactly how much they've contributed to a specific task and which clan member contributed the most to that task. Once a challenge has been completed, the challenge tile is replaced with the avatar of the player who contributed the most (Fig. 3.2a). Completing challenge boards unlocks rewards for the entire team.

The competitive aspect of the game is implemented through a clan leaderboard, where a clan's standing reflects the amount of points its members earned for completing the challenges (Fig. 3.2c). It is noteworthy that there are no in-game rewards for topping the leaderboard, short of bragging rights.

There are casual teams, who play as much or little as they wish, and there are serious teams, vying for a spot on the global leader board. Those teams often require players to log in daily and complete a certain number of missions or zombie-kills (information available on a player's profile page (Fig. 3.2e)).

In my time with the TWD: Our World community (a year spent in casual and serious teams, eventually leading a semi-competitive team of my own, and scouring Reddit forums for information) I found that players found the challenge boards a highly motivating mechanic. While most teams couldn't dream of making it to the top 200 list on the global leaderboards, completing the weekly challenges tended to keep players engaged. The high visibility of individual contributions gave players a sense of ownership and achievement over "their" tiles. Active players often make a friendly competition out of it, trying to get their avatar onto the board as many times as possible, or get a "Bingo" with their avatar taking up five tiles in a row. Beating the challenge boards often resulted in cheers all around.

The clan structure addresses several of the players' psychological needs, as well as employing what Yu-kai Chou called black-hat motivation. Collectively working toward a common goal (completing challenge boards and progressing through the tiers) gives a sense of belonging, a positive form of Social Interaction. The visibility of a top contributor on any given challenge plays into the Need to Achieve as well as giving potential to peer pressure of contributing to the challenges on equal footing. The boards' relatively short duration plays into the Limited Time Content, triggering FOMO and loss avoidance. Often, the teams make a hard push to complete a tier in the last day or two to avoid losing their progress when sent to the beginning of the tier. The board completion rewards contain an element of gacha, since they are a randomized arrangement of cards (higher rarity cards are awarded for higher tiers). Finally, the clan leader board provides a low-key competitive element to indulge the competitive players without discouraging the non-competitive ones.

3.1.3 Periodic Quests

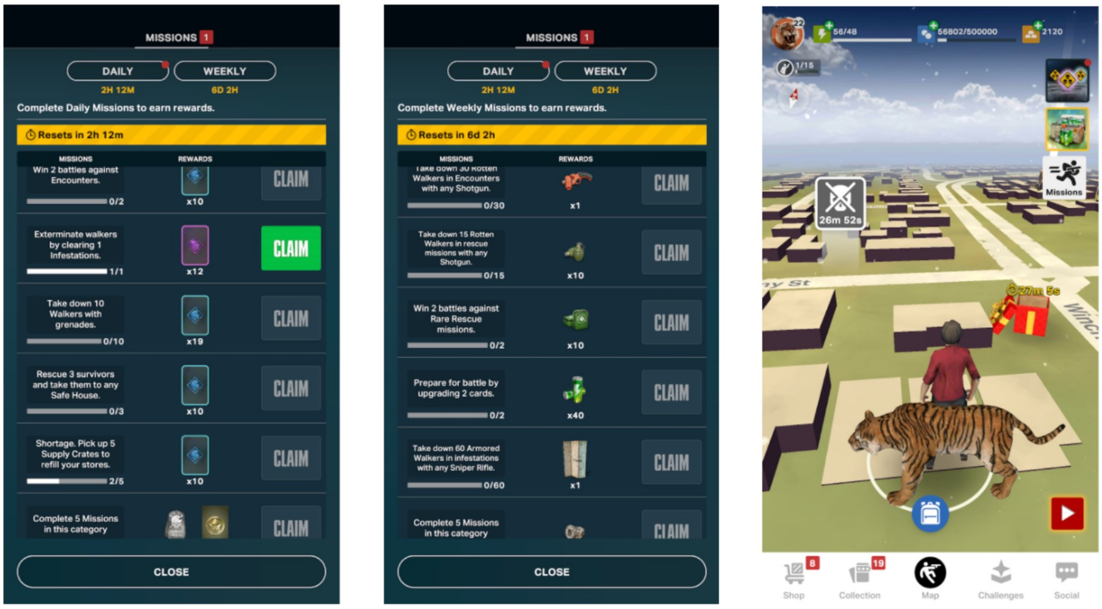


Figure 3.3: TWD: Our World screenshots. Left to right: Daily individual mission list; Weekly individual mission list; Overworld with interactable items on reset countdown.

TWD: Our World uses several methods to encourage players to log in at regular intervals (Fig. 3.3). For example, things like supply crates and minor missions respawn every 30 minutes, tempting the player to open the game every so often, even if they don't have an opportunity to go outside and play, just to "farm" goodies accessible from inside their home or workplace.

In addition to the clan challenges, the game also offers daily and weekly individual quests. The daily quests include easier tasks and smaller rewards, but the weekly ones allow the player an opportunity to get some in-game items that are otherwise available only with real-world money.

Aside from encouraging the player to play every day or at least every week, these quests serve another purpose. The wide variety of tasks ensure the player doesn't forget about the many mechanics available in the game. In a sense, this is very similar to the clan challenges, but has the benefit of engaging even the lone-wolf players who don't wish to participate in a clan. Also, if the player went through the trouble of starting up the game and completing quests, they may just keep playing a while longer, even after the quests are fulfilled [67].

3.1.4 Educational Game Application

Many of TWD: Our World clan structure elements could be employed in an educational game. For example, Duolingo uses leaderboards for some friendly competition to motivate its language learners to keep working on their lessons and earning more points, but without clans, it misses out on the motivational benefits of collaborative play, be it white-hat achievement and a sense of belonging, or the black-hat peer pressure.

The challenge boards and daily quests have a potential to lend themselves to a learning game. In a kanji-learning game, the possible challenges could include tasks like "practice writing hanzi 500 times," "unlock 10 kanji with more than 10 strokes," "successfully complete 200 reviews," etc. The challenge board set-back mechanic could help with a bit of loss aversion motivation. While making use of the aforementioned engagement elements, all of these tasks would directly contribute to a player's mastery of the language.

A player profile, with a detailed account of all of the player's activities similar to that of TWD: Our World is a convenient way to visualize the player's accomplishments. The stats could include things like number of kanji unlocked, number of strokes drawn correctly, review accuracy and so on. They can also help with a sense of personal accomplishment and growth.

The intra-clan leaderboard would provide the low-key competitive element.

3.2 Case Study 2. Cat Game – The Cats Collector!

3.2.1 Game overview

Cat Game (CG) is an idle game, where the core progression of collecting cats and crafting decorative items revolves around waiting and heavy use of Gacha and the Unfinished Business effect. The core gameplay includes the following elements: cats occasionally drop gold; the players use gold to buy cat food; after a cat is fed, a timer starts (Fig. 3.4 a.); after the wait time is over, the user gets a basket which may contain a new cat or crafting materials. Some cats can only be unlocked after certain decorative items have been crafted. The players use crafting supplies and gold to craft more sophisticated supplies (yarn into ribbons, wood into planks, etc), and then use those (and more gold) to craft the items (Fig. 3.4 b.). Both the cat basket timers, supply crafting times, item recipes and prices escalate dramatically as the player reaches higher and higher levels in their collection.

Figure 3.4: Screenshots of Cat Game: a. home screen; b. crafting screen; c. supplementary activities selection screen with a level-locked item; d. avatar and showroom customization screen.

3.2.2 Unlockable Content

Unlockable content is a hallmark of practically any game with a progression. There are many types of unlockables: character customization (outfits, animations, titles, etc.); visual and audio collectibles (unlocking items in galleries); achievement trophies and badges; access to new levels, characters, mechanics and so on.

There are two categories of unlockable content in CG. The first are the agency-restricting features. As mentioned in the Mobile Games section, it's easy to overwhelm a new player with a busy interface and a multitude of options. To prevent such overload, CG gradually introduces game features to the player (Fig. 3.4 c.). Initially, the player can feed the cats, do basic supply crafting, and participate in one daily contest. After gaining some levels (which requires proficiency in the game's base activities), the player gradually earns access to more contests, the club feature (just like clans or guilds in other games) and progressively more complex crafting recipes.

Next are the collectibles. The player earns stars each time they craft a decor item or unlock a new cat. The stars are used to unlock new rooms in the tower, along with a themed collection of cats and decor items to unlock. There is also the Premium Cat Tower. By completing daily tasks, the player might obtain a premium key. With those, they can unlock a random cat or decor item from the Premium Tower. Normally, the Premium Tower contents are only available at certain events and promo-packs (which cost real money). One level of the Premium Tower, the Club Kitties room is special. Elements of this room can only be unlocked by participating in an active club. Unlocking a premium item is a rare event, and players are often eager to show off their premium cats and items in their contest entries or customizable player rooms (Fig. 3.4 d., 3.5 a, b.).

3.2.3 Supplementary Activities

CG makes heavy use of the Unfinished Business mechanic. A few days into the game, a player might have over a dozen of baskets and materials in different stages of completion (Fig. 3.4 a, b.). To avoid losing players' interest as the wait times grow from minutes into long hours, CG offers the players plenty to do while they wait. There is a club system with member chat, where players can pass some time. Players can use their favorite cats and props to deck out their showroom - something visible to other players (Fig. 3.4 d.). They can participate and vote on themed daily contests, trying to select cats and props from their collection that best match the theme (Fig. 3.5 a.), and then, looking at two rooms at a time, try to guess community's favorite in return for a bit of gold (Fig. 3.5 b.). Finally, there are the minigames (Fig. 3.5 c, d.).

The minigames are nominally cat-themed, are hardly tied into the main game, or give a feeling of progression (there is a limit on how much gold a single play-through can award, no matter how skillful the player becomes at it, and those amounts become inconsequential as the player progresses through the game). Still, they are colorful and full of pleasant sound and visual effects, and satisfying combo-pops - the embodiment of intrinsic pleasure. Notably, Blocky Cats (Fig. 3.5 c.) is a logic puzzle and the player can spend unlimited time contemplating their each move, while the Poppy Cats minigame

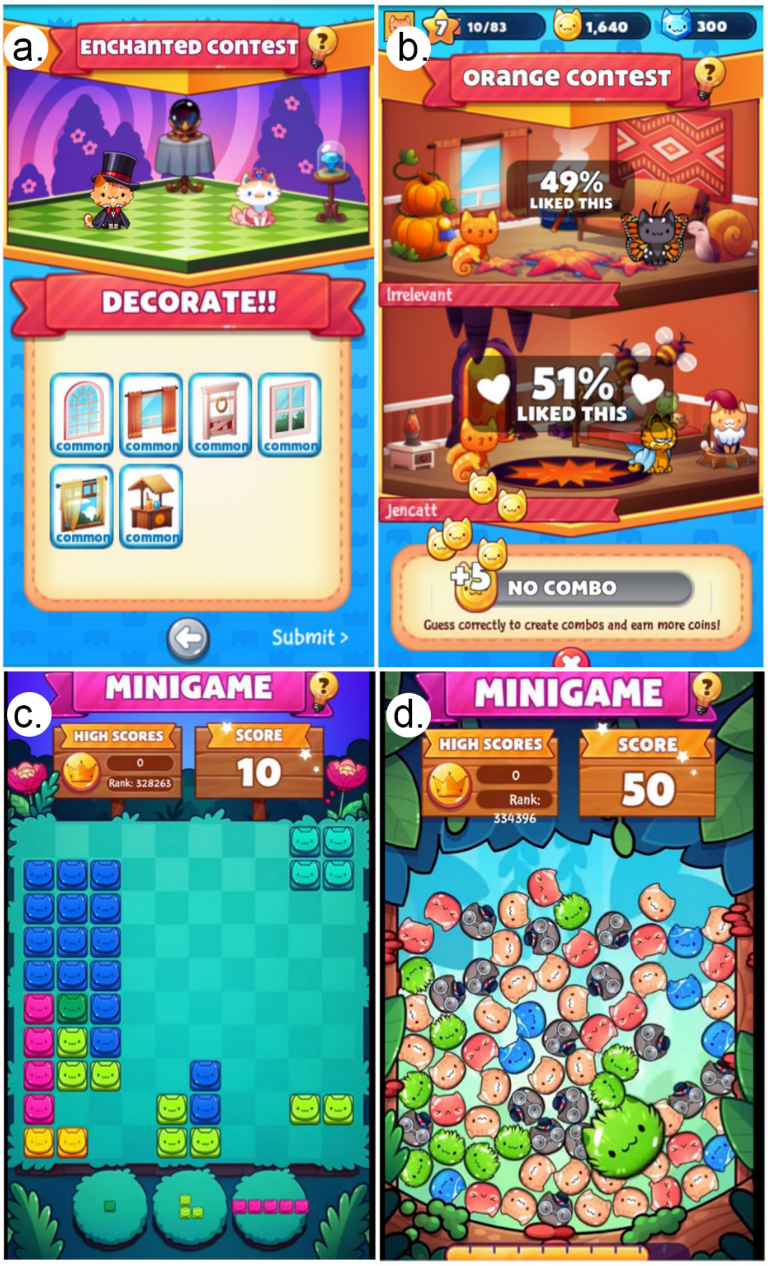


Figure 3.5: Screenshots of Cat Game: a. daily room decoration contest; b. contest voting screen; c. Blocky Cats minigame; d. Poppy Cats minigame.

is a fast hand-eye coordination game with a short timer for each round (Fig. 3.5 d.).

This array of things for the player to do outside of the game's main purpose (collection of adorable cats) is not unique to CG. It's a common feature among popular mobile games. The assortment of activities, each with a different level of mental energy required, allows the player to select one appropriate to the amount of time and attention they have at any given moment, and allows them to avoid boredom by switching up the activity whenever they wish.

3.2.4 Educational Game Application

The concept of unlockable content lends itself well to an educational game. A student should show some proficiency with the current material before proceeding to the next topic. In case of a language game, similar to the themed rooms in the CG tower (i.e. farm, nightmare, etc), there could be themed rooms of words and grammar a student could work on, gradually earning their right to start working on the next section. Similar to the special items in the Premium Tower, some of the more specialized words could be placed in the premium area. For example, only so many food words can be reasonably included in the Breakfast section. However, additional words could be unlocked through a bit of extra work.

Collectibles can be a bit more of a challenge to meaningfully integrate into an educational game, lest it becomes a case of thinly applied chocolate-covered broccoli. To be effectively motivating, the collectibles need to be desirable. To be desirable, they need to be intrinsically pleasing, which translates into high quality digital assets

Figure 3.6: Left: screenshot of Radical Tunes, a musical game for learning to write kanji [111]. Right: screenshot of Zen Hanzi, a puzzle-like hanzi-learning game [109].

(expensive and/or time consuming to produce). Costs aside, all the usual culprits can be included - avatar and color theme customizations, badges, etc.

If a language learning game were to use the Spaced Repetition System [178] (in use with many prominent language learning software like Rosetta Stone [182], Duolingo's Tinycards [53], and WaniKani [191]), wait times between the reviews are comparable to CG's idle waiting mechanic, because with each successful review, the wait time for the next review increases exponentially. This sets the stage for minigames and other activities. However, unlike the minigames in a generic F2P game meant for nothing but idle entertainment, minigames designed for educational software must contain educational value in addition to intrinsic pleasure.

Returning to a hypothetical LWS learning game, a few minigames come to mind: a fast-paced musical kanji-drawing game used as a core mechanic in Radical Tunes (Fig. 3.6) could be an analogue to Poppy Cats, and an at-your-own-pace hanzi-assembly game used in Zen Hanzi (Fig. 3.6) could serve as a Blocky Cats analogue.

Other possibilities could include the following mechanics: given a set of radicals, compose as many different hanzi as you can (this is similar to the English version of the game, where given a set of letters, the player is encouraged to come up with as many words as possible using nothing but those letters); and match hanzi into groups if they share a reading or tone.

The games can provide some small reward to the player. To keep the rewards in theme with the rest of the game, these could be various amounts of points that can be used to unlock new hanzi.

Having a variety of minigames offers the player a choice of productive pastimes, as each game would reinforce something they've learned. At the same time, the small but relevant rewards may encourage the player to play multiple rounds.

Some things can also be done with the contest mechanic. For example, players can be given a short proverb or poem to translate into Chinese to the best of their ability. The other players can then vote which translation was the most accurate and creative.

Once again, the key here is to offer a variety of activities without detracting from the primary purpose of the game - to teach hanzi.

Chapter 4

Radical Tunes: A Mnemonic Musical Game for Learning to Write Kanji

4.1 Game Overview

Radical Tunes is a musical game for learning to write kanji, the Japanese characters borrowed from Chinese language.

Many mnemonic devices exist for remembering the shape and meaning of a kanji [165], but to my knowledge, beyond several rules of thumb, fraught with exceptions, none of these are focused on learning the proper writing stroke order—which is vital for handwriting proficiency[130]. I created *Radical Tunes* to fill that void. Past studies have shown the positive effects of music on memorization of text, especially when it is rhythmically aligned with the material [196]. I wanted to explore if this music mnemonic effect could be extended to memorization of the stroke order of a kanji.

Radical Tunes implemented the intended musical mnemonic mechanism by employing a unique melody for each radical, thus giving each radical a melodic fingerprint of sorts. These melodies are further broken down, with each element of the melody corresponding to a stroke within the radical. As a result, as the student correctly draws a kanji, they will hear a unique melody corresponding to that kanji unfolding with every stroke. To further amplify the effect, I matched the changes in melody's pitch to the direction of the strokes (Fig. 4.1). E.g., a downward stroke would go down in pitch. This design leverages the concept of conceptual/embodied metaphors [118, 162], which have been shown to map movement to music effectively [8, 7]. The melody of the "tree" radical 木 consists of the melody for the "ten" radical 十 with the additional two stroke tunes.

The version of *Radical Tunes* used in the first study used pre-recorded melodic elements for each stroke, which allowed me to use different instruments for each radical, to further distinguish them from each other. The music for this version of the game was composed by Oscar Alfaro. However, if the player drew the stroke too quickly or too slowly, the sound and the drawing would be out of sync. For the second study I used procedural sound generation, which made for a more responsive drawing experience, as the sound could last for as long as the player's finger touched the screen. However, this limited the game to only the single type of melodic sound that I was able to procedurally generate. This is relevant, as an informal post-hoc survey after the second study revealed that many of the melodies were perceived as too similar, which may have muddled the mnemonic effect. This is further discussed in the Implications for Design section.

Figure 4.1: Melodies for the radicals “tree” (above) and “water” (below). I tried to make the pitch of the melody follow the direction of the strokes. A horizontal stroke is represented by a steady note. A long downward stroke is accompanied by a significant drop in pitch, while a shorter upward stroke is matched by a moderate rise in pitch.

4.2 Study 1

4.2.1 Methods

The goal of the pilot study was to investigate the impact that music would have on learning in *Radical Tunes.* Given that immersion can play an important role in learning for educational games [23, 40] and music has been shown to impact immersion [171], I also wanted to examine if the musical nature of *Radical Tunes* increased player immersion.

The study had two versions of the game:

1. The original *Radical Tunes* with unique tunes for each radical and melodic ele-

亻口木休回困

Figure 4.2: Radicals and kanji used in the first study. Left to right: person, mouth, tree, rest, episode, annoyed.

ments accompanying each stroke.

2. The control version where the music was replaced with the sounds of chalk writing on a blackboard.

For the initial study I selected three of the six most commonly occurring radicals in the Jōyō kanji (the 2,136 characters required for baseline literacy). I also selected three kanji which consisted of combinations of those radicals, to add increased complexity, for a total of six characters for the subjects to learn (Figure 4.2).

4.2.1.1 Procedure

The participants were told that the study was for an educational game where they would learn the meaning and writing of six Japanese characters. They were then randomly assigned to either the music or control condition and given an Android tablet with the appropriate version of *Radical Tunes* loaded. For the experiment, meaning and stroke order tests were built into the beginning and end of the game (Figures 4.3(a) and 4.3(b)).

For the meaning tests, the participants were shown the characters one at a

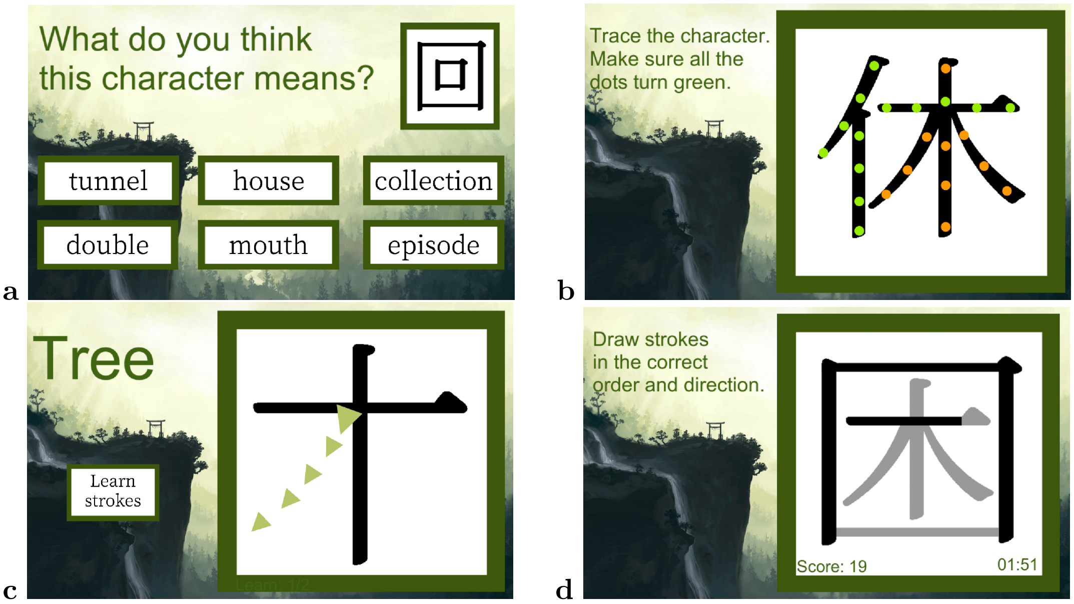


Figure 4.3: *Radical Tunes* screenshots. (a) Meaning pre- and post-test. (b) Stroke order pre- and post-test. (c) Stroke learning screen. (d) Timed and scored game portion.

time and had to choose an answer from 6 possible meanings (Figure 4.3a). The options were the same for pre- and post-tests. For the stroke order tests, the participants were asked to trace each character on the screen with their finger. I placed orange dots along the lines, and the dots turned green when touched, to indicate which sections of a kanji were already traced (Figure 4.3b).

No feedback was given on these evaluations, to prevent the participants from learning during the pretest via confirmation of their guesses.

After the pre-tests for meaning and stroke order, the participants were taken through the educational section of *Radical Tunes*. First, the subjects were shown the animation of the correct stroke order with accompanying sounds. Then, they were asked to draw the strokes by following the arrows on the screen (Figure 4.3c). Finally, they were asked to trace the character unassisted. A soft ping would let them know if they made a mistake, and they would have to redraw the latest stroke. To ensure that each participant received the same amount of instruction, I limited each step to two unskippable repetitions.

Next, participants played the *Radical Tunes* game by tracing the characters appearing on the screen as quickly and correctly as possible (Figure 4.3d). The game included each character twice in a randomized order. The same randomized order was used for all participants.

After the game was completed, the final score and time were displayed, and the participants were directed to take the meaning and stroke order post-tests. Lastly, they completed a demographic and experience questionnaire on a laptop.

4.2.1.2 Participants

Twenty four participants took part in the study. They were recruited through class announcements and by word of mouth outside of the university. There were 11 female, 12 male and one non-binary participant. The ages ranged from 19 to 51 with the average age of 25.6 (SD = 7.8). During the study, participants were randomly assigned to one of the two conditions: music (7 female, 5 male) and control (4 female, 1 non-binary, 7 male). All participants reported no prior knowledge of Japanese or Chinese.

4.2.1.3 Measures

- **Immersive Experience Questionnaire (IEQ)**

  I used the IEQ [98] to measure the level of immersion experienced by the subjects. IEQ takes into account five dimensions of immersion: cognitive involvement, emotional involvement, real world dissociation, control and challenge. The cognitive involvement dimension covers factors like attention and effort the subjects invested into the game. Emotional involvement covers the degree to which the subjects enjoyed the game and were interested in seeing more of it. The real world dissociation measures the degree to which the participants lost track of time and their surroundings. The control dimension assesses the perceived responsiveness and ease of use of controls. The challenge dimension measures the perceived difficulty of the game. In their paper, Jennett et al. suggested that immersion is an important component of what is perceived as a good game. For *Radical Tunes* to serve its educational purpose, it is important for it to be able to hold the players' attention and provide an enjoyable experience that entices them to return. With this in mind, I decided that IEQ would help me measure how well *Radical Tunes* performed in this sense.

- **Radical Tunes scores**

  The participants were given pre- and post-tests for meaning (Figure 4.3(a)) and stroke order (Figure 4.3(b)) for each of the six characters. I used the relative improvement between the pre/post tests to assess the learning outcome of the game.

4.2.2 Results

4.2.2.1 Prior Knowledge and Experience

According to a series of independent samples t-tests, participants in the two conditions did not differ with respect to prior video game experience ($p = .61$), prior rhythm game experience ($p = .81$), and on the pre-test learning outcome measures: kanji meaning ($p = .82$) and stroke order ($p = .31$).

It can therefore be assumed that participants in both groups had similar prior game experience and knowledge of kanji for the following analyses.

4.2.2.2 Immersion Experience

I first examine participants' experience of immersion during gameplay for both conditions (Table 4.1). To analyze differences between IEQ scores for the two conditions I used an independent samples t-test. Table 4.1 shows descriptive statistics for IEQ scores, as well as significant differences and effect sizes between conditions. Results found a significant difference in favor of the music condition of the game increasing *overall IEQ scores* ($p = .04$, $r = .41$), *cognitive involvement* ($p = .026$, $r = .44$), and *emotional involvement* ($p = .034$, $r = .42$). There were no significant differences for real world dissociation, challenge, or control (all p's $> .063$).

Table 4.1: IEQ mean scores, standard deviations, significant differences between the two conditions, and the effect size, which is in the medium to large range for significant differences.

	Control		*Music*		*Significance*	*Effect Size*
IEQ Measures	μ	σ	μ	σ	p	r
IEQ Overall	100.33	18.1	115.25	15.29	**.04***	**.41**
Cognitive Involvement	30.08	7.01	36.08	5.13	**.026***	**.44**
Real World Dissociation	25.58	6.49	28.5	5.81	.26	.23
Emotional Involvement	12.33	3.58	15.58	3.48	**.034***	**.42**
Challenge	16.33	1.5	17.67	1.83	.063	.37
Control	16	3.54	17.42	3.15	.31	.21

4.2.2.3 Learning Outcomes

To better understand players' learning outcomes from the game, I analyzed pre- and post-test scores for meaning and stroke order, as well as final scores from the game portion of *Radical Tunes.* Descriptive statistics for the three measures are shown in Table 4.2. A series of independent samples t-tests showed that both conditions had a significant increase in both meaning and stroke order scores from pre- to post-test: control (*meaning* - $p < .001$, $r = .95$; *stroke order* - $p < .001$, $r = .96$), and music (*meaning* - $p < .001$, $r = .92$; *stroke order* - $p < .001$, $r = .94$). This suggests that the game itself was indeed successful for short-term teaching of kanji meanings and writing stroke orders to players. However, there were no significant differences in improvement between conditions for meaning, stroke order, or in-game score (all p's $> .11$). As suggested by my second study, this was likely due to the limited number of kanji tested, not providing enough time and difficulty for the music mnemonic approach to have a

Table 4.2: Descriptive statistics of learning outcomes (meaning tests, stroke order tests, and in-game scores).

	Control		*Music*	
Learning Outcome Measures	μ	σ	μ	σ
Meaning Pre-test	1.08	.9	1.16	.84
Meaning Post-test	5.5	1	5.58	.67
Meaning Improvement	+4.42	1.73	+4.42	1.17
Stroke Order Pre-test	.33	.65	.59	.52
Stroke Order Post-test	5.41	.9	5.67	1.16
Stroke Order Improvement	+5.08	.9	+5.08	1.38
In-Game Score	102.17	8.93	106.58	2.19

significant impact on learning outcomes.

4.2.2.4 Perception of Usefulness

I asked participants to rate if they would recommend this game to someone trying to learn Japanese or Chinese on a 7-point Likert scale (1 - not at all to 7 - very likely). The scores were notably higher for the music condition ($\mu = 6.58$, $\sigma = .67$) than the control condition ($\mu = 5.42$, $\sigma = 1.78$). An independent samples t-test showed that there was a significantly higher likelihood of recommending the game for the music condition ($p > .045$, $r = .4$).

4.3 Study 2

4.3.1 Rationale

With the second study I wanted to address the limitations of the *Radical Tunes* pilot by expanding the learning content, increasing the pool of participants and adding a longitudinal post-test that the participants were encouraged to take about 24 hours after the completion of the learning task. Additionally, since music improved immersion in the pilot study, I decided to add music to both versions of the app in this study. I used procedural generation to create ever-changing melodies for the control group and hand authored stroke-specific melodies for the treatment group (the procedural sound generation code was written by Dr. Adam Smith). This was done so that both groups could benefit from the music's immersive properties (or be equally hindered by the increased cognitive load) while only the treatment group was exposed to radical-specific mnemonic melodies.

4.3.2 Methods

4.3.2.1 Participants

A total of 69 participants completed the first step of the study. They were recruited via Amazon's Mechanical Turk and were randomly assigned to either control or treatment group. The participants were provided a link to an Android version of the app and given instructions to not take any notes and only complete the study once. This was done to ensure each participant received an equal amount of instruction and review. Of the 69 first round participants, 47 returned for the follow-up test. Of these, two were eliminated for admitting to prior knowledge of kanji, and one for completing step one of the study twice. Using the Inter Quartile Range test for outliers, I identified and removed two mild lower outliers on the stroke knowledge score, leaving me with the final count of 22 and 20 people in control and treatment groups respectively. There were 28 male, 13 female, and one participant who chose not to disclose their gender. The ages ranged from 18 to 50. There were no statistically significant differences in age and gender breakdown across the two groups (all $p > 0.4$).

4.3.2.2 Procedure

The participants downloaded and installed either a control or treatment version of the app. Next, within the app, they provided basic demographic data, including their age, gender, and level of experience with Japanese/Chinese characters. After that, they took a pretest on their knowledge of the meaning of the 15 kanji selected for the study (Figure 4.4). The participants could not be expected to draw the characters even approximately before having learned them, so the meaning test was the only pretest administered.

The learning part of the experiment consisted of three steps for each kanji. First, the participants would watch an animated demo of how the kanji was meant to be drawn. Next, they would practice drawing the character twice by drawing over the dotted lines, one stroke at a time (Fig. 4.4, center). Finally, they would practice drawing the character three times without dotted guidelines (a hint would appear if the participant attempted to draw the wrong stroke several times in a row). During the demo and each instance of drawing of the kanji, a melody would play, a small part of it corresponding to each stroke. In the treatment group, the melody would always be the same, while for the control group, the melody would change every time.

After the learning part was complete, the participants were instructed to draw

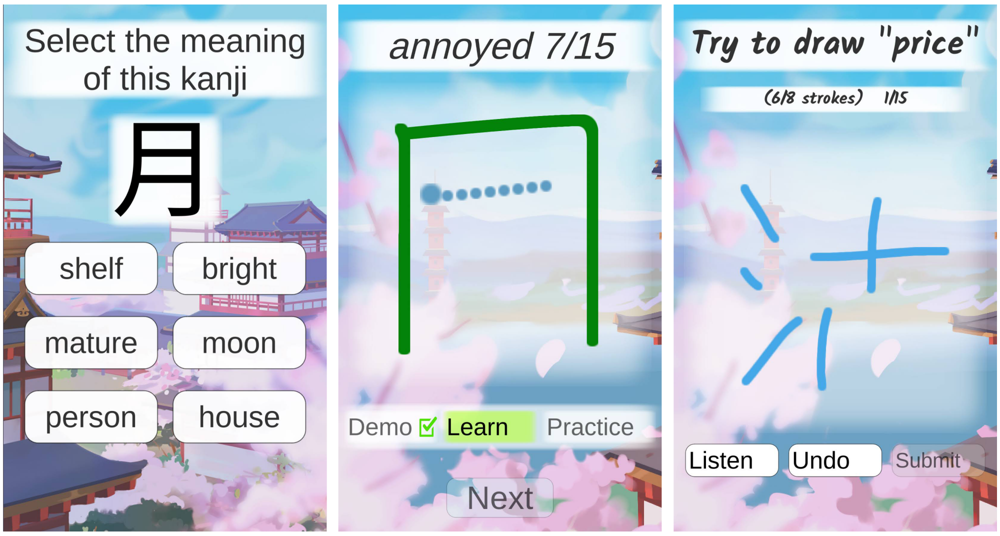


Figure 4.4: Screenshots of Radical Tunes. Left: the meaning test screen. Middle: the learning screen. The dotted line indicates where the next stroke should be drawn, with the biggest dot indicating the beginning of the stroke. Right: the writing test screen. The second line of text displays how many strokes the target kanji has and how many the participants already drew. The “listen” button in the treatment condition would play the kanji’s melody. The button is absent in the control version of the app.

the 15 kanji. For this, they were told the meaning of the kanji, and the number of strokes it had (Figure 4.4, right). Both groups could use the "undo" button to make corrections to their drawing before clicking "submit." The treatment group had to listen to the kanji's melody before attempting to write it, and they could listen to the melody as many times as they liked. I made the stroke number explicitly known on the screen to make up for the fact that the melodic reminder in the treatment group could well give a clue as to how many strokes were in a kanji. Next, the participants were given the meaning post-test, identical to the pretest.

The participants were encouraged to take the follow-up test in the following days. The follow-up test was identical to the post-test, consisting of the kanji writing followed by the meaning test.

One of the limitations of this study is that the Mechanical Turk provided limited means for controlling when the participants would return, and there was significant variance in the interval between the first session and the follow-up test, with the control group taking on average 14 hours longer to take the test with $p=0.03$. However, this did not have an impact on the study results (see Fig 4.5).

4.3.2.3 Kanji

For this study I selected six common radicals and nine kanji that consist of combinations of those radicals for a total of 15 characters (Fig. 4.6). In the treatment group, each radical was accompanied by a designated melody each time it appeared.

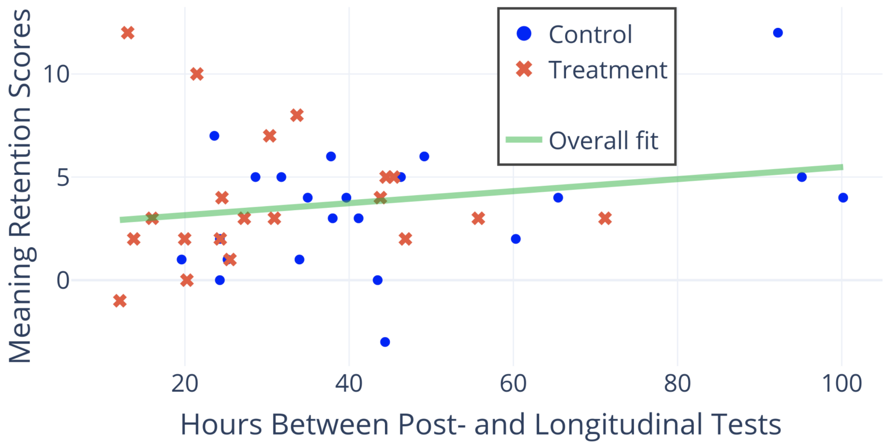


Figure 4.5: The scores for meaning retention (calculated by subtraction of post-test meaning scores from the longitudinal test scores) plotted against the time that passed between the two tests. The linear fit function shows no negative correlation between the length of time and the changes in scores.

Figure 4.6: The 15 characters selected for this study. All of the kanji consist of some combination of the six common radicals (highlighted).

Table 4.3: Descriptive statistics of meaning tests.

	Control		*Treatment*		*Significance*	*Effect Size*
Results for the Meaning tests	μ	σ	μ	σ	p	g
Pre-test	2.86	2.23	3.5	2.96	0.43	0.24
Post-test	10.22	4.31	12.6	2.74	**0.04***	0.65
Longitudinal Test	10.86	4.22	13	2.13	**0.048***	0.63

4.3.2.4 Measures and Scoring

My main objective was to determine whether music could be used as a mnemonic tool to help students acquire stroke knowledge. Stroke knowledge consists of stroke order and stroke direction. Thus, each character was scored on two aspects: the number of strokes drawn in the correct order and the number of strokes drawn in the correct direction. I also collected pre-, post- and longitudinal test scores for the meaning of characters. Even though I didn't expect music to affect meaning retention, I was surprised to see a statistically significant advantage in the treatment group. I discuss this unexpected finding in the Discussion section.

I calculated the scores for the stroke knowledge manually by looking at the images the participants drew during the post- and longitudinal tests and counting the number of strokes drawn in the correct order and direction (Figure 4.7).

4.3.3 Results

I observed a number of statistically significant differences in learning outcomes between the two groups, both for remembering the meaning (Table 4.3) and stroke knowledge (Table 4.4) of the characters.

Figure 4.7: Two of the post-test images for the kanji "foreign." For scoring purposes, the color indicates the order in which the strokes were drawn, while the large initial dot indicates the direction of the stroke. The participant on the left drew the strokes in the correct order, but the first stroke of the moon radical (purple) is drawn in the wrong direction. The participant on the right drew all the strokes in the correct direction (with the exception of the missing last stroke), but switched the order of the first two strokes in the "ten" radical. The participant on the left received 9/9 on order and 8/9 on direction. The participant on the right received 6/9 for order and 8/9 for direction.

Table 4.4: Descriptive statistics of combined stroke knowledge scores.

	Control		*Treatment*		*Significance*	*Effect Size*
Stroke Knowledge test results	μ	σ	μ	σ	p	g
<7 strokes Post-test Order	0.78	0.24	0.81	0.19	0.58	-
7+ strokes Post-test Order	0.51	0.24	0.69	0.25	**0.02**	0.73
<7 strokes Post-test Direction	0.82	0.2	0.87	0.16	0.36	-
7+ strokes Post-test Direction	0.54	0.24	0.72	0.24	**0.02**	0.75
<7 strokes Longitudinal Order	0.79	0.24	0.88	0.14	0.14	-
7+ strokes Longitudinal Order	0.62	0.24	0.7	0.22	0.27	-
<7 strokes Longitudinal Direction	0.81	0.21	0.93	0.13	**0.04**	0.69
7+ strokes Longitudinal Direction	0.66	0.23	0.77	0.19	0.09	-

With respect to learning kanji meaning, the treatment group performed better on the meaning post- and longitudinal tests (p's < 0.05) with a medium effect size (g's > 0.6).

For ease of comprehension I combined the stroke knowledge scores for characters with 2-6 strokes and 7+ strokes (respectively 8 and 7 total kanji in this experiment). The full list of individual kanji scores are at the end of the chapter (Tables 4.5, 4.6), and figure 4.8). The differences between the two groups appear correlated with the complexity of the kanji, with the differences becoming more pronounced for characters with a higher stroke count, at least in the short term (See table 4.4). In Jōyō kanji[1], over 84% have 7+ strokes. Thus, while the results are mixed, the positive effect on complex kanji is more important than the lack of perceptible effect in simpler kanji.

Our results showed that consistent melodies indeed helped the participants remember the characters' stroke composition better, particularly when it came to stroke direction and characters with many strokes in them. It is possible that the simpler characters with fewer strokes were easy enough to remember even without music, and that is why the difference in scores wasn't as pronounced there.

[1] The list of 2000+ kanji deemed necessary for daily life reading proficiency

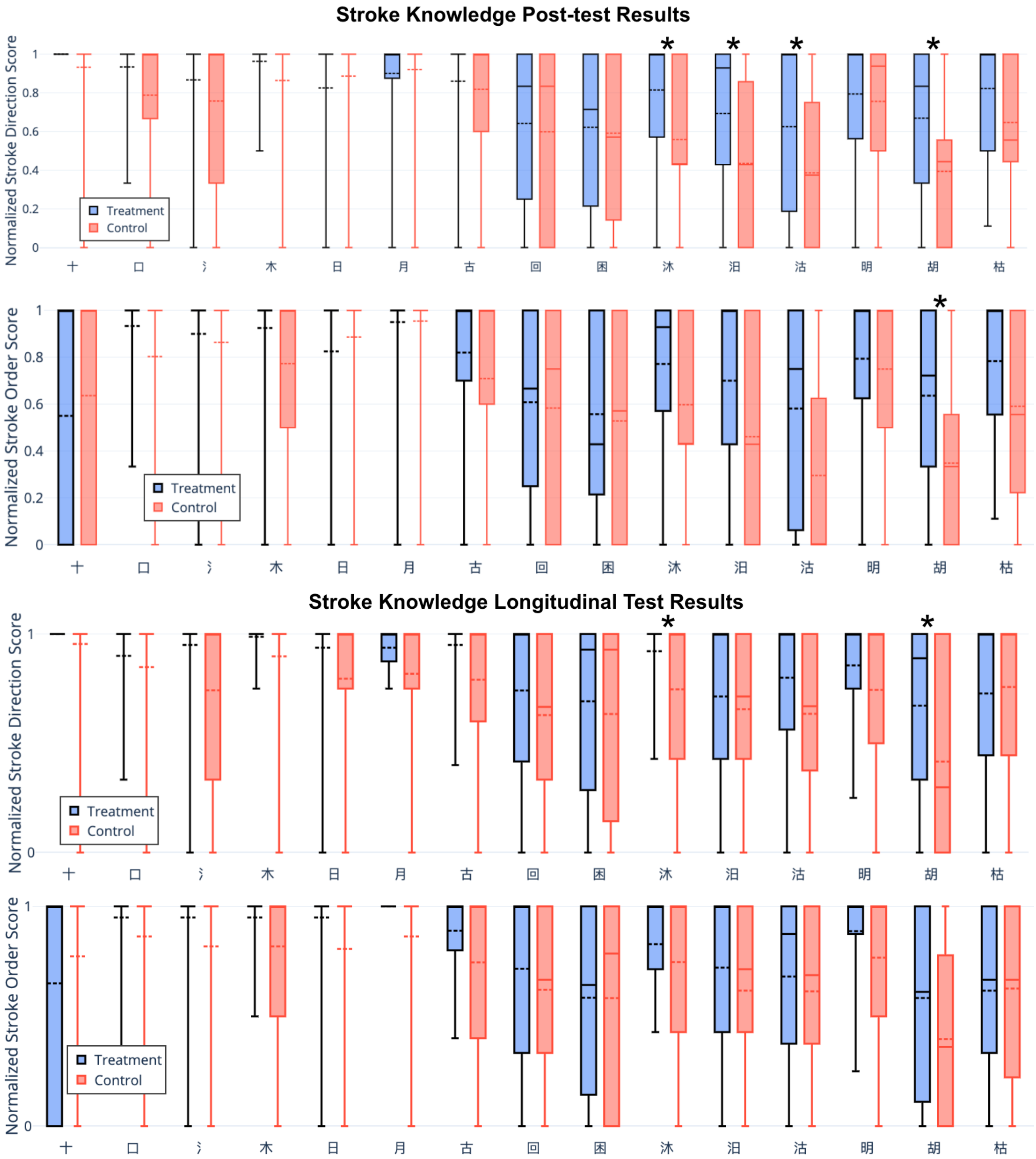


Figure 4.8: Normalized scores for stroke knowledge of individual characters. Top to bottom: stroke direction post-test scores, stroke order post-test score, stroke direction longitudinal test score, stroke order longitudinal test score. The scores with statistically significant differences are marked with an asterisk.

4.4 Discussion

4.4.1 Music and Immersion

The first study found a significant difference in favor of the music condition of the *Radical Tunes* increasing *overall IEQ scores*, *cognitive involvement*, and *emotional involvement*, but found no significant differences for real world dissociation, challenge, or control (see Table 4.1). I expected challenge and control to be similar across the two groups since the game controls and tasks were identical. Overall, this indicates that adding music to *Radical Tunes*—and likely SLA games in general—can improve player factors (cognitive involvement and emotional involvement) but not game factors (challenge and control) of immersion [98]. Furthermore, these findings fall in line with existing studies showing increased immersion with the addition of pleasing music [171].

4.4.2 The Meaning Mystery

The second study was meant to explore whether music could help people remember the stroke order and direction of several Japanese kanji. Yet, I found that the treatment group performed better on the meaning tests as well. I have a possible explanation: During early testing of my app I found that people tended not to look at the meaning of the character while going through the learning section of the app, and would consequently not learn any of the meanings even though they learned to write the characters. To help with this, I added a recorded spoken meaning that played each time after the last stroke of a character was drawn during the learning phase. For the treatment group, that meant that the spoken meaning was always attached to the unique melody assigned to that character, while the control group heard the meaning attached to a new melody every time. A study by Wallace et.al. showed that accompanying text with a repeated melody helped participants memorize text better than when a variety of melodies was used [196]. I hypothesise that the treatment group benefited from the same effect when it came to the spoken meaning.

4.4.3 Scalability

While there are thousands upon thousands of logographic characters, they all consist of a finite number of radicals (around 80) and even fewer stroke types. It is therefore feasible for a human musician to compose unique melodic elements either for the individual strokes or radicals. An algorithm could then combine them (with pitch adjustment, where needed) based on the composition of each kanji to generate a unique melody for it. As a result, this work presents the groundwork for design and creation of a scalable approach towards creating music for an otherwise intractable number of characters.

4.4.4 Implications for Design

The study results illustrate that music has the potential to serve as a useful mnemonic tool in stroke knowledge acquisition, with increases to learner immersion and perceptions of usefulness. This provides initial evidence for the value of incorporating consistent melodies into the design of logographic learning software. The results also showed that the mnemonic effect was more pronounced in characters with a greater number of strokes. This suggests that a musical mnemonic design is particularly useful for learning more complex logographic characters, while other memorization techniques, such as repetition, may serve learning simple characters well enough. This is not to say that a musical mnemonic design approach is not beneficial at an introductory level, however, since even among the characters that are taught early on there are some with 9 strokes and more (e.g., 前 "before," which is part of the word "name," has 9 strokes, and 時 "time" has 10). Therefore, the musical mnemonic isn't something useful only to advanced learners.

4.5 Limitations and Future Works

The two primary limitations of this study were with the chosen platform for participant recruitment, and the music. Amazon's Mechanical Turk is a platform where one can hire workers to perform remote tasks for a set reward. The workers are incentivised to complete tasks as quickly as possible to maximize their profit, and that sometimes leads to sloppy performance. The evidence for this is some participants' unwillingness to use the "undo" button to correct a clearly accidentally misplaced stroke (see the white stroke on the lower right of Fig. 4.7). Because I wasn't able to observe the participants, I can not be sure of the diligence they applied to the task.

As for the second limitation, I used procedurally produced sound in my app to create a responsive musical experience. Without a designated sound generation expert on the team, I was limited in the variety of sounds I could generate. While pre-recorded sound could offer infinite variety, I lacked the expertise to make it play in accordance to the speed and direction of the stroke, the way I could with procedurally produced sound. Additionally, I should have taken more care in designing melodies that were clearly distinct from each other. If, as some informal feedback suggested, many of the melodies sounded too similar, that would certainly dampen the music's mnemonic effect. Future work could explore different approaches to music creation and the effects they have on the melody's memorability.

In this study I made a point of using kanji consisting of a small subset of radicals. Future work could explore whether music is equally effective for memorizing characters with no common elements.

4.6 Conclusion

In this chapter, I described *Radical Tunes*—a musical game for learning to write Japanese kanji, which uses unique melodies to aid with stroke order memorization. I also described the results of two studies. One compared the musical version of *Radical Tunes* with a version which instead had non-musical sound effects (the sound of chalk writing on a blackboard). The second compared a version of *Radical Tunes* with consistent melodies to a version with procedurally generated ever-changing melodies.

Through use of the *Immersive Experience Questionnaire*, I found that incorporation of music in *Radical Tunes* significantly increased player immersion. The participants from both conditions showed significant score improvement between pre- and post-tests, which shows that *Radical Tunes* was effective in teaching the kanji to players, at least in the short-term. The small scope of the pilot study—only six characters—didn't expose any mnemonic effects the music may have on players' ability to retain the kanji long term. However, the follow up study showed that exposure to consistent melodies lead to significantly higher post-test results, particularly in kanji with a higher number of strokes. I believe that if designed and implemented well, the ideas described in this chapter could make a positive difference for the millions of novice learners of LWS.

Table 4.5: Descriptive statistics of post-test stroke knowledge scores.

Post-test Stroke Knowledge Scores	*Control* μ	*Control* σ	*Treatment* μ	*Treatment* σ	*Significance*
十 Order	0.64	0.49	0.55	0.51	0.58
十 Direction	0.93	0.23	1	0	0.2
口 Order	0.8	0.39	0.93	0.21	0.19
口 Direction	0.79	0.39	0.93	0.21	0.15
氵 Order	0.86	0.35	0.9	0.31	0.72
氵 Direction	0.76	0.39	0.87	0.33	0.34
日 Order	0.89	0.3	0.83	0.36	0.55
日 Direction	0.89	0.3	0.83	0.36	0.55
月 Order	0.96	0.21	0.95	0.22	0.95
月 Direction	0.92	0.24	0.9	0.24	0.78
木 Order	0.77	0.37	0.93	0.25	0.13
木 Direction	0.86	0.32	0.96	0.12	0.2
古 Order	0.71	0.38	0.82	0.33	0.32
古 Direction	0.82	0.33	0.86	0.33	0.68
回 Order	0.58	0.45	0.61	0.4	0.85
回 Direction	0.6	0.44	0.64	0.41	0.75
困 Order	0.53	0.43	0.56	0.38	0.82
困 Direction	0.59	0.42	0.62	0.39	0.81
汨 Order	0.46	0.43	0.7	0.38	0.06
汨 Direction	0.44	0.41	0.69	0.37	0.04
沐 Order	0.6	0.35	0.77	0.29	0.09
沐 Direction	0.56	0.35	0.81	0.3	0.02
明 Order	0.75	0.37	0.79	0.36	0.7
明 Direction	0.76	0.32	0.79	0.32	0.7
沽 Order	0.3	0.38	0.58	0.44	0.03
沽 Direction	0.39	0.4	0.63	0.44	0.08
枯 Order	0.59	0.39	0.78	0.31	0.09
枯 Direction	0.65	0.33	0.82	0.29	0.07
胡 Order	0.35	0.34	0.64	0.4	0.02
胡 Direction	0.39	0.36	0.67	0.35	0.02

Table 4.6: Descriptive statistics of longitudinal post-test stroke knowledge scores.

Longitudinal test	*Control*		*Treatment*		*Significance*
Stroke Knowledge Scores	μ	σ	μ	σ	
十 Order	0.77	0.43	0.65	0.49	0.39
十 Direction	0.96	0.21	1	0	0.35
口 Order	0.86	0.35	0.95	0.16	0.32
口 Direction	0.85	0.35	0.9	0.22	0.58
氵 Order	0.82	0.4	0.95	0.22	0.2
氵 Direction	0.74	0.41	0.95	0.22	0.05
日 Order	0.81	0.38	0.95	0.22	0.15
日 Direction	0.8	0.38	0.94	0.23	0.15
月 Order	0.86	0.32	1	0	0.06
月 Direction	0.82	0.33	0.94	0.11	0.13
木 Order	0.82	0.33	0.95	0.15	0.11
木 Direction	0.9	0.25	0.99	0.06	0.13
古 Order	0.75	0.38	0.89	0.2	0.13
古 Direction	0.79	0.37	0.95	0.16	0.08
回 Order	0.62	0.39	0.72	0.39	0.43
回 Direction	0.63	0.36	0.74	0.39	0.34
困 Order	0.58	0.45	0.59	0.43	0.99
困 Direction	0.63	0.44	0.69	0.37	0.65
泪 Order	0.62	0.41	0.72	0.39	0.4
泪 Direction	0.66	0.37	0.71	0.38	0.62
沐 Order	0.75	0.34	0.83	0.22	0.37
沐 Direction	0.75	0.33	0.92	0.18	0.04
明 Order	0.77	0.37	0.89	0.22	0.21
明 Direction	0.74	0.36	0.86	0.23	0.24
沽 Order	0.61	0.39	0.68	0.4	0.59
沽 Direction	0.64	0.38	0.8	0.34	0.15
枯 Order	0.63	0.39	0.62	0.37	0.94
枯 Direction	0.76	0.32	0.73	0.35	0.78
胡 Order	0.4	0.4	0.58	0.42	0.15
胡 Direction	0.42	0.43	0.67	0.37	0.05

Chapter 5

Zen Hanzi (怎汉字): A Component-Focused Game for Learning Hanzi

5.1 Game Overview

Zen Hanzi is an experimental component-focused game for learning traditional Chinese characters, i.e., hanzi. The title of the game is a play on the word "Zen", which—in addition to the western adaptation of the Japanese term derived from the Chinese word chan (meaning "meditation")—is a homonym of the Chinese word for "how," making the title into "how to hanzi". It was inspired by research claiming that component awareness is beneficial to learning hanzi/kanji [60], and the evidence that recognition practice isn't usually sufficient to memorize hanzi well enough to be able to produce them [163]. In *Zen Hanzi*, the player must assemble a given hanzi by selecting

Figure 5.1: *Zen Hanzi* screenshot. The player must drag the components from three rows at the bottom into one of the four structures above to assemble the hanzi indicated by the English translation and pinyin pronunciation at the top of the screen.

correct components and placing them into correct slots in a correct hanzi structure (see Figure 5.1). In addition to drawing the player's awareness to the component-based nature of the characters, *Zen Hanzi* also makes sure to include look-alike options for each correct component, ensuring the students are aware of some of the possible mistakes they could make when writing (see Figure 5.2).

In terms of presentation of educational content, *Zen Hanzi* has a "Learn"

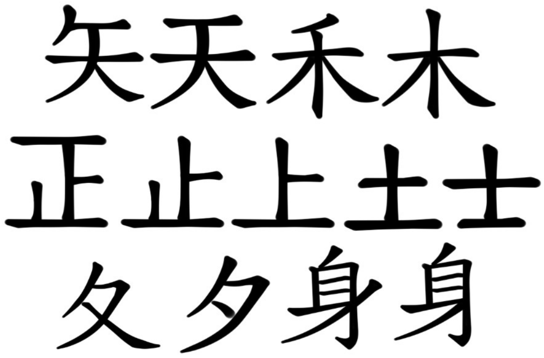


Figure 5.2: Some examples of look-alike components. Some differ by a single extra stroke, some by comparative length of two strokes, some by whether a stroke crosses another.

and a “Review” mode. In “Learn” mode, the player is shown a picture of the hanzi before assembling it (see Figure 5.3), while in review mode they are guided only by the English meaning and Chinese pronunciation of the hanzi—requiring players to recall the structure of the hanzi without guidance. Additionally, the “Hint” button allows the player to view the hanzi one more time if they are stuck. The game also implements

Figure 5.3: In the learn mode, the player is shown the hanzi before they attempt to assemble it.

a number of assistance features to help players keep track of options they previously tried, such as by changing the color of the components after each attempt based on their correctness (see Figure 5.4). Finally, the game has scoring and "lives" mechanics. While those aren't universally effective motivators, they are known to provide motivation for some of the more competitive players [25].

5.1.1 Rationale

Acquiring and maintaining working knowledge of hundreds of hanzi/kanji needed for LWS literacy requires frequent exposure to the characters, or, barring that, frequent reviews and practice. This can often be a monotonous, tedious task, and many students rapidly lose their reading and writing proficiency after they are done taking courses [164]. While there are many learning and reviewing apps that offer structure and even written practice, the persisting target language literacy deficiency of LWS students indicates that the current technology isn't quite enough to help them gain and retain hanzi/kanji knowledge [206]. The apps often lack sufficient motivational elements, and the existing games predominantly use exogenous flashcard-style multiple-choice exercises disconnected from the game content and offer no component emphasis. The challenge of designing a good educational game lies in the necessity of three things: effective educational content, an enjoyable/motivating game, and a seamless connection between the two [48, 200]. As a result, I researched methods for learning hanzi, educational game design, and Free-to-Play (F2P) mobile game design. I came up with the following principles, which I applied when designing *Zen Hanzi*.

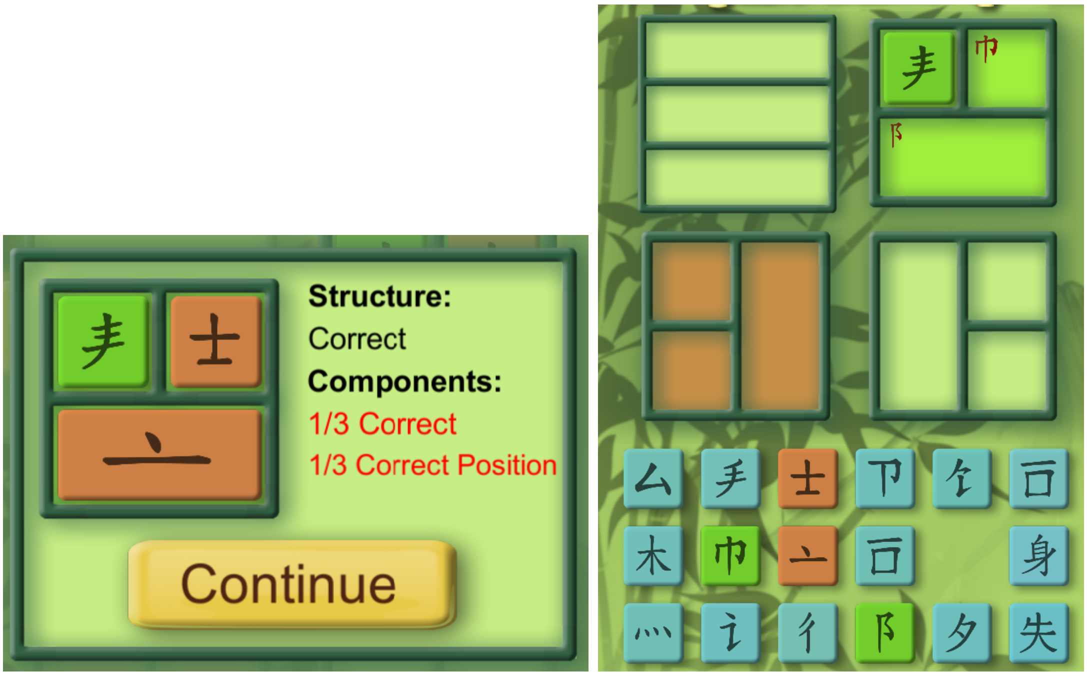


Figure 5.4: The player receives feedback on their attempt, including whether they used the correct structure, how many components are correct, and whether any of the components were in the correct position. The game helps the player remember which options they tried by coloring incorrect components and structures red, and correct ones green. Once the correct structure is identified, the game also marks which components were incorrectly tried in each slot.

5.1.2 *Zen Hanzi* Design Principles

1. Intrinsic pleasure is a must [131]. This can be aided by appealing graphics/sounds and plentiful feedback to player's actions with visual and audio effects [103, 84, 177, 106, 34, 139, 47], as well as designing the game to be playful overall [153].

2. The game must have short rounds in order to fit into micro-leisure intervals of contemporary life [122].

3. The learning content must be endogenously integrated into the game [77, 200, 154, 174]. That is to say, no *Math Blaster*-like "add two numbers to shoot at asteroid" [12]. The core game mechanic must be inseparable from the learning content [152, 151].

4. The game must avoid orthogonal mechanics [193]. Orthogonal mechanics are those that detract from the primary purpose of a serious game, in this case – learning LWS. An example would be requiring the player to become good at aiming and shooting in order to access the learning content. Every action should serve to reinforce or increase knowledge.

5. The game must be scalable in terms of authoring burden. Hand-authoring separate exercises for thousands of characters to support learners of different skill levels would make the game production logistically infeasible.

6. The game must raise students' hanzi component awareness as developing component awareness at the beginning of a student's hanzi learning quest can aid them

```
"想":{
"pinyin":"xiǎng",
"hanzi":"想",
"pinyin_sound":"xiang3",
"english":"to want",
"english_sound":"to_want",
"layout":{"type":"s_H","part_1":"s_V",
"part_2":"s_none"},
"components":[
    {"name":"113","position":0},
    {"name":"208","position":1},
    {"name":"115","position":2}]}
```

Figure 5.5: An example of one hanzi encoding

in developing their writing skills [164] and has a positive impact on a student's ability to memorize characters [60].

5.1.3 Designing for Content Scalability

Zen Hanzi's design would be very impractical if one had to hand-author levels for each hanzi, selecting look-alike components, and laying out hanzi structure frames. To get around this, I designed a succinct way to encode all the necessary information about each hanzi, including it's structure and composition (Fig. 5.5). I also created my own mini-database of hanzi components and their look-alikes. This way, the game can randomly include at least one look-alike for each component within the hanzi.

Much of the information needed to encode any hanzi can be scraped from free

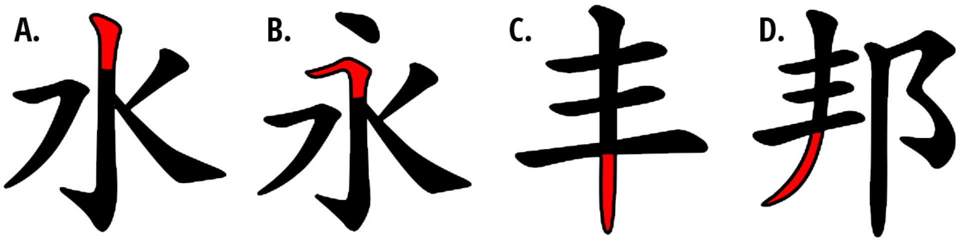


Figure 5.6: Component variations with differences highlighted in red. The 'water' component (A) is written with a vertical line at the top in some of the hanzi (B). It's still considered the same component according to databases I examined, but it would be incorrect to omit that line or add it where it isn't needed. Similarly, the 'great' component (C) is occasionally written in a slanted form (D). In *Zen Hanzi*, I include both versions of each component to draw the player's attention to the possible mistakes they could make.

databases online, such as the Chinese Character Decomposition[1]. The sounds can also be obtained from free online datasets, like the Mandarin Voice Soundset[2].

Things get trickier when it comes to the component database. While some lists of look-alike characters and components can be found online, they typically don't address component variations, as in Figure 5.6, or include incorrect versions of components. The latter ones are useful in highlighting mistakes a novice learner might make.

I used Fordham's list of 800 most common components [64] as a foundation for my own component database (Fig. 5.7). However, I only used the basic components that couldn't be easily broken up into smaller components, and added some of my own,

[1] `https://commons.wikimedia.org/wiki/Commons:Chinese\_characters\_decomposition`
[2] `http://www.chinese-lessons.com/`

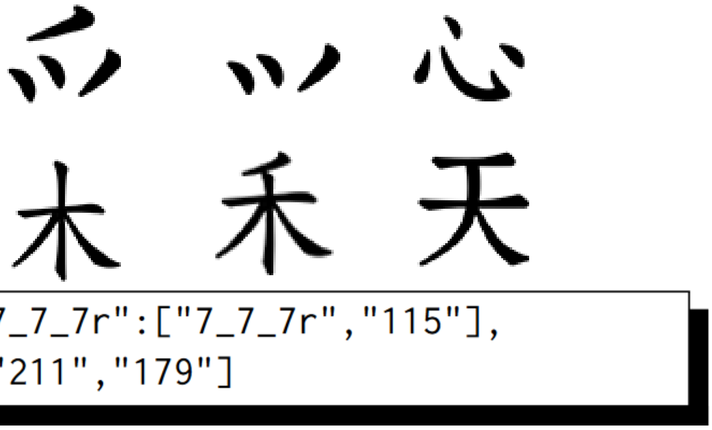


Figure 5.7: Two entries from my component database. The principal component is on the left. The name is composed of numbers in Fordham's list, Chinese names for specific strokes, horizontal or vertical division of the component, and a suffix 'r,' indicating the component is reversed left to right. On the right are the names of other components within the database, that can be considered look-alikes for this component.

to account both for slight but important variations in some components; and some of the common wrong versions of components.

With all this in mind, the bulk of work in making content scalabe in future games/designs would be in the component database as it has to be assembled largely from scratch—as I had to for *Zen Hanzi*. Notably, the hanzi database would also have to be checked to replace components with their variants where necessary.

5.1.4 Unconventional Hanzi Component Design

After designing and prototyping *Zen Hanzi*, I showed it to three domain experts (i.e., Chinese language professors) to receive feedback on the design. All of them

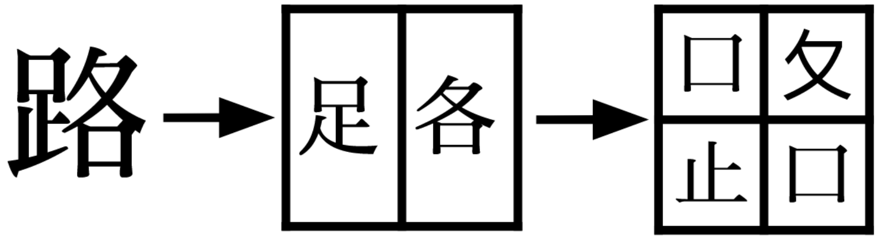


Figure 5.8: The conventional way of hanzi structure classification would identify this hanzi as a vertically-split two-part hanzi. However, each of those hanzi parts consist of two basic components, each carrying its own potential for a mistake. It's also worth noting that division into two parts in this case results in the transformation of the lower-left component into a different variant of itself, which may confuse the student about the proper shape within the original hanzi. Because of this, *Zen Hanzi* breaks this hanzi into four parts.

expressed their surprise over how I divided hanzi into components (Figure 5.8). Rather than using the conventional hanzi structure classification methods which split hanzi into 1-3 major parts, I broke them down into their more basic components[3]. This was done as each small component holds its own potential for errors and is therefore worthy of attention. Additionally, if I used complex components, my component database would have to hold exponentially more components in order to achieve the same granularity of possible errors and their combinations. I believe this would make both the design and scaling of the component database logistically infeasible. It would also likely present the student with a less effective learning experience, since they would have fewer incorrect

[3]There is no one comprehensive list of all hanzi components. I used Fordham's list [64] and my best judgement to identify which parts of the hanzi should be considered as independent components for my purposes.

options to eliminate at each stage.

5.1.5 What *Zen Hanzi* is and isn't

Zen Hanzi isn't meant as a stand-alone method of learning Chinese. While it teaches pronunciation and meaning of included hanzi, it is a supplemental activity intended specifically to minimize component substitution mistakes in writing. It is meant to informally supplement formal classroom learning by deepening a student's awareness of the components within each hanzi. Those interested in learning the language as a whole would need to learn sentence structure and nuances of using hanzi in various combinations elsewhere. Furthermore, *Zen Hanzi* is intended to be a supplemental informal learning activity that students might engage in on their own time, thus avoiding the hurdle of classroom integration.

5.2 Methodology

For this study, I wanted to see how the highly granular component-focused design of *Zen Hanzi* compared to the whole-hanzi approach of typical existing Chinese as a Foreign Language (CFL) learning tools with respect to players' learning of hanzi component composition and hanzi meaning. To this end, I performed a comparative between-subjects study of *Zen Hanzi* and *Quizlet*, a popular flashcard app frequently used in Chinese courses.

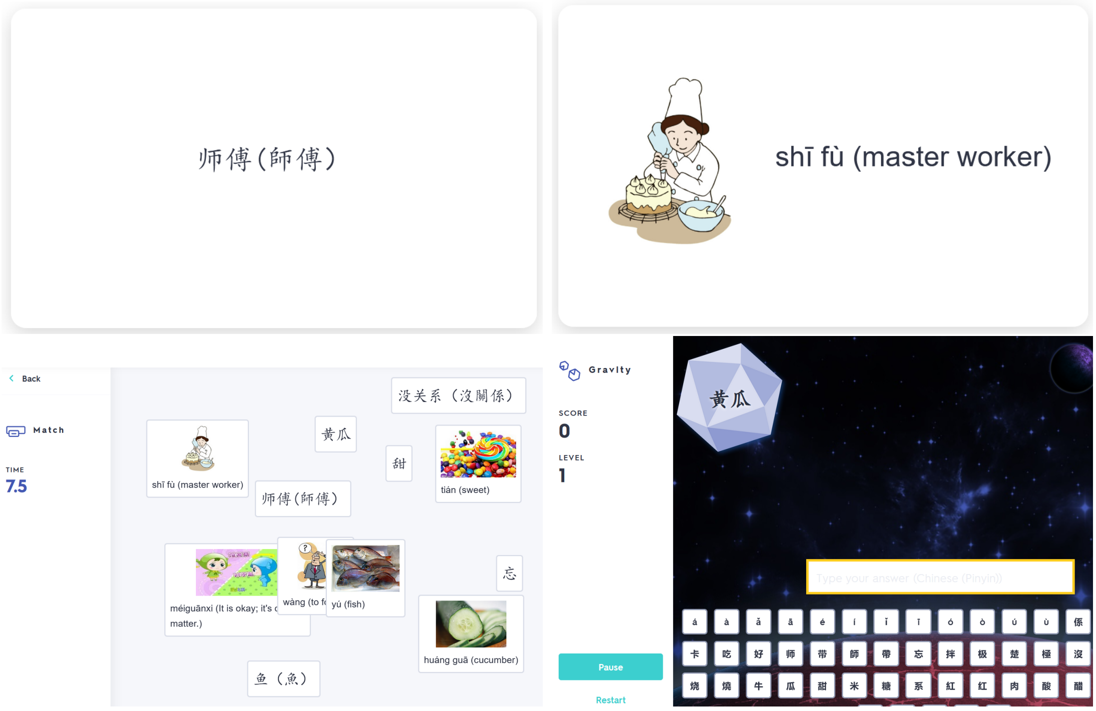


Figure 5.9: Quizlet screenshots. On the top is a typical two-sided *Quizlet* flashcard from a Chinese 1 course. On the lower left is *Quizlet*'s timed matching game, on the right is their Asteroid game, where entering the corresponding option destroys the asteroids speeding towards Earth.

5.2.1 Quizlet

Quizlet is a web-based digital flashcard software. I selected *Quizlet* for the control group because several Chinese professors I spoke to in multiple universities use it to assign hanzi and vocabulary homework for their students. It also takes the whole-hanzi design approach that is common in most CFL learning apps where players are presented with the entire hanzi to memorize/learn. *Quizlet* offers its users an ability to create their own digital flashcard decks and several ways to study them, including quizzes and two gamified options (Figure 5.9, bottom). A typical flashcard from a Chinese 1 course includes the hanzi or a multi-hanzi word, its pinyin pronunciation, and an English translation (Figure 5.9, top). The students can also hear both the English and Chinese words provided via *Quizlet*'s built in sound database.

5.2.2 Hypotheses

I set out to investigate the educational and experiential aspects of *Zen Hanzi* and took an exploratory approach with the following research question:

- **Research Question**: Are there notable differences between user experience perception and learning outcomes for the component-focused design of *Zen Hanzi* and whole-hanzi design of current popular learning apps such as *Quizlet*?

To that end, I identified the following hypotheses:

- **H1.1**: *Zen Hanzi*'s component-focused gameplay will result in higher scores on hanzi component composition post-test, as compared to the whole-hanzi approach of existing hanzi learning tools.

- **H1.2**: since the approach to teaching hanzi meaning is roughly the same between *Zen Hanzi* and *Quizlet* (i.e., both provide written and audio pronunciation of hanzi meaning), they will have comparable effects on the hanzi meaning post-test scores.

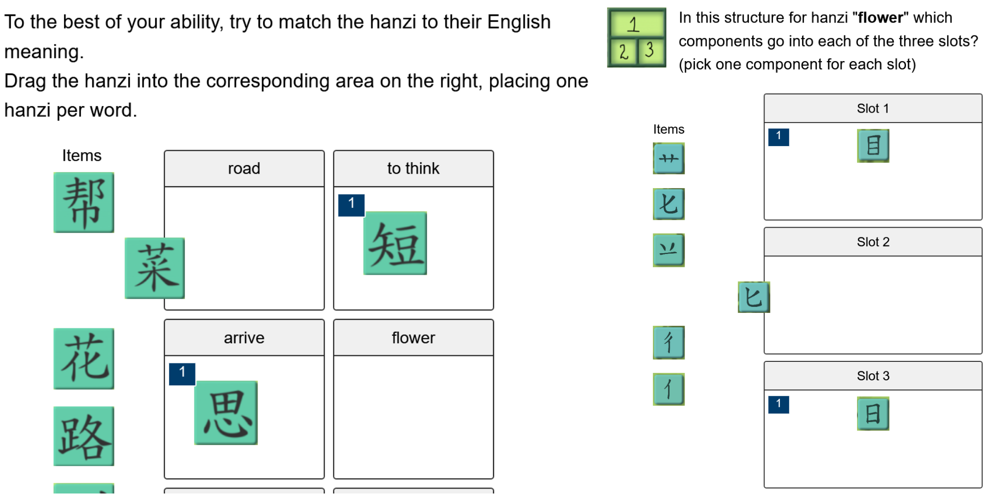


Figure 5.10: *Qualtrics* pre- and post-test screenshots. On the left is the hanzi meaning test, on the right is the individual hanzi components test.

5.2.3 Participants

A convenience sample of participants were recruited for this study by online announcements in large language-learning groups and Chinese drama fan-clubs on *Discord*, *Facebook* and *Tumblr*; as well as by word of mouth. In total, 73 people completed the study. Of those, 10 participants were excluded from the final analysis, as they received 25 or more points out of 29 on the pre-test. This indicated a high level of prior hanzi knowledge, which in turn meant they didn't have much room for improvement. Of the remaining participants, there were 43 female, 13 non-binary, six male and one of undisclosed gender. The ages ranged from 18 to 74, with the average age of 40.2 and median age of 30. The participants were also asked to self-report their reading and writing proficiency in any languages that use hanzi or hanzi-derived written systems.

The participants were randomly distributed into two conditions (*Zen Hanzi* treatment group or *Quizlet* control condition) with the final count of 27 people in the control group and 36 in the treatment group[4]. There were no statistically significant differences in age, gender or prior knowledge distribution between the two groups, with the exception of there being 10 non-binary participants in the treatment group vs. only three in the control group. The participants were not paid or compensated for their participation in any way.

5.2.4 Procedure

The participants were provided with digital instructions and links to the five steps of the experiment. First, they were shown a short video tutorial explaining how to use the survey platform, *Qualtrics*[5], and how the drag-and-drop questions within it work. Second, they were directed to a *Qualtrics* survey where they consented to the study, filled out basic demographic information and took a hanzi pre-test. This test consisted of matching 10 hanzi with their English meaning, and identifying correct individual components and their placement in each of the 10 hanzi (see Fig. 5.10).

For the third step, they were provided a 2-minute YouTube video tutorial of the app they were about to use — *Zen Hanzi* for the treatment group and *Quizlet* for the control group. The participants were then directed to use their assigned app for 10 minutes to study the 10 hanzi. For *Quizlet*, this meant that participants could use either digital flash cards or the two gamified quizzes to study (see Figure 5.9). As the final

[4] I slightly increased the odds of a participant being assigned to the treatment group, as I wanted to collect as much experiential feedback on *Zen Hanzi* as possible.

[5] `https://www.qualtrics.com`

Table 5.1: Questions Used in Experience Analysis

Questionnaire	Subscales Used	Modifications
PENS v1.6	Competence Intuitive Controls	I replaced the word "game" with "app," and "play" with "use," so the questions made sense both in the context of *Zen Hanzi* and *Quizlet*. I chose not to use the Immersion, Relatedness, and Autonomy subscales, because the questions were geared toward story-driven or multi-player games.
GEQ-R	Positive Affect Negative Affect Competence Flow	I chose not to use the Immersion subscale from GEQ-R, because it was geared toward a story-driven game (i.e. "I was interested in the game's story.")
SI	Exploration Intention Instant Enjoyment Attention Quality Challenge Novelty	None

step, upon completion of the 10 minute study session, the participants were once again asked to complete a *Qualtrics* survey. This survey consisted of a post-test identical to the pre-test, followed by a 48-question survey consisting of select questions from Player Experience Needs Satisfaction (PENS) [159], Situational Interest (SI) [20], and modified Game Experience Questionnaire (GEQ) [102].

Table 5.2: Hanzi Test Scores and significance values from ANCOVA with pre-test scores as covariates

Test	Treatment (avg / std)	Control (avg/std)	p	95% CI
Meaning pre-test	3.14 / 3.16	3 / 3.37	0.86	
Component pre-test	8.78 / 6.97	7.19 / 6.39	0.28	
Meaning post-test	4.33 / 3.36	8.48 / 2.72	0.17	-2.64 / 0.06
Component post-test	21.06 / 5.52	15.81 / 5.99	**0.0006**	1.97 / 6.89

Table 5.3: Hanzi Test Scores and significance values from ANCOVA with pre-test scores as covariates

Test	Treatment (avg / std)	Control (avg/std)	p	effect-size g
Meaning pre-test	3.14 / 3.16	3 / 3.37	0.86	
Component pre-test	8.78 / 6.97	7.19 / 6.39	0.28	
Meaning post-test	4.33 / 3.36	8.48 / 2.72	0.17	
Component post-test	21.06 / 5.52	15.81 / 5.99	**0.0006**	0.91

Table 5.4: Experience Evaluation results and significance values from ANCOVA with pre-test scores as covariates

Subcategory	Treatment (avg/std)	Control (avg/std)	p	95% CI
PENS Competence	11.67 / 3.54	13.07 / 2.7	0.08	-3.21 / 0.14
PENS Controls	15.11 / 2.5	14.41 / 2.24	0.25	-0.46 / 2.03
SI Exploration	13.92 / 5.32	13.96 / 5.49	0.97	-3 / 2.66
SI Enjoyment	19.69 / 5.86	20.26 / 5.47	0.7	-3.65 / 2.32
SI Attention	18.28 / 3.57	18.44 / 4.25	0.87	-1.99 / 2.02
SI Challenge	6.94 / 4.16	8.26 / 4.75	0.25	-3.29 / 1.07
SI Novelty	13.14 / 2.92	11.19 / 4.23	**0.03**	0.062 / 3.72
GEQ Positive Aspect	20.97 / 6.02	21.59 / 5.14	0.67	-3.48 / 2.45
GEQ Negative Aspect	12.56 / 7.52	10.89 / 4.71	0.32	-1.26 / 5.22
GEQ Competence	14.44 / 4.38	16.67 / 3.89	**0.04**	-4.53 / 0.53
GEQ Flow	18.44 / 5.59	18.07 / 7.31	0.82	-2.38 / 4.06

Table 5.5: Experience Evaluation results and significance values from ANCOVA with pre-test scores as covariates

Subcategory	Treatment (avg/std)	Control (avg/std)	p	effect-size g
PENS Competence	11.67 / 3.54	13.07 / 2.7	0.08	-
PENS Controls	15.11 / 2.5	14.41 / 2.24	0.25	-
SI Exploration	13.92 / 5.32	13.96 / 5.49	0.97	-
SI Enjoyment	19.69 / 5.86	20.26 / 5.47	0.7	-
SI Attention	18.28 / 3.57	18.44 / 4.25	0.87	-
SI Challenge	6.94 / 4.16	8.26 / 4.75	0.25	-
SI Novelty	13.14 / 2.92	11.19 / 4.23	**0.03**	0.55
GEQ Positive Aspect	20.97 / 6.02	21.59 / 5.14	0.67	-
GEQ Negative Aspect	12.56 / 7.52	10.89 / 4.71	0.32	-
GEQ Competence	14.44 / 4.38	16.67 / 3.89	**0.04**	0.53
GEQ Flow	18.44 / 5.59	18.07 / 7.31	0.82	-

5.2.5 Educational Content

Both *Zen Hanzi* and the *Quizlet* deck provided to the participants included the following ten hanzi: 到, 路, 帮, 菜, 花, 想, 思, 对, 短, 谢. These are all taught in the first year Chinese courses at a large university and were chosen for their multi-component structure.

5.2.6 Measurements

I scored the participants on the number of hanzi they correctly matched to their English meaning (max score of 10), as well as the total number of components correctly placed into the hanzi structures (max score of 29). I tested them both before and after the treatment, to be able to use pre-test scores as a covariate in analysis.

To evaluate how *Zen Hanzi* measures up to *Quizlet* experience-wise, I used 48 questions from three game/activity evaluation questionnaires — GEQ-R, SI, and PENS — to assess user experience. The list of subscales used from each of the questionnaires and minor modifications I made (if applicable) are in Table 5.1.

5.3 Results

Using the pre-test meaning and component scores as covariates to control for individual differences in prior hanzi knowledge, I ran a series of ANCOVA tests to examine the two post-test knowledge result sets (i.e., hanzi meaning and component composition) and the 11 experience questionnaire subscale result sets.

The hanzi meaning and component test scores are listed in Table 5.3. The treatment group displayed significantly higher post-tests scores for correct placement of individual components ($p < .001$). The pre-test component scores were also significant predictors of the post-test component scores with $p = 0.002$. I found no other significant covariate interactions.

The experience evaluation results are listed in Table 5.5. Notably, I found significantly higher scores with respect to novelty for *Zen Hanzi* ($p = .03$), and significantly higher scores with respect to GEQ competence for *Quizlet* ($p = .04$). Although not quite significant, PENS competency scores trended similarly to the GEQ competence scores ($p = .08$). I discuss these results further in the Discussion.

5.4 Discussion

5.4.1 Component-focused vs. Whole-hanzi Design

The treatment group did significantly better on the individual components post-test, thus confirming the first hypothesis **H1.1**. This demonstrates that *Zen Hanzi* was better at helping students recall details of each hanzi, rather than only their approximate shape—at least in the short term and with visual cues. The second hypothesis, **H1.2**, was similarly confirmed as there was no significant difference in post-test meaning scores between the two groups. This indicates that *Zen Hanzi* was not significantly worse or better than current popular hanzi learning apps such as *Quizlet* for teaching hanzi meaning.

In an informal feedback I received from some of the participants, some reported feeling frustrated at having to differentiate between components that differed by a single stroke or even relative length of a single stroke (some examples are in Figure 5.2). Unfortunately, this frustration is shared by any student tackling an LWS. The lack of any such feedback from *Quizlet* users only highlights the fact that those participants weren't as aware of the shape of components within the hanzi. For example, some explained to me that they would memorize one or two components within the hanzi and then just look for that component to identify the hanzi, without bothering to give any attention to the other components. That is a bad strategy, as there are many hanzi that differ by a single component, or even have the exact same components arranged in a different way.

I suggest that hanzi and kanji learning app and game creators could include *Zen Hanzi*'s fairly simple to implement mechanic in their products to help their users learn the characters on a more than just superficial level.

5.4.2 Novelty

On the self-reported experiential side, novelty was significantly higher for *Zen Hanzi* while competence was significantly higher for *Quizlet*, leaving the exploratory research question about user experience without a definitive answer. With respect to higher perceived novelty scores for situational interest, this is a good sign, as situational interest is critical to the design of education tools (and education overall) since it is essential for developing personal interest in learners which can in turn positively impact learning [157, 85, 140]. I hypothesize that the novelty of *Zen Hanzi* may be due to the novelty of the component-focused approach to learning. Many learners study for a variety of topics, including English vocabulary, using flash cards and similar forms of SRS tools. The logographic nature of hanzi makes learning a word and meaning more complex, as shape becomes a factor of memorization—an area of emphasis in *Zen Hanzi's* design. Therefore, tools that focus on learning components rather than whole words/whole-hanzi are much less commonly encountered by most learners.

5.4.3 Competence

It would be worthwhile to further investigate the users' reported feelings of competence, which trended lower in the treatment group. A person's feeling of competence plays an important part in their motivation to persist with a task [55]. Adopters of this design should be aware of the possible competence penalty that comes with increased awareness of hanzi components, at least at the early stages of learning. I speculate that having to memorize each individual component in a hanzi is more difficult than merely remembering the approximate shape of a hanzi. As a result, it's possible that *Zen Hanzi* players got the answer wrong more frequently than the *Quizlet* users, and therefore felt less competent.

Overall, the results highlight some clear advantages of a component-focused design over traditional whole-hanzi designs, as well as provide some areas that need careful consideration (i.e., improving feelings of competency) in the design of hanzi learning tools.

5.5 Limitations and future work

5.5.1 Design Limitations

While scalable by design, in its current form *Zen Hanzi* supports only a subset of possible hanzi structures, so it does not include every possible hanzi at this time. Additionally, many Chinese words are composed of multiple hanzi, so *Zen Hanzi* could be further improved by an ability to practice multi-hanzi words. Furthermore, the images for each component currently need to be drawn by hand. However, there are a finite number of basic components, and once the assets for all of them exist, adding new hanzi with the scalable component and hanzi databases is relatively easy. Finally, I was disappointed to see that *Zen Hanzi* players didn't report higher enjoyment than *Quizlet* users. The game would benefit from further user testing and design adjustments to facilitate learner engagement and habit formation. For instance, adding more "juicy" feedback may help to increase player feelings of enjoyment, engagement, and competence [103, 84, 177].

5.5.2 Evaluation Limitations

There are some notable limitations to the study presented here, i.e., a small sample size that might not be fully representative of CFL learners. However, on a small scale, *Zen Hanzi* demonstrated notable improvement in short-term memorization of hanzi component structure over the widely used *Quizlet* app. These promising results warrant further research with a larger sample size, expanded learning content, longitudinal testing, and testing of participants' ability to produce hanzi unassisted. This will provide a better understanding of the long-term effects that a component-focused design might have in comparison to whole hanzi methods. Future work should also investigate the possibility of lowered feelings of competence and ways to counteract that effect—as feelings of competence have been shown to be important for the efficacy of educational games [76, 169].

5.6 Conclusion

In this section I described my experimental hanzi learning game, along with evaluations of its educational efficacy and user experience as compared to a popular flashcard app. The game's component-focused design resulted in participants achieving higher scores on the immediate post-test of hanzi component structure.

I discussed the possible drawbacks for the beginner students, and other aspects of the game that bear further design improvement and investigation.

I shared both the visual design and the underlying data structures. I hope my design has the potential to serve as a useful free resource for students of LWS in any stage of their education, as well as to serve as a reference for the design of future LWS learning tools.

Chapter 6

Discussion

When I discuss my research, the first reaction I often face is doubt that something so "evil" could be used in a positive way. While it would be outside of the scope of this work to provide a detailed cross-examination of each of the game mechanics described in this dissertation, it is my hope that a thorough analysis of two of the most notorious and pervasive F2P mechanics will provide a compelling argument in favor of my chosen topic.

6.1 Variable Rate Reinforcement, Gacha, and the Loot Box Controversy

Among all of the engagement and monetization techniques I've discussed, loot boxes are arguably the most controversial in-game mechanic [126], and thus warrant a closer look.

In the following section, I delve into the history of loot boxes; the psychological underpinnings that made them such an effective engagement and monetization mechanic in digital games; and the ensuing controversy around their use. Next, for a more well-rounded understanding on the subject, I look into loot boxes from the perspective of game makers and players. Finally, I share my ideas on potential uses of this mechanic to encourage learning through digital games.

Figure 6.1: Gachapon: modern day toy capsule vending machines outside a game arcade in the Namba entertainment district of Osaka, Japan. Some are featuring a set of Harry Potter-themed necklaces [46].

The UK Parliament defines Loot Boxes as “items in video games that may be bought for real-world money, but which provide players with a randomised reward of uncertain value" [201]. The suspected first instance of loot boxes in a game was the 2003 South Korean MMORPG, MapleStory [126]. In the game, the players could spend real money to purchase an item called "gachapon ticket," which could be exchanged for a random in-game item. Gachapon (or gashapon) are a type of a mystery toy capsule vending machine that originated in Japan in 1965 [72] (Fig. 6.1). With time, the concept expanded to include any sort of blind or mystery box toys and items, called gacha. When the concept migrated into the world of digital games, it became known as the gacha mechanic. Henceforth, gacha and loot boxes will be used interchangeably. The defining feature of gacha is that they always come in a thematically connected set, be it a brood of cute kittens or a roster of characters from a popular anime or film. Collecting each item in a set, or trying to get a specific toy can be an exciting, frustrating and expensive experience, since the customer pays for the box without knowing its contents, and often some items are more rare than others. With bated breath, the customer opens a box, tears open a blister pack, or glimpses the clear capsule dispensed from a gachapon, and then comes either elation or disappointment. This brings us to the variable rate reinforcement (VRR).

Variable rate reinforcement or variable rate reward is a concept applicable to elicitation of behavioral changes, such as, for instance, training a dog to sit on command [62]. If one gives a dog a treat every time she sits on command, they are rewarding the desired behavior and the dog learns the "sitting on command equals reward" concept. If they stop giving treats, the dog will quickly stop obeying the command, since the incentive is gone. However, if the dog is rewarded only once in a while, the model in her mind is "sitting on command sometimes leads to a treat," and she is likely to continue to obey the command, hoping that the next time is the charm. The concept was first explored by Fester and Skinner in 1957, who evaluated VRR schedules with pigeons and found that VRR produced stronger and more lasting responses than fixed rate results. A subsequent productivity study with human subjects confirmed VRR's efficacy and added that VRR schedule in combination with a varied reward amount lead to even better results [155]. VRR, also known as Probabilistic Reward, has been shown to successfully effect desired behavioral changes in a variety of contexts, from inducing better recycling habits [45] to increasing the number of purchases of a target product [120]. In the latter case lies the crux of the loot box controversy: paying for a chance to obtain a desirable in-game item is akin to placing a bet, and thus, loot boxes are frequently likened to gambling [49].

Gambling is a classic example of probabilistic reward and is known to cause addiction [147]. It is thus regulated in most countries, particularly in regards to the minimal age of persons allowed to participate in it [2]. Loot boxes are a common feature in games targeting younger audiences and there have been repeated calls for their legal regulation [49] . Some countries have forbidden loot boxes altogether, while others imposed various restrictions [78]. For example, Japan outright banned Kompu Gacha in 2012 [37]. "Kompu" stands for "Complete" and requires players to collect a set of common items in order to unlock the ultimate prize. It was prohibited under the false advertisement clause, since the odds of getting the final prize weren't disclosed and some players (including children [167]) spent equivalents of hundreds of dollars trying to complete their collection without any guarantee of eventual success. In contrast, Box Gacha, a legal alternative to Kompu, has a limited pool of items, so the player's odds improve with each draw [192]. Though it may not be cheap, it is ultimately possible to buy every item in the virtual box and guarantee the final prize.

While gacha mainly got its bad rep as a morally dubious way to get money out of players, monetization is not the only reason game makers employ this mechanic. In a blog on ethics in the video game industry, one game maker suggests VRR and loot boxes are only problematic if used for monetization [87], so what are the other uses? In a talk bluntly titled "Let's go whaling," a mobile game company CEO claims gacha is a lot more exciting and provides additional content, since it takes the players about five times as long to unlock content through gacha than by collecting soft currency and buying it outright [100]. Notice here that he talks about collecting soft currency, which is usually obtained through in-game grinding as opposed to real money purchases, and is thus not directly related to monetization. Others discuss the proper way to design gacha in a way that will provide value to all players (both paying and non-paying), no matter which random reward they get [3]. For example, one could augment gacha with a fusion mechanic, which allows fusing less-desirable items to either get new items, or to level up some other items. The same blog discusses seasonal and special event gacha sets as a way to encourage players to keep returning to the game.

What do the players have to say about the positive sides of gacha? One friendly gamer debate on loot boxes points out that loot boxes, when done right, are a great way for game developers to cater to audiences of all income levels. For example, Overwatch [58] only provides cosmetic items in their loot boxes, such as character outfits and poses. While these don't affect the gameplay in a tangible way, players of means are happy to invest into obtaining those unique looks, and in return, the game makers can release significant new content, such as level maps and new characters, free of charge, for everyone to enjoy [36]. Another gacha game enthusiast suggests that despite all the flaws, there is a unique appeal to gacha-heavy games where the loot box contents have a direct influence on the player's progress (such as characters, weapons and equipment) [88]. Namely, it is the uniqueness of every player's experience. As opposed to classic gacha-free RPG's where a player can follow a 100% completion guide and end up with the exact same equipment and character roster as any other player following the same guide, there's no one-size-fits-all for gacha games. As a result, those games foster communities with extended wikis and forums where people discuss strategies built around their unique sets of in-game assets (e.g. the fan-made wiki for Final Fantasy Brave Exvius, a gacha-heavy game, has accumulated over 36,000 pages since the game's launch in 2015 [1]). The blogger also mentions that those who can't afford to pay cash for gacha can get there by grinding, and that as long as the grinding core loop is enjoyable and further spiced up with things like special quests, the prospect doesn't have to be daunting.

With all these points considered, what are the potential applications of VRR and gacha in an educational game? Since monetization is outside of the scope of this work, I'm focusing on loot boxes obtainable through grinding. Once again, I'm using a hypothetical hanzi learning game as an example below.

Encouraging target behavior: Anyone who's ever tried to learn a new language knows that there's no getting around frequent reviewing and practice when conquering new vocabulary [178]. Thus, getting students to review is a target behavior. If our LWS learning game's grinding core loop involves reinforcement of previously learned material, the VRR will be tied to that target behavior.

Pacing student progress: If a student tries to cram too much material at once, it is likely that a lot of it won't stick. Research on distributed learning suggests that breaking the material into spaced out study sessions is more beneficial for learning [180]. With that, gacha's ability to pace the player's progression is a good fit to prevent the student from overwhelming themselves.

Gacha contents: For gacha to have a positive effect on engagement, the potential prizes need to be desirable. In an educational game without orthogonal mechanics, these prizes also need to contribute to the student's progress in mastering the language. One potential set of such contents can consist of supplementary vocabulary words or parts of such words that can be later assembled. Event and seasonal gachas can offer unique sets of words, such as names of folklore characters or various festival-related words. Word level up tokens can grant sample sentences for the upgraded word. The game's other features and themes can suggest other relevant loot options, for example, if the game has a social element, avatar customization options could be added to the pool.

In conclusion, after reviewing the history and the controversial nature of RVV mechanics, such as loot boxes, I maintain that there is more to gacha than exploiting players for money. To my knowledge, there are no educational games employing gacha to enhance learning outcomes. However, gacha's capacity to generate excitement, encourage target behavior, and promote retention, among other things, could conceivably be utilized in an educational game and thus warrant further investigation.

6.2 Zeigarnik (a.k.a. the Unfinished Business) effect

Figure 6.2: Screenshot of a typical quest list in World of Warcraft.

Unlike with Gacha, there are no current attempts to legislate the use of Zeigarnik effect in video games, though it's been named among the candidates for such legislature due to the power it can hold over players [142]. Despite its wide use in modern mobile and online games, it receives a lot less media coverage, and the general public appears a lot less aware of it and it's potentially exploitative nature. Even when it is talked about, it often isn't mentioned by name [175], so let's take a look into the origins of the term.

Zeigarnik effect was first observed by a Russian psychologist of the same name. She noticed that the waiters remembered the order contents and customers only until the order was complete and the guests were out the door [42]. In a subsequent series of experiments, the participants were given around 25 tasks. They were allowed to finish some tasks, while for others they were interrupted. At the end of the experiment, the subjects were asked to recall as many tasks as possible. On average, the participants were twice as likely to recall the interrupted tasks. The ratio grew the closer a particular task was to completion when interrupted. Zeigarnik hypothesised that the unfinished task created a "psychic" tension in a person's mind, which served as an impetus to complete the task. Not only were participants more likely to remember unfinished tasks, some were even compelled to return and finish the tasks, unbidden [158], as completing the task can provide closure and release of the tension.

Ending a book chapter on a cliff-hanger and advertising a news segment with a question can be considered examples of employing Zeigarnik effect to capture an audience.

In digital games, it is often implemented either through a plethora of count-down timers of various lengths, like in the Cat Game, or through lengthy quest logs, like in TWD: Our World [197].

Through an experimental non-educational game, Alexandrovsky et al. showed that the mechanic of waiting (e.g. for resources to gather, or for a basket to arrive in the Cat Game) promoted snacking [1] behavior in players, which is linked with establishing long-lasting behaviour change, and is ideal for getting the students into the habit of reviewing the material on a regular basis [4]. Murphy et al. suggest that Zeigarnik effect is well suited for motivating players in training games to improve their skills through lists of skill-focused quests/tasks [146].

As with gacha, the main objection players have with the waiting mechanic is in the way it's been used for monetization [156]. One infamous case is the 2014 EA game Dungeon Keeper, which made players wait up to 24 hours to excavate a single tile, all the while taunting the player to speed up construction with real world money [199]. Most F2P games allow for watching ads, waiting, or grinding as a way to progress without spending money. If monetization isn't an objective, a variety of supplementary skill-focused activities can make the grinding and/or waiting a more agreeable experience for the players.

(add section about application to educational game: pacing, both snacking and extended play)

A note of caution: research showed that one way to diminish Zeigarnik effect is by promising a reward, whether or not the task was completed [136]. In that study, the participants who received a reward were less likely to return to attempt to finish the

[1] A pattern of brief periodic interactions [149]

task than those who were given no reward (58% vs. 86%, $p<0.05$). This is something to keep in mind when incorporating Zeigarnik effect into a game.

6.3 Ethics

F2P games are rife with exploitative elements, like gacha, Zeigarnik effect, peer pressure, FOMO, etc., trapping players in the game and targeting their time and money. Possibly due to their infamy, many of these elements have not been widely used in serious games. I suggest that these techniques could be used in educational games to promote extended play and learning. Chou [25] warned that black hat motivation devices can leave a player with a bad taste. However, I suggest the bitterness comes from the realization of the time and money wasted with little to show for it. In a quality educational game, time spent in-game should correlate with knowledge gained, thus counterbalancing the stereotypical harmful effects the addictive game mechanics may cause [73]. While commercial developers seem to have no such reservation, I advise caution against using black hat motivation in serious games targeting children in light of their cognitive and developmental immaturity [57]. However, this still leaves the door open for white-hat gamified learning opportunities like Radical Tunes and Zen Hanzi.

Chapter 7

Conclusion

Acquisition of LWS languages such as Chinese and Japanese is a relevant subject in the modern world. Learning and retaining the thousands of characters needed for functional literacy in these languages is an ongoing struggle for foreigners and native speakers alike. In this dissertation I addressed some of the ways to ease this burden through development of LWS learning games. Since learning hanzi and kanji is an ongoing months or even years-long quest, I also presented my ideas for using black hat engagement and retention techniques that are widely used in commercial games, but have been largely overlooked by the serious game community. It is my hope that the ideas presented in this dissertation can be of use to the educational game research and development community in general and LWS learning games in particular.

7.1 Revisiting the Research Questions

At the beginning of this dissertation I outlined four research questions. In this section I revisit them, along with ways I addressed them in my work.

- RQ1 **Task Definition**: Identifying the disconnect between current research on effective LWS learning approaches, educational game design, and the commercially available LWS learning technologies. My literature review on LWS learning, and media research of popular commercially available games and apps showed that some of the major disconnections included lack of component-focused learning apps and games, while LWS educational research confirms again and again that component awareness is instrumental in developing solid reading and writing skills. Another area has to do with endogenous material integration, which research shows is more effective, but the current learning games tend to overlook.

- RQ2 **Operationalization**: Creating LWS learning apps employing the missing elements identified in RQ1. I designed, developed, and evaluated prototypes of two LWS learning games with endogenous learning content integration, component-focused mechanics and melody-based mnemonics. My findings conformed to the current consensus of the efficacy of component-focused approach to learning hanzi. Radical Tunes, the musical kanji-drawing game also demonstrated the promising potential of melody as a mnemonic device in stroke knowledge acquisition, especially for characters with higher stroke number count.

- RQ3 **Scalability**: There are thousands of characters in LWS with complex subsystems of structure and components within those characters. How can the authoring burden be minimized to allow for inclusion of more characters in LWS learning software?

Open source svg databases of Chinese and Japanese characters can be scraped for information such as stroke order, number and direction. This information can be used to auto-generate unique melodies or other accompanying features for each character. When it comes to focus on components to target substitution or stroke addition/omission production errors, I designed a paired database of hanzi and components. While one part succinctly stores the structural layout information of the character, the other part keeps a record for each component and its variants, along with a list of look-alike components, for easy retrieval for component-manipulation based mechanics. The existing databases don't make a distinction between common component variants, or describe the character structure in terms of its smallest components, so these databases have to be created by hand. However, once created, they can be used in many useful ways for LWS education.

- RQ4 **Future work** : Application of black hat motivational techniques to educational games in general and LWS in particular. I analyzed two moderately popular mobile F2P games, identifying their engagement and retention mechanics, and discussing their application to a hypothetical LWS learning game. Focusing on two of the more controversial game mechanics: gacha and Zeigarnik effect, I propose that the potential of employment of these powerful devices in educational games warrants further investigation.

7.2 Looking to the Future

There are several areas in which my research could be expanded.

1. Comparative studies of several versions of an educational game with various engagement features enabled in each, similar to Alexandrovsky's work with non-educational games [4].

2. Studies in the wild to assess the long-term efficacy and efficiency of the two mini-game mechanics I described via longitudinal studies.

3. Completion of the hanzi and component databases, along with manuals for suggested use.

Bibliography

[1] Final Fantasy Brave Exvius Wiki. https://exvius.gamepedia.com/Final_Fantasy_Brave_Exvius_Wiki, 2016. [Online; accessed 23-February-2023].

[2] Gambling laws licensing authorities around the world. https://www.gamblingsites.org/laws/, 2016. [Online; accessed 01-December-2022].

[3] Pepe Agell. Getting gacha right: Tips for creating successful in-game lotteries. https://www.chartboost.com/blog/getting-gacha-right-tips-for-creating-successful-in-game-lotteries/, 2016. [Online; accessed 02-December-2022].

[4] Dmitry Alexandrovsky, Maximilian Achim Friehs, Max V. Birk, Rowan K. Yates, and Regan L. Mandryk. Game dynamics that support snacking, not feasting. CHI PLAY '19, page 573–588, New York, NY, USA, 2019. Association for Computing Machinery.

[5] Sultan Alharthi, Olaa Alsaedi, Zachary Toups, Theresa Tanenbaum, and Jessica Hammer. Playing to wait: A taxonomy of idle games. pages 1–15, 04 2018.

[6] Khaled Mohammad Alomari, Cornelius Ncube, and Khaled Shaalan. Predicting success of a mobile game: A proposed data analytics-based prediction model. In *Recent Trends and Future Technology in Applied Intelligence*, pages 127–134. Springer International Publishing, 2018.

[7] Alissa N Antle, Greg Corness, and Milena Droumeva. What the body knows: Exploring the benefits of embodied metaphors in hybrid physical digital environments. *Interacting with Computers*, 21(1-2):66–75, 2009.

[8] Alissa N Antle, Milena Droumeva, and Greg Corness. Playing with the sound maker: do embodied metaphors help children learn? In *Proceedings of the 7th international conference on Interaction design and children*, pages 178–185, 2008.

[9] Anurag Anurag. 8 steps: Your guide to successful mobile game development. `https://www.newgenapps.com/en/blogs/steps-successful-mobile-game-development`, 2017. [Online; accessed 03-October-2022].

[10] Overpass Apps: Super-Human Apps and Games. Chinese SPY: Learn Mandarin, 2016.

[11] Argeworld. Kanji Battle, 2015.

[12] Davidson & Associates. Math Blaster!, 1983.

[13] Nick Babich. A comprehensive guide to mobile app design. `https://www.smashingmagazine.com/2018/02/comprehensive-guide-to-mobile-app-design/`, 2018. [Online; accessed 03-October-2022].

[14] Xiaojun Bi, Yang Li, and Shumin Zhai. Ffitts law: Modeling finger touch with fitts' law. CHI '13, page 1363–1372. Association for Computing Machinery, 2013.

[15] Donald Bitzer, Peter Braunfeld, and W. Lichtenberger. Plato: An automatic teaching device. *IRE Transactions on Education*, 4(4):157–161, 1961.

[16] Broderbund. Where in the World is Carmen Sandiego?, 1985.

[17] Eric Butler, Erik Andersen, Adam M. Smith, Sumit Gulwani, and Zoran Popović. Automatic game progression design through analysis of solution features. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, page 2407–2416, New York, NY, USA, 2015. Association for Computing Machinery.

[18] Paul Cairns, Anna Cox, and A Imran Nordin. *Immersion in digital games: review of gaming experience research*, volume 1. Wiley-Blackwell, 2014.

[19] Adam Carpenter. Why the first ten minutes are crucial if you want to keep players coming back. `https://medium.com/googleplaydev/why-the-first-ten-minutes-is-crucial-if-you-want-to-keep-players-coming-back-to-your-mobile-game-4a89031b6308`, 2017. [Online; accessed 03-October-2022].

[20] Ang Chen, Paul W. Darst, and Robert P. Pangrazi. What constitutes situational interest? validating a construct in physical education. *Measurement in Physical Education and Exercise Science*, 3(3):157–XXX, 1999.

[21] Alan Cheng, Lei Yang, and Erik Andersen. Teaching language and culture with

a virtual reality game. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 541–549. ACM, 2017.

[22] M.-T. Cheng, H.-C. She, and L.A. Annetta. Game immersion experience: its hierarchical structure and impact on game-based science learning. *Journal of Computer Assisted Learning*, 31(3):232–253, 2015.

[23] M-T Cheng, H-C She, and Leonard A Annetta. Game immersion experience: its hierarchical structure and impact on game-based science learning. *Journal of Computer Assisted Learning*, 31(3):232–253, 2015.

[24] J. C. Chou, C. Hung, and Y. Hung. Design factors of mobile game for increasing gamer's flow experience. In *2014 IEEE International Conference on Management of Innovation and Technology*, pages 137–139, 2014.

[25] Yu-kai Chou. Learn gamification with yu-kai chou: cheat codes to win the game of life. `https://yukaichou.com/`. [Online; accessed 10-February-2020].

[26] Tsung-Yu Liu; Yu-Ling Chu. Using ubiquitous games in an english listening and speaking course: Impact on learning outcomes and motivation. In *Computers Education*, pages 630–643. Elsevier BV, 2010.

[27] Hui Ya Chuang. Mobile assisted language learning apps for the chinese classroom. *Journal of Technology and Chinese Language Teaching*, 2016.

[28] Looi-Chin Ch'ng, Hie-Ling Ting, and Kee-Man Chuah. Common stroke mistakes in chinese character writing among business administrative students/ch'ng looi-

chin, ting hie-ling and chuah kee-man. *Academic Journal of Business and Social Sciences (AJoBSS)*, 2:1–10, 2018.

[29] Thomas M Connolly, Elizabeth A Boyle, Ewan MacArthur, Thomas Hainey, and James M Boyle. A systematic literature review of empirical evidence on computer games and serious games. *Computers & education*, 59(2):661–686, 2012.

[30] Thomas M. Connolly, Mark Stansfield, and Thomas Hainey. An alternate reality game for language learning: Arguing for multilingual motivation. *Comput. Educ.*, 57(1):1389–1415, August 2011.

[31] Iam Cooper and Royston Shufflebotham. PDA web browsers: Implementation issues, 1995.

[32] Dan Crawley. Making educational games is tough, especially if you want to make money. `https://venturebeat.com/2015/06/24/making-educational-games-is-tough-especially-if-you-want-to-make-money/`, 2015. [Online; accessed 23-February-2023].

[33] Mihaly Csikszentmihalyi. *Beyond boredom and anxiety.* Jossey-Bass Publishers, 1975.

[34] Marjorie Ann M Cuerdo and Edward F Melcer. "I'll be back": A taxonomy of death and rebirth in platformer video games. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2020.

[35] Gabriel Culbertson, Shiyu Wang, Malte Jung, and Erik Andersen. Social situational language learning through an online 3d game. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 957–968. ACM, 2016.

[36] Oscar Dayus. Are loot boxes really that bad? — versus. `https://www.youtube.com/watch?v=Y8Bog1GSY1M`, 2018. [Online; accessed 02-December-2022].

[37] Kathleen De Vere. Japan officially declares lucractive kompu gacha practice illegal in social games. `https://www.adweek.com/digital/japan-officially-declares-lucractive-kompu-gacha-practice-illegal-in-social-games/`, 2012. [Online; accessed 01-December-2022].

[38] Edward Deci. *Intrinsic Motivation.* Springer, 1975.

[39] Edward Deci and Richard Ryan. *Intrinsic Motivation and Self-Determination in Human Behavior.* Springer, 1985.

[40] Chris Dede. Immersive interfaces for engagement and learning. *science*, 323(5910):66–69, 2009.

[41] Jonathan deHaan, W. Michael Reed, and Katsuko Kuwada. The effect of interactivity with a music video game on second language vocabulary recall. *Language Learning & Technology*, 14(2):74–94, 2010.

[42] Florence L. Denmark. *Zeigarnik Effect.* John Wiley Sons, Ltd, 2010.

[43] Tamara Denning, Adam Lerner, Adam Shostack, and Tadayoshi Kohno. Control-alt-hack: The design and evaluation of a card game for computer security awareness and education. In *Proceedings of the 2013 ACM SIGSAC Conference on Computer Communications Security*, CCS '13, page 915–928. Association for Computing Machinery, 2013.

[44] Apple Developer. UI design do's and don'ts. `https://developer.apple.com/design/tips/`. [Online; accessed 03-October-2022].

[45] William D. Diamond and Ben Z. Loewy. Effects of probabilistic rewards on recycling attitudes and behavior1. *Journal of Applied Social Psychology*, 21(19):1590–1607, 1991.

[46] Mimi Dietderich. Gachapon: Japan's irresistible capsule toys you never knew you needed. `https://blog.gaijinpot.com/gachapon-japans-irresistible-capsule-toys-you-never-knew-you-needed/`, 2018. [Online. Accessed 29-November-22].

[47] John Dominic S Diez and Edward F Melcer. Cookie mania: a serious game for teaching internet cookies to high school and college students. In *Joint International Conference on Serious Games*, pages 69–77. Springer, 2020.

[48] Mary Jo Dondlinger. Educational video game design: A review of the literature. *Journal of applied educational technology*, 4(1):21–31, 2007.

[49] Aaron Drummond and James D. Sauer. Video game loot boxes are psychologically akin to gambling. *Nature Human Behaviour*, 2:530–532, 2018.

[50] Sleepy Duck. Learn Japanese to Survive! Kanji Combat, 2018.

[51] Mark Dunlop and Stephen Brewster. The challenge of mobile devices for human computer interaction. *Personal Ubiquitous Comput.*, 6(4):235–236, jan 2002.

[52] Duolingo. Duolingo, 2007.

[53] Duolingo. Tinycards, 2016.

[54] Carol S. Dweck and Ellen L. Leggett. A social-cognitive approach to motivation and personality. *Psychological Review*, 95:256–273, 1988.

[55] Carol S Dweck and Daniel C Molden. Their impact on competence motivation and acquisition. *Handbook of competence and motivation: Theory and application*, 135, 2017.

[56] Darren Edge, Kai-Yin Cheng, and Michael Whitney. Spatialease: Learning language through body motion. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, New York, NY, USA, 2013. ACM.

[57] Alan M Emond and Mark D Griffiths. Gambling in children and adolescents. *British Medical Bulletin*, 136(1):21–29, 09 2020.

[58] Blizzard Entertainment. Overwatch, 2016.

[59] Rovio Entertainment. Angry Birds 2, 2015.

[60] Toyoda Etsuko. Teaching kanji by focusing on learners' development of graphemic awareness. *Australian Review of Applied Linguistics. Supplement Series*, 15:155–168, 1998.

[61] Xiangmin Fan, Wencan Luo, and Jingtao Wang. Mastery learning of second language through asynchronous modeling of native speakers in a collaborative mobile game. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 4887–4898. ACM, 2017.

[62] Charles Fester and Burrhus Skiner. Schedules of reinforcement. *Science*, 127(3296), 1957.

[63] Vicente Filho, Átila Valgueiro Malta Moreira, and Geber Ramalho. Deepening the understanding of mobile game success. In *Proceedings of the XIII Brazilian Symposium on Computer Games and Digital Entertainment (SBGames)*, 2014.

[64] Carl Gene Fordham. 800 most common chinese character components (simplified Chinese version). `http://www.carlgene.com/800_Most_Common_Chinese_Character_Components_Simplified_Chinese_carlgene.pdf`, 2015. [Online. Accessed 05-September-2021].

[65] Nick Fortugno. Soapbox: Learning to play to learn - lessons in educational game design. `https://www.gamasutra.com/view/feature/130686/soapbox_learning_to_play_to_learn_.php`, 2005. [Online. Accessed 03-October-2022].

[66] David Fried. This is the formula for creating an insanely successful mobile game. https://www.forbes.com/sites/quora/2016/07/07/this-is-the-formula-for-creating-an-insanely-successful-mobile-game/, 2016. [Online. Accessed 03-October-2022].

[67] Julian Frommel and Regan L. Mandryk. Daily quests or daily pests? the benefits and pitfalls of engagement rewards in games. volume 6 of *CHI PLAY*, New York, NY, USA, oct 2022. Association for Computing Machinery.

[68] GameAnalytics. Mobile gaming industry analysis for H1 2019. https://gameanalytics.com/resources/item/mobile-gaming-industry-analysis-h1-2019, 2019. [Online; accessed 15-December-2019].

[69] Mino Games. Cat Game - The Cats Collector!, 2017.

[70] Next Games. The Walking Dead: Our World, 2018.

[71] Riot Games. League of Legends, 2009.

[72] Andrea Gee. The first gacha craze in japan. https://web.archive.org/web/20190822162045/http://www.gachapontoys.com/category/history/, 2018. [Online. Accessed 29-November-2022].

[73] James Gee. *Good Video Games and Good Learning: Collected Essays on Video Games, Learning and Literacy.* Peter Lang Inc., International Academic Publishers, 2007.

[74] Lance W. Gibbon. *Effects of LEGO Mindstorms on convergent and divergent problem-solving and spatial abilities in fifth and sixth grade students.* PhD thesis, Seattle Pacific University, Washington, January 2007.

[75] Herbert Gingold and Eugene Abravanel. Music as a mnemonic: The effects of good-and bad-music settings on verbatim recall of short passages by young children. *Psychomusicology: A Journal of Research in Music Cognition*, 1987.

[76] Katelyn M Grasse, Max Kreminski, Noah Wardrip-Fruin, Michael Mateas, and Edward F Melcer. Using self-determination theory to explore enjoyment of educational interactive narrative games: A case study of academical. *Frontiers in Virtual Reality*, page 74, 2022.

[77] Jacob Habgood and Shaaron Ainsworth. Motivating children to learn effectively: Exploring the value of intrinsic integration in educational games. *The Journal of the Learning Sciences*, 20:169–206, 04 2011.

[78] Leana Hafer. The legal status of loot boxes around the world, and what's next in the debate. `https://www.pcgamer.com/the-legal-status-of-loot-boxes-around-the-world-and-whats-next/`, 2016. [Online. Accessed 01-December-2022].

[79] Susan Hallam, John Price, and Georgia Katsarou. The effects of background music on primary school pupils' task performance. *Educational Studies*, 28(2):111–122, 2002.

[80] Andrew Head, Yi Xu, and Jingtao Wang. Tonewars: Connecting language learners and native speakers through collaborative mobile games. In *International Conference on Intelligent Tutoring Systems*, pages 368–377. Springer, 2014.

[81] Carolyn Hebert. Movement memory: How we learn, retain and remember dance. `https://www.thedancecurrent.com/feature/movement-memory`, 2017. [Onlin; accessed 06-September-2020].

[82] Karen Hewitt. Blocks as a tool for learning: Historical and contemporary perspectives. *Young Children*, 56(1):6–10, 1 2001.

[83] C.C. Heyde. Central limit theorem. In *Encyclopedia of Actuarial Science*. American Cancer Society, 2006.

[84] Kieran Hicks, Kathrin Gerling, Patrick Dickinson, and Vero Vanden Abeele. Juicy game design: understanding the impact of visual embellishments on player experience. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, pages 185–197, 2019.

[85] Suzanne Hidi and K Ann Renninger. The four-phase model of interest development. *Educational psychologist*, 41(2):111–127, 2006.

[86] A.B. Hill and R.E. Perkins. *Towards a model of boredom*, volume 76, pages 235–240. 1985.

[87] Celia Hodent. Ethics in the videogame industry: A mythbusting and scientific approach. `https://www.gamedeveloper.com/business/ethics-in-the-`

videogame-industry-a-mythbusting-and-scientific-approach, 2019. [Online. Accessed 01-December-2022].

[88] Christopher Hovermale. Why do people enjoy, yet loathe, gacha games? https://www.Destructoid.com/stories/why-do-people-enjoy-yet-loathe-gacha-games--472728.phtml, 2017. [Online; accessed 02-December-2022].

[89] Chen Hsueh-Chih, Hsu Chih-Chun, Chang Li-Yun, Lin Yu-Chi, Chang Kuo-En, and Sung Yao-Ting. Using a radical-derived character e-learning platform to increase learner knowledge of chinese characters. *Language Learning & Technology*, 17:89–106, 2013.

[90] C Huang. The graphabet and bujian approach at acquiring hanzi (chinese character) writing skill. *L1 Educational Studies in Language and Literature*, (SI ReadWritMult), 2017.

[91] Chen hui Tsai, Chin-Hwa Kuo, Wen-Bing Horng, and Chun-Wen Chen. Effects on learning logographic character formation in computer-assisted handwriting instruction. *Language Learning & Technology*, 16(1):110–130, February 2012.

[92] Kai Huotari and Juho Hamari. Defining gamification: A service marketing perspective. In *Proceeding of the 16th International Academic MindTrek Conference*, MindTrek '12, page 17–22, New York, NY, USA, 2012. Association for Computing Machinery.

[93] IBM. The Sumerian Game, 1966.

[94] Sarah Impey. Mobile vs desktop UI: Key differences in design. https://gameanalytics.com/blog/mobile-desktop-ui-design.html, 2018. [Online. Accessed 03-October-2022].

[95] Gumi inc. Final Fantasy: Brave Exvius, 2015.

[96] Mostafa Janebi Enayat and Mohsen Haghighatpasand. Exploiting adventure video games for second language vocabulary recall: a mixed-methods study. *Innovation in Language Learning and Teaching*, pages 1–15, 2017.

[97] Japan Foundation. Survey report on Japanese-language education abroad 2015. https://www.jpf.go.jp/j/project/japanese/survey/result/dl/survey_2015/Report_all_e.pdf, 2015. [Online; accessed 06-September-2020].

[98] Charlene Jennett, Anna L. Cox, Paul Cairns, Samira Dhoparee, Andrew Epps, Tim Tijs, and Alison Walton. Measuring and defining the experience of immersion in games. *Int. J. Hum.-Comput. Stud.*, 66(9):641–661, September 2008.

[99] Charlene Jennett, Anna L. Cox, Paul Cairns, Samira Dhoparee, Andrew Epps, Tim Tijs, and Alison Walton. Measuring and defining the experience of immersion in games. *International journal of human-Computer studies*, 66(9):641–661, sep 2008.

[100] Torulf Jernström. Let's go whaling: Tricks for monetising mobile game players

with free-to-play. https://www.youtube.com/watch?v=xNjI03CGkb4, 2016. [Online; accessed 02-December-2022].

[101] Chen Jia. 40 million foreigners learning Chinese. https://www.chinadaily.com.cn/china/2010-12/13/content_11689902.htm, 2010. [Online; accessed 06-September-2020].

[102] Daniel Johnson, M. John Gardner, and Ryan Perry. Validation of two game experience scales: The player experience of need satisfaction (pens) and game experience questionnaire (geq). *International Journal of Human-Computer Studies*, 118:38–46, 2018.

[103] Martin Jonasson and Petri Purho. Juice it or lose it. https://www.gdcvault.com/play/1016487/Juice-It-or-Lose, 2012. [Online; accessed 18-November-2020].

[104] Terry Joyce, Hisashi Masuda, and Taeko Ogawa. Jōyō kanji as core building blocks of the japanese writing system: Some observations from database construction. *Written Language Literacy*, 17:173–194, 10 2014.

[105] Kari Kallinen. The effects of background music on using a pocket computer in a cafeteria: Immersion, emotional responses, and social richness of medium. In *Extended Abstracts on Human Factors in Computing Systems*, CHI '04. ACM, 2004.

[106] Dominic Kao, Rabindra Ratan, Christos Mousas, Amogh Joshi, and Edward F Melcer. Audio matters too: How audial avatar customization enhances visual

avatar customization. In *CHI Conference on Human Factors in Computing Systems*, pages 1–27, 2022.

[107] Ajay B. Karthic, Gopinath Bharathi, Conrad S. Singh, Abhinav Tucker, and Harriet B. Nembhard. Knowledge discovery of game design features by mining user-generated feedback. *Computers in Human Behavior*, 60:361–371, 2016.

[108] Michail Katkoff. Supercell's Spooky Pop and the six rules of a hit puzzle game. `https://www.deconstructoroffun.com/blog//2015/01/supercells-spooky-pop-and-six-rules-of.html`, 2014. [Online; accessed 18-December-2022].

[109] Oleksandra Keehl, Dominic Kao, and Edward Melcer. Zen Hanzi: A game for raising hanzi component awareness. In *Proceedings of the International Conference on the Foundations of Digital Games*, 09 2022.

[110] Oleksandra Keehl and Edward Melcer. Radical Tunes: exploring the impact of music on memorization of stroke order in logographic writing systems. In *Proceedings of the 14th International Conference on the Foundations of Digital Games*, pages 1–6, 2019.

[111] Oleksandra G Keehl and Edward F Melcer. Music to my ears: Developing kanji stroke knowledge through an educational music game. *Multimodal Technologies and Interaction*, 5(12):83, 2021.

[112] May-Li Khoe. Chocolate-covered broccoli. `https://klr.tumblr.com/post/`

149091244433/chocolate-covered-broccoli, 2016. [Online. Accessed 03-October-2022].

[113] Amy Jo Kim. What makes games compelling?, 2014. [Online; accessed 03-October-2022].

[114] King. Candy Crush Saga, 2012.

[115] Raph Koster. *Theory of Fun for Game Design, 2nd Edition.* O'Reilly Media, Inc., 2005.

[116] Markus Krause, Aneta Takhtamysheva, Marion Wittstock, and Rainer Malaka. Frontiers of a paradigm: Exploring human computation with digital games. In *Proceedings of the ACM SIGKDD Workshop on Human Computation*, HCOMP '10, page 22–25, New York, NY, USA, 2010. Association for Computing Machinery.

[117] Anuj Kumar, Pooja Reddy, Anuj Tewari, Rajat Agrawal, and Matthew Kam. Improving literacy in developing countries using speech recognition-supported games on mobile devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1149–1158. ACM, 2012.

[118] George Lakoff and Mark Johnson. *Metaphors we live by.* University of Chicago press, 2008.

[119] Effie Lai-Chong Law and Xu Sun. Evaluating user experience of adaptive digital educational games with activity theory. *International Journal of Human-Computer Studies*, 70(7):478–497, 2012.

[120] Chang-Yuan Lee, Carey K. Morewedge, Guy Hochman, and Dan Ariely. Small probabilistic discounts stimulate spending: Pain of paying in price promotions. *Journal of the Association for Consumer Research*, 4(2):160–171, 2019.

[121] Lukas Legner, Chloe Eghtebas, and Gudrun Klinker. Persuasive mobile game mechanics for user retention. In *Extended Abstracts of the Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts*, CHI PLAY '19 Extended Abstracts, page 493–500. Association for Computing Machinery, 2019.

[122] Louis Leung. Using tablet in solitude for stress reduction: An examination of desire for aloneness, leisure boredom, tablet activities, and location of use. *Computers in Human Behavior*, 48:382–391, 2015.

[123] Michael Li. A systematic review of the research on chinese character teaching and learning. *Frontiers of Education in China*, 15(1):39–72, 2020.

[124] Xiao Li. *Investigation of teaching stroke order of Chinese characters in beginner CSL learners (in Chinese)*, pages 23–25. 2006.

[125] S. Librenjak, K. Vučković, and Z. Dovedan. Multimedia assisted learning of japanese kanji characters. In *2012 Proceedings of the 35th International Convention MIPRO*, pages 1284–1289, 2012.

[126] William Lim. Blood in the water: A history of microtransactions in the video game industry. `https://williamlim3.medium.com/blood-in-the-water-a-history-`

of-microtransactions-in-the-video-game-industry-e5bf9e3de4da, 2018. [Online; accessed 28-November-2022].

[127] C Lin, Keyi Zhou, and Shiyun Yang. A survey of mobile apps for learning chinese vocabulary. *Journal of Technology and Chinese Language Teaching*, 2018.

[128] Zheng Liu, Holden. Analyzing students' language learning experience in an augmented reality mobile game: An exploration of an emergent learning environment. In *Procedia - Social and Behavioral Sciences*, pages 369–374. Elsevier BV, 2016.

[129] ConcernedApe LLC. Stardew Valley, 2016.

[130] Lap-yan Lo, Pui-sze Yeung, Connie Suk-Han Ho, David Wai-ock Chan, and Kevin Chung. The role of stroke knowledge in reading and spelling in chinese. *Journal of Research in Reading*, 39(4):367–388, 2016.

[131] Nicholas Lovell. The pyramid of free-to-play game design. https://www.gamasutra.com/blogs/NicholasLovell/20130919/200606/The_Pyramid_of_FreetoPlay_game_design.php, 2013. [Online; accessed 12-November-2020].

[132] Karen Ludke, Fernanda Ferreira, and Katie Overy. Singing can facilitate foreign language learning. *Memory cognition*, 07 2013.

[133] Amanda Mason and Wenxin Zhang. An exploration of the use of mobile applications to support the learning of chinese characters employed by students of chinese

as a foreign language. *Beyond the language classroom: researching MOOCs and other innovations*, pages 99–112, 2017.

[134] Joshua McCoy, Mike Treanor, Ben Samuel, Aaron A. Reed, Michael Mateas, and Noah Wardrip-Fruin. Prom week: Designing past the game/story dilemma. In *FDG*, 2013.

[135] Patrick McGinnis. Social theory at HBS: McGinnis' two FOs. `http://www.harbus.org/2004/social-theory-at-hbs-2749/`, May 2004. [Online; accessed 02-December-2020].

[136] Kenneth O. McGraw and Jirina Fiala. Undermining the zeigarnik effect: Another hidden cost of reward. *Journal of Personality*, 50(1):58–66, 1982.

[137] James McMillan and Donelson Forsyth. What theories of motivation say about why learners learn. *New Directions for Teaching and Learning*, 1991:39 – 52, 03 1991.

[138] MECC. The Oregon Trail, 1971.

[139] Edward F Melcer and Marjorie Ann M Cuerdo. Death and rebirth in platformer games. In *Game User Experience And Player-Centered Design*, pages 265–293. Springer, 2020.

[140] Edward F Melcer and Katherine Isbister. Bots & (main) frames: exploring the impact of tangible blocks and collaborative play in an educational programming

game. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2018.

[141] Walid Mestadi, Khalid Nafil, Raja Touahni, and Rochdi Messoussi. An assessment of serious games technology: toward an architecture for serious games design. *International Journal of Computer Games Technology*, 2018, 2018.

[142] Christian Montag, Bernd Lachmann, Marc Herrlich, and Katharina Zweig. Addictive features of social media/messenger platforms and freemium games against the background of psychological and economic theories. *International Journal of Environmental Research and Public Health*, 16(14), 2019.

[143] Átila Valgueiro Malta Moreira, Geber Ramalho, et al. Which features matter to make a successful mobile game? *Journal on Interactive Systems*, 5(2), 2014.

[144] Ezzeldin A. Mostafa and Ashraf A.H. El Midany. Review of mnemonic devices and their applications in cardiothoracic surgery. *Journal of the Egyptian Society of Cardio-Thoracic Surgery*, 25(1):79 – 90, 2017.

[145] Aline Moussard, Emmanuel Bigand, Sylvie Belleville, and Isabelle Peretz. Music as a mnemonic to learn gesture sequences in normal aging and alzheimer's disease. *Frontiers in Human Neuroscience*, 8:294, 2014.

[146] Curtiss Murphy, Dustin Chertoff, Michael Guerrero, and Kerry Moffitt. Design better games: Flow, motivation, and fun. *Design and development of training games: Practical guidelines from a multidisciplinary perspective*, page 1773, 2014.

[147] John B. Murray. Review of research on pathological gambling. *Psychological Reports*, 72(3):791–810, 1993.

[148] Musa Nushi, Samaneh Shafeie, and Meisam Shafiei. Game-based teaching of stress placement on multi-syllabic english words. *Journal of Teaching English Language Studies*, 5:46–64, 01 2017.

[149] Antti Oulasvirta, Sakari Tamminen, Virpi Roto, and Jaana Kuorelahti. Interaction in 4-second bursts: The fragmented nature of attentional resources in mobile hci. CHI '05, page 919–928, New York, NY, USA, 2005. Association for Computing Machinery.

[150] Joana Pais. Ux for mobile: The rise of fat-finger design. `https://www.outsystems.com/blog/posts/ux-for-mobile-fat-finger-design/`, 2018. [Online; accessed 23-February-2023].

[151] Jan L Plass, Bruce D Homer, Elizabeth O Hayward, Jonathan Frye, Tsu-Ting Huang, Melissa Biles, Murphy Stein, and Ken Perlin. The effect of learning mechanics design on learning outcomes in a computer-based geometry game. In *E-learning and games for training, education, health and sports*, pages 65–71. Springer, 2012.

[152] Jan L Plass, Bruce D Homer, Charles Kinzer, Jonathan Frye, and Ken Perlin. Learning mechanics and assessment mechanics for games for learning. *G4LI White Paper*, 1:2011, 2011.

[153] Jan L Plass, Bruce D Homer, and Charles K Kinzer. Foundations of game-based learning. *Educational psychologist*, 50(4):258–283, 2015.

[154] Jan L Plass, Richard E Mayer, and Bruce D Homer. *Handbook of game-based learning*. Mit Press, 2020.

[155] Robert D. Pritchard, Dale W. Leonard, Clarence W. Von Bergen, and Raymond J. Kirk. The effects of varying schedules of reinforcement on human task performance. *Organizational Behavior and Human Performance*, 16(2):205–230, 1976.

[156] Fanny Ramirez. Affect and social value in freemium games. *Social, casual and mobile games: The changing gaming landscape*, pages 117–132.

[157] K Ann Renninger, Suzanne Hidi, Andreas Krapp, and Ann Renninger. *The role of interest in learning and development*. Psychology Press, 2014.

[158] Maria A Rickers-Ovsiankina. *The resumption of interrupted actions*. PhD thesis, Julius Springer Publishing House, 1928.

[159] Scott Rigby and Richard Ryan. Pens v1.6 – subscale scoring, 2007.

[160] Mary-Russel Roberson. Dance memory: Studying how the mind remembers physical movement. `https://research.duke.edu/dance-memory`, 2010. [Online; accessed 06-September-2020].

[161] DellaFave Robert. Designing RPG mini-games (and getting them right). `https://gamedevelopment.tutsplus.com/articles/designing-rpg-mini-`

games-and-getting-them-right--gamedev-14226, 2014. [Online; accessed 18-November-2020].

[162] Tim Rohrer. The body in space: Dimensions of embodiment. *Body, language and mind*, 1:339–377, 2007.

[163] Heath Rose. *The Japanese writing system: Challenges, strategies and self-regulation for learning kanji.* Multilingual Matters, 2017.

[164] Heath Rose. *Unique challenges of learning to write in the Japanese writing system*, pages 78–94. 05 2019.

[165] Michael Rowley. *Kanji pict-o-graphix: over 1,000 Japanese kanji and kana mnemonics.* Stone Bridge Press, 1992.

[166] Michael Rowley. *Kanji Pict-o-Graphix: Over 1,000 Japanese Kanji and Kana Mnemonics.* Stone Bridge Press, 1992.

[167] Jon Russel. Japan mulls ban on controversial but lucrative game genre. https://thenextweb.com/asia/2012/05/06/japans-mobile-gaming-industry-faces-a-ban-on-lucrative-kompu-gacha-games/, 2012. [Online; accessed 01-December-2022].

[168] Richard Ryan and Edward Deci. Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemp. Educ. Psychol.*, 25(1):54–67, 1 2000.

[169] Richard M Ryan, C Scott Rigby, and Andrew Przybylski. The motivational pull

of video games: A self-determination theory approach. *Motivation and emotion*, 30(4):344–360, 2006.

[170] William Ryan and Dennis Charsky. Integrating serious content into serious games. In *Proceedings of the 8th International Conference on the Foundations of Digital Games*, pages 330–337, 2013.

[171] Timothy Sanders and Paul Cairns. Time perception, immersion and music in videogames. In *Proceedings of the 24th BCS interaction specialist group conference*, pages 160–167. British Computer Society, 2010.

[172] Claudio Scolastici and David Notle. *Theory of Fun for Game Design, 2nd Edition.* Packt Publishing, 2013.

[173] Richard Ryan Scott Rigby. *Glued to Games: How Video Games Draw Us in and Hold Us Spellbound.* Praeger, 2011.

[174] Rhea Sharma, Edward F Melcer, and Dominic Kao. Exploring relevance, meaningfulness, and perceived learning in entertainment games. In *DiGRA*, 2022.

[175] Mike Shea. World of Warcraft and the Addiction of Task Lists. `https://mikeshea.net/World_of_Warcraft_and_the.html`, 2008. [Online; accessed 06-December-2022].

[176] Skritter. Skritter, 2009.

[177] Jolie HK Smets and Erik D Spek. That sound's juicy! exploring juicy audio effects

in video games. In *International Conference on Entertainment Computing*, pages 319–335. Springer, 2021.

[178] P. Smolen, Yili Zhang, and J. Byrne. The right time to learn: mechanisms and optimization of spaced learning. *Nature Reviews Neuroscience*, 17:77–88, 2016.

[179] Asji Software. Japanese Kanji Tree Pro, 2018.

[180] Lisa Son and Dominic Simon. Distributed learning: Data, metacognition, and educational implications. *Educational Psychology Review*, 24:379–399, 09 2012.

[181] Lisa K. Son and Dominic A. Simon. Distributed learning: Data, metacognition, and educational implications. *Educational Psychology Review*, 24(3):379–399, 2012.

[182] Rosetta Stone. Rosetta Stone, 1992.

[183] Terry Young Studio. Chinese Dungeon: Learn C-Word, 2017.

[184] PocApp Studios. Castle Cats - Idle Hero RPG, 2017.

[185] PocApp Studios. Dungeon Dogs - Idle RPG, 2020.

[186] Michael B Syson, Ma Regina E Estuar, and Karl T See. Abkd: Multimodal mobile language game for collaborative learning of chinese hanzi and japanese kanji characters. In *Proceedings of the The 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology-Volume 03*, pages 311–315. IEEE Computer Society, 2012.

[187] Wanda T. Wallace. Memory for music: Effect of melody on recall of text. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20:1471–1485, 11 1994.

[188] June P. Tangney, Roy F. Baumeister, and Angie Luzio Boone. High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. *Journal of Personality*, 72(2):271–324, 2004.

[189] TeamHoppi. Learn Japanese Yami, 2017.

[190] Techland. Companion for Dying Light, 2015.

[191] tofugu. Wanikani, 2011.

[192] Serkan Toto. How japanese mobile game makers go after whales: 5 popular gacha mechanics. `https://www.serkantoto.com/2016/03/14/gacha-monetization-japan/`, 2016. [Online; accessed 01-December-2022].

[193] Kathleen Tuite. Gwaps: Games with a problem. In *FDG*, 2014.

[194] Foreign Service Institute U.S. Department of State. Foreign language training. `https://www.state.gov/foreign-language-training/`. [Online; accessed 22-December-2022].

[195] A. Wald and CENTER FOR NAVAL ANALYSES ALEXANDRIA VA OPERATIONS EVALUATION GROUP. *A Reprint of 'A Method of Estimating Plane Vulnerability Based on Damage of Survivors*. Defense Technical Information Center, 1980.

[196] Wanda T Wallace. Memory for music: Effect of melody on recall of text. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(6):1471, 1994.

[197] Eddy Webb. Dopamine and zeigarnik: Game loops. `https://medium.com/@eddy.webb/dopamine-and-zeigarnik-game-loops-34ff2ae56960`, 2014. [Online; accessed 03-December-2022].

[198] Riah Werner. Music, movement and memory: Pedagogical songs as mnemonic aids. *TESOL Journal*, 9(4):e00387, 2018.

[199] Nick Wingfield. Free video games say pay up or wait, testing players' patience give this article. `https://www.nytimes.com/2014/07/06/technology/free-video-games-say-pay-up-or-wait-testing-players-patience.html`, 2014. [Online; accessed 16-December-2022].

[200] Brian M Winn. The design, play, and experience framework. In *Handbook of research on effective electronic gaming in education*, pages 1010–1024. IGI Global, 2009.

[201] John Woodhouse. Loot boxes in video games. `http://researchbriefings.files.parliament.uk/documents/CBP-8498/CBP-8498.pdf`, 2022. [Online. Accessed 29-November-22].

[202] Kelly Yap, Clement Zheng, Angela Tay, Ching-Chiuan Yen, and Ellen Yi-Luen Do. Word out!: Learning the alphabet through full body interactions. In *Proceedings*

of the 6th Augmented Human International Conference, AH '15, pages 101–108, New York, NY, USA, 2015. ACM.

[203] Haiwei Zhang. A review of stroke order in hanzi handwriting. *Language Learning in Higher Education*, 4:423–440, 10 2014.

[204] LI Zhe, Cheng Meng, and Maesako Takanor. A bibliometrical analysis of comparative studies on chinese hanzi and japanese kanji characters. *ARTS, SOCIAL SCIENCES*, page 3, 2013.

[205] Zynga. Farmville 2: Country Escape, 2014.

[206] Tina Čok. Ict-supported language learning tools for chinese as a foreign language. *Journal of Elementary Education*, 9(3):103–120, Jun. 2016.