

Social Situational Language Learning through an Online 3D Game

Gabriel Culbertson¹, Shiyu Wang², Malte Jung¹, and Erik Andersen²

¹Department of Information Science, ²Department of Computer Science, Cornell University
{grc74, sw673, mfj28, ela63}@cornell.edu

ABSTRACT

Learning a second language is challenging. Becoming fluent requires learning contextual information about how language should be used as well as word meanings and grammar. The majority of existing language learning applications provide only thin context around content. In this paper, we present work in *Crystallize*, a language learning game that combines traditional learning approaches with a situated learning paradigm by integrating a spaced-repetition system within a language learning roleplaying game. To facilitate long-term engagement with the game, we added a new quest paradigm, “jobs,” that allow a small amount of design effort to generate a large set of highly-scaffolded tasks that grow iteratively. A large-scale evaluation of the language learning game “in the wild” with a diverse set of 186 people revealed that the game was not only engaging players for extended amounts of time but that players learned an average of 8.7 words in an average of 40.5 minutes.

Author Keywords

language learning, video games, situated learning

ACM Classification Keywords

H.5.0. Information Interfaces and Presentation: General

INTRODUCTION

Finding the right balance of abstract learning and situated, communicative approaches has long been a concern for language learners. In an older investigation of “good language learners” teachers found that successful second language learners are willing to experiment in social interactions and make mistakes, but also spend a significant amount of time on abstract practice exercises such as flashcards [36]. Since these findings, the expansion of computer applications has made the abstract learning exercises much more accessible to learners. For example Duolingo [2] enables users to learn by translating isolated sentences and websites such as WaniKani [8] that allow for memorization of individual words.

However, designing a computer application for a situated approach is much more difficult. It requires shifting the focus towards creating social engagements that “provide the

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CHI 2016, May 7–12, 2016, San Jose, California, USA.

Copyright © 2016 ACM ISBN 978-1-4503-3362-7/16/05 ...\$15.00.
<http://dx.doi.org/10.1145/2858036.2858514>

proper context for learning to take place” [29]:14. A situated approach also emphasizes the practices involved in becoming skillful. Through his studies of Icelandic fishing practices Gisli Palsson explains the advantages of conceptualizing learning as active engagement with one’s environment:

“Becoming skillful means to attend to the task at hand, actively engaged with a social and natural environment. This suggests a notion of enskilmement that emphasizes immersion in the practical world, being caught up in the incessant flow of everyday life, and not simply, as many cognitive studies have assumed, the mechanistic internalization and application of a mental script, a stock of knowledge or a ‘cultural model’, ‘what one needs to know in order to behave as a functioning member of one’s society’” [32], citing [27], which cited [25].

Specifically in the context of computer applications for learning, Felix [20] argues that taking this perspective is essential to furthering language education in an online environment, but also identified key social, conceptual, political and pedagogical dilemmas that come with adopting a situated approach. It is impossible to fully realize this perspective in an application. However, the question remains, what can we learn from a situated perspective on learning to improve the design of language learning tools?

Some have taken the view that context is primarily visual. For example, *Rosetta Stone* uses images to give learners context without requiring translation. Another game, *Influent*, uses a virtual 3D world to help students learn vocabulary. Yet another tool, Micro-Mandarin [19], pushes the perspective further and uses real-world location for context.

However, we believe a purely visual approach is insufficient in creating a situated learning experience. As we argued above, learning and especially language learning depends on social engagement with others and when we lose the sense of who we are talking to and why we are talking to them, language learning becomes disconnected from reality. We therefore want learners to believe that they are interacting with other real people. To be more precise in our interpretation of situated learning, we describe this as design for social situational language fluency.

Games have a unique potential to provide situated learning experiences. Peterson describes these benefits, citing Gee: “Proponents of game-based learning argue that as games engage players in authentic situated problem solving, they facilitate learning that can be transferred to other domains [21]. Gee claims that as computer games enable language to be put in the context of dialog, they enable language to be situated.

This provides an ideal environment for language learning as verbal information is given just in time and is provided in an appropriate context [22]” [33]. Allowing students to engage and shape imagined scenarios can make them more real and engaging.

Furthermore, engagement is a key factor in second language learning. Games have tremendous potential to engage users. Applications such as DuoLingo [40] have made use of game elements to keep users engaged. Games give learners the ability to learn through experimentation rather than through instruction which can improve learning [11].

In order to explore the potential of design for social situational fluency, we created *Crystallize*. An early version of the game is described in [15], and focused mainly on the design and evaluation of joint quests that facilitated collaboration between two players in a highly constrained laboratory setting. This previous work left open the question of whether the ideas behind the game can succeed in the wild and provide a fruitful context for language learning independent of a highly structured laboratory test setting. This work also left open the question of whether we can develop game mechanics that are sufficiently engaging to motivate players who are playing under their own volition.

In this paper, we discuss the challenges and lessons learned from building a socially situated language learning game that can engage players for a long time. The challenges include 1) how to design and maximize the impact of long-term engagement mechanisms, 2) how to represent real-world situations precisely enough that players feel immersed, and 3) how to use game mechanics and player interactions to give the player the feeling of social engagement. We describe the iterative design and development of the game and two user-studies that tested the game in a controlled classroom setting and a final open field-setting on the internet. Through interviews with players, survey responses, direct observation of chat-logs, and analyses of quantitative gameplay data we found that the game engaged learners in socially rich and complex ways for an extended amount of time (40 minutes on average), and with a measurable improvement in language learning. This paper makes three core contributions:

- a language learning game that combines traditional learning approaches with a situated learning paradigm by integrating a spaced-repetition system within a language learning roleplaying game.
- a new quest paradigm to facilitate long-term engagement with the game through, “jobs,” that allow a small amount of design effort to generate a large set of highly-scaffolded tasks that grow iteratively.
- a large-scale “in the wild” evaluation with a diverse set of 186 people recruited through Reddit.

RELATED WORK

Language learning tools

Many language learning tools and websites have leveraged context to enhance learning. Some tools make use of content originally designed for native speakers for learning. In

order to enable learning while browsing the web in a second language, tools such as Rikaichan [6] allow dictionary look-up of words. Sites such as FluentU [3] make use of publicly available videos with subtitles for learning. Duolingo [2] also has a section where learners can work on translating content for intended for native speakers. These methods allow learners to interact with natural language and have some sense of context, but these methods do not allow for the learner to interact and receive feedback or provide motivation.

Other sites have taken a more social approach to language learning. Italki [4] uses a social network structure to match language learners with native speakers. Another site, Live-Mocha [5] encourages users to develop content in their native language for learners of that language to consume. For example, native speakers can create sets of vocabulary combined with audio recordings or provide one-on-one lessons. These methods allow for communicative learning, but the need to coordinate with teachers can make these interactions less accessible. Other work has looked at creating artificial social interactions through a chatterbot [37]. However, the effectiveness of this method needs further study. Furthermore, all of these methods provide only limited motivation frameworks.

Some games have also been developed for language learning to increase engagement for learners. For learning basic skills such as writing, games such as *My Japanese Coach* have been developed. Another game, *Influent* [28] teaches vocabulary in a virtual home and has game modes where the player must rush to find items in the home. These games, while they have some motivational framework, lack real or simulated interpersonal interactions.

Spaced Repetition Systems

Spaced repetition systems (SRS) make use of Ebbinghaus’ [18] finding that memory decays over time, but a memory takes longer to decay when it is reviewed at increasingly long intervals. Edge et al. [19] have provided a concise description of the process of learning using an SRS: “Unlike lesson-oriented learning, flashcards operate at the granularity of facts (e.g., word translations) and measure the learners ability to provide the correct response (e.g., a Chinese word) in the presence of a stimulus (e.g., an English word).” SRS are incredibly popular among second language learners. Systems as old as the Pimsleur system [34] and more recently sites such as Duolingo [2] and WaniKani [8] have made use of SRS. Some systems such *Anki* [1] and *SuperMemo* [7] are spaced repetition systems exclusively.

Teachers have examined the use of SRS in classrooms, but found it very difficult to fit SRS in with traditional curricula [12]. Classroom testing prioritizes recall in the short term, which incentivizes cramming strategies. Ebbinghaus’ [18] findings have shown these strategies to be inefficient in long term study. Thus SRS is best suited for learners who are self-motivated. We suggest, based on findings that that games can foster intrinsic motivation [16], that a game is an ideal context for SRS learning. Recently, a flash card game called *Picard* combined an SRS with social challenges to create mnemonic devices for remembering language material [39]. We build

on this work by addressing the challenges of merging an SRS into a game with a rich quest structure.

Roleplaying and learning

Language is best learned in through experiences. Barsalou [13] argues that people learn language in terms how it relates to real actions and outcomes. Memory of language is connected to memories of the physical surroundings at the site of learning and the outcomes of speaking. In a study by Glenburg et al. [24], it was shown that children manipulating physical figurines while reading or imagining manipulating the figurines led to increased memory and comprehension. Simply imagining scenarios while learning is enough to enhance learning.

Gee [23] drew a connection between these findings and learning in digital games. Games present fictional worlds and simulations. Although there is still a gap between reality and the game scenario, the game enhances the context around the content both by including a virtual world and associating language with an outcome. McGonigal [31] argues that for serious players, there are instances where the lines between game worlds and the real world start to blur.

Some studies have explored the idea of leveraging immersive online games for language learning. In a review of studies of online roleplaying games used to improve language ability [33], Peterson showed that these games have the potential to improve language skills. In one of the studies [35] discussed, intermediate English as a Second Language (ESL) students vastly improved their vocabularies and confidence by playing *Everquest II*. In our previous work on the prototype for *Crystallize*, we found that players with no prior language experience could learn in a collaborative game environment. However, because previous work has been in a lab setting, we lack information about how game design decisions affect experiences in the wild, how engaging these methods are, and what game features are useful to learners.

GAME DESIGN

Overview

Crystallize uses the rich context of a 3D environment with both human and computer controlled characters to create socially situated language learning experiences. Players collect words that they see computer-controlled characters use in the game and complete language puzzles with these collected words to progress in the game. The game also offers players the possibility for social interaction and collaboration with other players in the virtual environment. The game can be downloaded and tried on the game website: crystallize-online.com/.

We designed the game to engage learners of Japanese who are tech-savvy and interested in using online tools and games to improve their language skills. We designed the mechanics to have nuance so that learning the game would be engaging in and of itself. However, this complexity makes the game difficult to learn for very young learners. The game is designed to be engaging for anyone old enough to understand the rules. Furthermore, we assume that learners are doing

some learning independently, but have difficulty structuring their learning and staying motivated. Thus we provide both a motivational framework and scaffold the learning.

Our first priority is to provide the player with a framework of goals, obstacles and tools to overcome those obstacles. Goals are provided using a questing system. While exploring the world, players can find NPCs (non-player characters or computer controlled characters) who offer tasks for the player to complete. For example, a NPC may ask the player “what is this?” while looking at a bicycle. The player must search the area for other NPCs that can tell the player how to say the word bicycle. All of the quests involve the player responding to prompts using the words that they have collected.

In order to simulate social interaction through conversations using game mechanics, we introduce a system where players construct phrases in order to interact with NPCs. These interactions enable the player to complete quests. First, the player must have the correct words to respond to the prompts provided by the NPCs. In order to better simulate conversational flexibility, there are usually multiple correct responses, and choosing a more complex response will yield a greater reward, but at minimum, the player must be able to construct the simplest response. For example, in Japanese, the subject can be excluded from the sentence resulting in a simple 2-word response to the bicycle prompt: “*jitensha desu*” (literally is bike). The player may also choose a more complex response for a greater reward: “*kore ha jitensha desu*” (this is a bike).

In order to encourage reviewing and integrate an SRS into a seamless gameplay experience, we introduced a new game mechanic called *confidence*. When the player sees words that they have not seen before or responds incorrectly to prompts, the player’s confidence is reduced. The player begins with some amount of confidence and when the player’s confidence score falls to zero, the player is ejected from the current conversation and sent to their home where they can review words. Each time an NPC uses a word that the player has not yet collected, the player’s confidence will drop by 1 point. The player can recover that confidence by collecting the word, but because the player’s inventory is limited, the player will not be able to complete some conversations before collecting enough words beforehand. Choosing a less desirable response (e.g. if the NPC asks “Can you do this job well?” and the player responds with “Um...”) will also cause a reduction in confidence.

To complete quests, the player must (1) collect words and phrases, and (2) review those words and phrases. Collecting words can be accomplished by listening to dialogues of pairs of NPCs, or through the quest conversations. Strategically choosing words that appear more frequently can help the player complete challenges more quickly, but there is no penalty for collecting more words than necessary. Reviewing words and phrases can increase the amount of confidence that the player begins a level with. Therefore, even without having the correct words, players that have spent a long time with the game can complete more difficult quests.



Figure 1. An early prototype of *Crystallize* as used for the laboratory study described in [15]. Because conditions needed to be controlled for the laboratory setting, the game was linear and lacked mechanisms needed for long term motivation and learning.

Iterative development

Crystallize was developed and tested in three stages. First, an early version of the game was designed with a focus on creating a short immersive language learning experience that can be studied in a laboratory setting and is described in [15]. During the second design stage that we report on here we redesigned the game to include mechanics and content to support long term play for an in-class deployment. The goal of the third and final design stage, the outcome of which is the main focus of this paper, we redesigned *Crystallize* to support self-motivated learners on the web and to be conducive to large-scale distributed data collection.

Stage 1 - *Crystallize* for laboratory deployment

The first prototype of the game was designed for preliminary exploration of the idea for the game in a laboratory based user study. Through the user study we found that participants without any prior Japanese knowledge who played the game for 25 minutes were able to learn, but perceptions of how engaging the game was were highly varied depending on the specific participants interests and background [15]. This initial implementation had several drawbacks. First, the game was highly linear. The game was designed to be played in a lab setting, so we needed participants to follow similar progressions through the game. This meant that players lacked the freedom to explore specific areas of interest in depth. Second, the game lacked mechanisms needed to sustain long-term play. Although players were able to complete challenges in the game, players retained relatively little in the short-term, and we lacked data to make any conclusions about long-term retention. The previous prototype also lacked mechanisms to support long term motivation. The players only goals were to complete the current level and proceed to the next. In the following iterations, we explored methods for addressing these issues.

Stage 2 - *Crystallize* for classroom deployment

Our main design goal in this iteration was to create mechanics and content to support long term play and engagement. The iteration culminated in a two-week long user study for which we deployed the game in a classroom setting. We recruited 26 high-school aged students from a summer learning



Figure 3. While exploring the world encountering new words or making language mistakes results in reduced confidence (top). When the player runs out of confidence or chooses to go home, they can review (center) to restore confidence (bottom).

program at Cornell University. These students were recruited by email before the learning program began. All of the students came from a single class, but participation was optional and not all students participated. The study was conducted in ten 30 minutes sessions over three weeks. Throughout the study, we conducted interviews with the students, and responded to student comments by adjusting the design. The feedback from this group led to the addition of aesthetic and game-mechanic-related rewards, additional elements of the Japanese language (Hiragana, Katakana and Kanji scripts), and implemented mechanics to reduce repetitiveness.

Stage 3 - *Crystallize* for field deployment

Our primary goals for this iteration were to enrich social interaction in the game and give players more freedom in the game environment. Unlike the classroom study where participants were co-located, the online release of the game required additional features so that players could have a sense of social presence. For this, we expanded on the ideas of the initial prototype. Unlike the initial lab study, we had no control over participants' play schedules and play time. Thus, rather than connecting two players in an interdependent relationship, we designed a more open interaction system where players could come and go as they pleased and have the freedom to interact with players if they desired. This change also required the redesign of the level structure from the earlier prototype. In order for players to have the opportunity to encounter other players, we required a single, central hub for players gameplay rather than a linear progression from one level to the next.

FIELD STUDY

We conducted a field study to evaluate the three design challenges outlined in the introduction. Each challenge has one or more research questions associated with it. The first challenge was about long-term motivation and learning and ad-

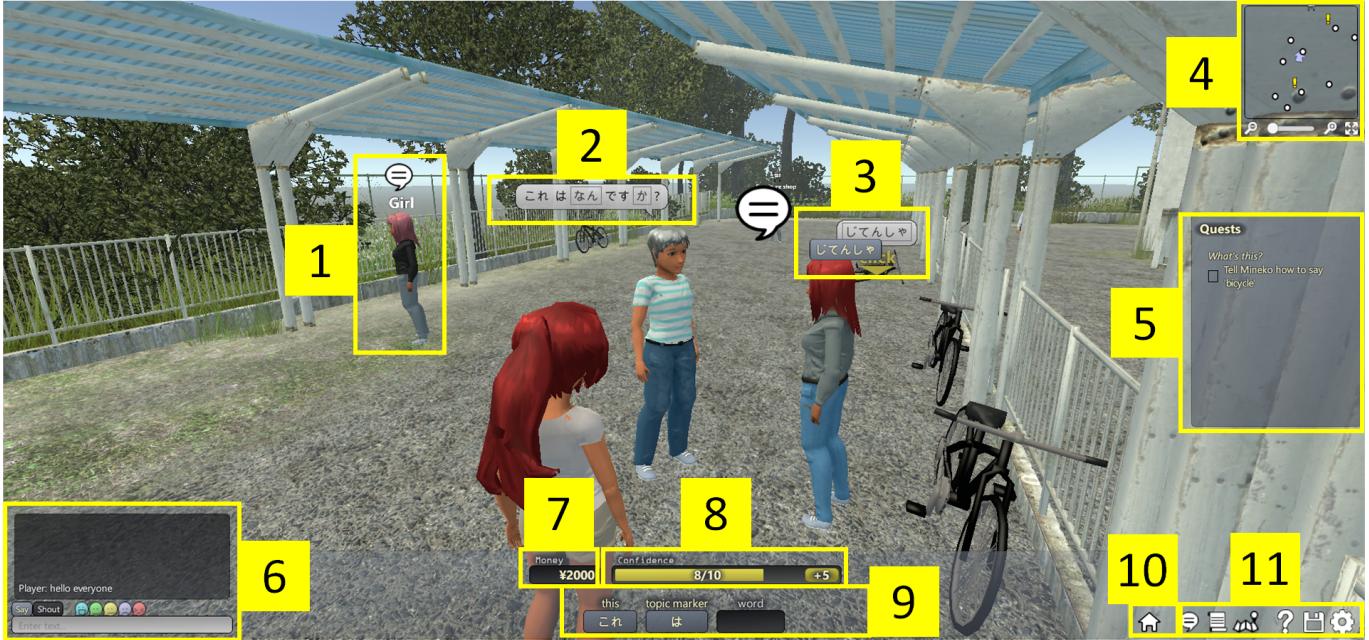


Figure 2. In *Crystallize* players explore an open world where they can find and talk to NPCs (1). Players can overhear conversations of groups of NPCs. Speech bubbles show what the NPCs are saying and words that the player has not collected yet are shown as outlined (2). These words can be dragged from the speech bubbles (3) to the player's inventory (9). A map shows the locations of normal NPCs and quest NPCs relative to the player's location (4). Progress on current quests is shown in the quest panel (5). Players can chat with other players using the chat box (6). The players money (7) and confidence (8) are displayed. The player can see how many reviews they have currently available and go home using the home button (10). Other settings can be adjusted and the game can be exited using the control panel (11).

dressed the following research questions: How does the game engage an audience of language learners in the field that do not have the incentives and context that a laboratory study and classroom deployment provide (R1)? Would we be able to measure actual improvements in language learning skills given that we have no control over the makeup of the players in comparison to our laboratory and classroom studies (R2)? The second challenge was about creating immersive as well as situated learning experiences and addressed the following research question: What benefits do learners perceive in a socially situated approach to learning (R3)? The third challenge was about social engagement and addressed the following research question: Would the game give learners the experience of social engagement (R4)?

Reddit Deployment

The game was made publicly available on the web, and advertised using a forum post to the Japanese Learning sub-Reddit (www.reddit.com/r/LearnJapanese). We chose Reddit as a deployment channel because we believed we could reach our desired audience this way. Sub-Reddit communities are focused and active, and the Reddit website provides a simple interface for providing comments and upvotes.

To play *Crystallize*, players clicked the link in the post that led them to the game website. Upon starting the game, players were informed that the game was part of a research project and consented to have their gameplay data collected. After providing a username and password, participants could create a customized avatar. During the 5 to 10 minute tutorial, users could not see each other or chat with other users. After the

tutorial players were free to explore the world and complete quests in an order chosen by the player. At this point, the chat tool was also unlocked and other players in the game world were made visible.

Before beginning the actual game, a prompt appeared asking participants to complete an optional survey. Surveys were completed through a web browser. Participants provided information about perceptions of language learning (e.g. Learning language is easier in a classroom, Language learning causes anxiety) and attitude toward learning Japanese (e.g. I feel excited about learning the Japanese language). All responses used a seven point Likert scale.

Following 15, 30 and 60 minutes of play, participants provided their general impressions of the game (e.g. How much did you like the game?) and rated how well certain words described their experiences (e.g. frustrating, motivating, fun, exciting). The survey then gave participants the opportunity to provide free response feedback. Finally participants were prompted to provide their email if they were willing to participate in a follow-up interview.

The follow-up semi-structured interviews inquired how learners viewed the game relative to other tools they used (e.g. How do other methods you have used compare to the game?), motivation to play the game (e.g. If you were motivated to play, was there something about the game that motivated you?), immersion (e.g. What moment in the game did you feel the most immersed? Least immersed?), how players felt about other players (e.g. How did other players in the game impact your game experience?) and about the mechanisms

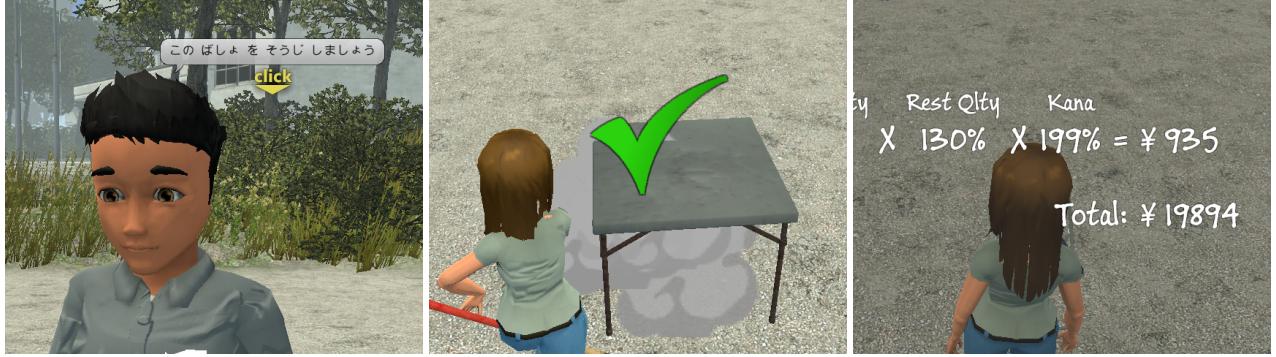


Figure 4. Players complete jobs by first initiating a conversation with the “boss NPC” (left), then the player must clean the item specified by the boss (center) and after completing several iterations receives a reward based on performance and progress in the game (right).

| Quest or job class | Description |
|------------------------|---|
| Introduction quest | The player learns basic introductions and how to interact with NPCs in the game |
| Learning from context | In these quests, the NPC will prompt the player to tell him or her how to identify a thing in the 3D world. For example, the NPC may ask “what is this?” while looking at a tree. The player then goes to other NPCs in the game world and listens to them discussing the thing to learn that word. The player then returns to the original NPC and tells them how to say the word. |
| Person finding | Player learns to identify people using descriptive characteristics. The prompts are initially very simple. For example: “I lost my friend, he is a boy.” The player would then need to search nearby for a male NPC. Later prompts become complex. For example, the quest prompt may become “I lost my friend. He has red hair and is wearing a T-shirt”. |
| Get-to-know-someone | The player learns how to get to know someone in Japanese. Initially, the player will go through an introduction sequence. If the player uses language that is too informal, or does not respond politely to the NPCs prompts, the player will be penalized. After the initial meeting, the player will randomly find the same NPC again in the game world. In following meetings, the NPC may request information from the player (“do you remember my name?”, “do you remember where I’m from”) or allow the player to ask for new information (“what’s your hobby?”, “how old are you?”). |
| Take someone somewhere | In this type of quest, the player takes a NPC to a specific location. Initially the NPC prompts the player with a specific location. For example, the NPC may say “I need to go to the restroom, can you take me there?”. Later, the NPC will make the request more indirectly. For example, “I’m late to class!”. The player completes the quest by navigating to the correct location and telling the NPC that they have arrived. |
| Chain of information | Players must follow clues provided by NPCs in sequence to solve problems. For example, an NPC may prompt the player with “I can’t find my cat... I wonder if Sakura has seen it.” The player must then find Sakura and after talking with her she may say “I know where the cat is, but I will only tell you if you help me find out when Kenji’s birthday is”. The player follows the clues until they reach the final goal. In the example given, this would be finding the cat. |
| Janitor job | The job is started by completing a kind of interview. By making it difficult to begin the job, we provide a gating mechanism to prevent players from doing the job too early while giving the player the freedom to try if they wish. If the player lacks sufficient confidence they will not be able to complete the interview. Once the interview is passed, the boss NPC will tell the player to clean specific items. For example, the boss may ask the player to “clean the table”. After doing the job several times, the items that can be cleaned are expanded. |
| Cashier job | Similar to the janitor job, the player must first complete an interview before beginning the job. Once the job is started, people will appear who want to buy things. The player must tell the NPC the correct amount of money needed for the items they wish to buy. As the player advances in the job, NPCs will request more complex combinations of items. |

Figure 5. Description of activities in the game. After the first quest, players are free to do any of the activities listed here in any order.

for long term learning (e.g. How did you feel about the review system in the game? Did you find it helpful? Annoying?).

Throughout the 15 days during which we collected data, a total of 186 participants created accounts. 69 participants filled out the initial survey and 23 participants filled out the survey in one of the later prompts. Seven participants filled out the survey both at the initial prompt (before starting to play) and at one of the later prompts (15, 30, and 60 minutes into the game). 22 participants provided their email addresses which resulted in 6 participant interviews.

Measures

We collected data through gameplay logs, chat-logs, surveys and interviews to understand (1) players’ engagement with the game and (2) learning outcomes. Comparisons between groups were made using independent sample T-tests and correlations were found using Pearson’s r.

Measures of engagement: Engagement was measured using the total amount of time that players spent in the game and the number of times the player returned to the game after waiting an hour or more. Many previous “in the wild” experiments with games have used play time to measure engagement [14, 30, 11, 9, 10]. Since players can play many other free games or use other language learning tools if they lose interest in this one, getting players to spend a significant amount of time on a game suggests that the game is engaging.

Measures of progress in the game: After the initial tutorial section of the game, players are free to pursue any content they wish. Thus we measure progress as the number of words and phrases that the player has collected.

Measures of learning: The spaced repetition system used in the game also serves as a continuous test of learning. Each review in the spaced repetition system is a multiple choice quiz

where the player must choose the correct English meaning for a Japanese word. We acknowledge that multiple choice questions are not ideal for measuring learning, but the focus of the study was to collect as much information as possible from participants, so we did not wish to drive participants away with mandatory pre and post tests. Thus we use these scores to show evidence that learning occurred rather than a measure of the amount of learning.

For this study, we define learning as the memorization of previously unknown word or phrase meanings. Thus for each word or phrase we first evaluate whether the learner knew the word prior to playing the game. We do this using the first two SRS reviews. Because the review system is multiple choice with 5 choices, there is a 1 in 5 or better chance that the learner can guess the correct answer without knowing the translation. By taking the first two reviews we have a finer value for the probability that the learner knew the word prior to playing the game. All of the following reviews in the spaced repetition system are used to calculate the probability that the learner knew the word after playing the game. Note that this is a conservative measure, because it is possible for the player to see the definition of a word before beginning the review session, so answering incorrectly requires that the player both not know a given word or phrase beforehand and forget the meaning during the first review.

Measures of prior knowledge: A prior knowledge score was calculated by taking the total number of reviews that the player responds to correctly divided by the total number of reviews seen to construct *review accuracy*. We would expect those with less prior knowledge to receive lower scores.

Measures of social engagement: Previous work ([17]) has shown that social engagement is a key motivator in players of online roleplaying games. It can be enough just to see that others are playing the game. Thus we look at players in two groups: those that were online at the same time as another player and those that only played alone. We also take lines of chat to be an indicator of social engagement.

RESULTS AND DISCUSSION

Expert players

We found that many of the players had extensive prior knowledge of Japanese before beginning play. The 58 participants who took the pre-test had an average score of 16.31 ($SD=10.67$) out of 30 possible on words and phrases available in the game. To better understand the behavior of expert and non-expert behavior, we divided the data in to groups using *review accuracy*. Those with greater than 90% accuracy were considered experts and those with less than 90% accuracy were considered non-experts.

There are 98 experts and 88 non-experts. Number of words learned per hour was higher in non-experts ($M = 12.1$, $SE = 1.6$) than in experts ($M = 5.8$, $SE = 0.4$, $p < 0.001$). In survey responses, enjoyment of the game was found to be higher in experts ($M = 6$, $SE = 0.3$) than in non-experts ($M = 4.6$, $SE = 0.3$, $p = 0.010$). Playtime was also more than three times as long for expert players ($M = 65$ minutes, $SE = 7.4$) than non-expert players ($M = 17$, $SE = 3.0$, $p < 0.001$). Return rate was

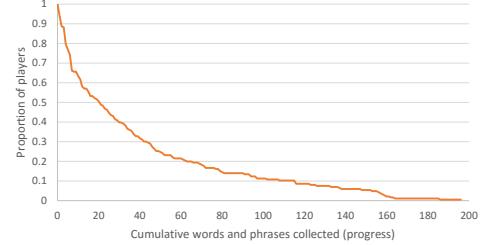


Figure 6. Graph of progress for all 186 players. The x-axis shows numbers of words and phrases collected, and the y-axis shows the percentage of the players who collected at least that many words and phrases. This graph shows that players were motivated to collect many words; the top 20% collected more than 60.

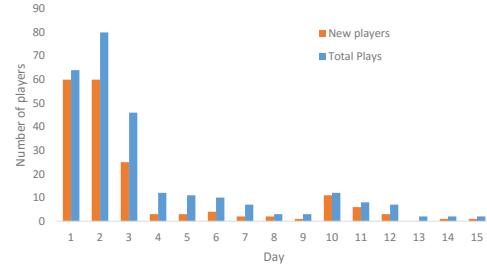


Figure 7. Players over the duration of the study. The majority of the players came in the first two days when the Reddit post was on the front page of the sub-Reddit.

significantly higher in experts ($M = 22\%$, $SE = 2.4\%$) than in non-experts ($M = 6\%$, $SE = 4\%$, $p = 0.002$).

We believe integrating the efficiency-driven SRS with the contextual learning offered by exploring the game environment creates an effective hybrid learning approach. Many learners were already doing this. All but one of the interviewees mentioned that they had used some SRS in the past, the most common being *Anki*.

Evidence that the game is engaging

Addressing R1, we examined various data sources for indicators of engagement. The game was received very positively by the Japanese learning sub-Reddit community. The game thread received an up-vote minus down-vote score of around 115 (for comparison, the posts on the front page of the same sub-Reddit at the time of writing have an average score of 8). There were over 60 comments posted by the sub-Reddit users and half of the people who played the game responded to at least one of the optional surveys. The 186 players spent an average of 40.5 minutes playing the game. Twenty-five (13%) of the players returned to the game after taking a break. We define a break as being logged off for at least 1 hour. For the players that returned at least once, those players returned an average of 2.36 times ($SE=0.443$). The players that returned at least once played an average of 142 minutes ($SE=2.426$) while the players that played the game only once played for an average of 24 minutes ($SE = 15.925$). Charts of play time and progress can be found in Figures 9, 6, and 7.

Many of the comments on Reddit thread were supportive. For example these comments were posted on the Reddit thread:



Figure 8. Heat map of player movement. A value closer to red indicates that more time was spent by players in those areas. Players began at (1) and we also see a lot of activity around (2) and (3) which were the popular jobs: janitor and cashier.

*This game is really fun; originally, I felt burdened by learning vocabulary (especially in Japanese, when phrases are so long) but this game makes it very exciting to learn new words and phrases! and I'm actually really enjoying this game, it even covers some vocab that isn't in my textbook!. These comments were often made despite a long list of issues that the player may have had. Many participants also expressed a long term interest in the project. We also take the high number of participants who responded that they were willing to be contacted for follow-up (22 of the 92 survey-takers) to be indicative of engagement. Some participants asked us to set up a way to communicate updates. For example this was posted on the Reddit thread: *It looks very promising and interesting, but I really miss a changelog on the website. I want to know every bug that was patched and every new feature!**

Of the those who completed the survey after playing the game, 11% said they did not like the game (7-point Likert scale with a score of 3 or less), 18% were neutral (score = 4) and 71% liked the game (score > 4).

We also find evidence for support of the game in the chat log:

maya: I hope this game can become great, though
 Breathless: yeah, I hope so
 maya: eee
 Breathless: it's looking great
 Breathless: I hope it doesn't die
 ...
 username: wish their was a korean ver of this
 username: would try

Evidence of learning in the game

To gain insight into research questions 2 and 3 that addressed learning, we performed quantitative analyses into the gameplay data. Players collected an average of 29.75 words ($SD=33.83$) and 7.43 phrases ($SD=11.86$). For words that player did not know previously, we find a significant difference between review accuracy before ($M=0.548$, $SD=0.142$) and after ($M=0.873$, $SD=0.149$) the initial set of reviews, $t(184)=-14.68$, $p<0.001$. Learning gains were greater for those who returned ($M = 18.3$, $SE = 2.7$) than for those who did not ($M = 5.6$, $SE = 0.7$, $p < 0.001$). Using our learning measure, players learned on average 8.7 words and phrases

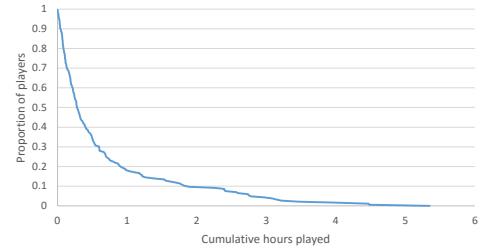


Figure 9. Graph of time played for all 186 players. The x-axis shows lengths of time since starting the game, and the y-axis shows the percentage of the players who played for at least that much time. This graph shows that many players were motivated to continue playing for a long time. 20% of players played for more than one hour.

($SD=10$). The large difference between words and phrases collected and words and phrases learned is due to three factors: (1) players entered the game with extensive prior knowledge, (2) our learning measure was conservative and (3) we lacked sufficient data to conclude that some of the words had been learned. As we would expect, this is strongly correlated ($r(184)=0.620$, $p < 0.001$) with time spent playing the game. We also measured learning using the pretest and posttest in the survey, but because the surveys were optional and could be taken at any time, only 6 participants completed both the pretest and posttest. All of the participants who took both tests had an increase in score, but the difference was not significant, likely due to the small sample size.

Evidence that social aspects improved the experience

Our analyses also provided insights into R4 and R5. Although many of the users did not make use of the social features of the game or were always online alone, we believe that designing for social play greatly enhanced the game for some players. Length of play was significantly higher for those learners who entered a line of chat in search of someone to talk to ($M = 78.61$, $SD=62.64$) than those who did not ($M=32.97$, $SD=53.27$), $t(184)=-4.120$, $p < 0.001$. We looked further at players who found another player to chat with, and those that found someone had longer average playtime, but the results were not statistically significant. We speculate that this is because the number of players who found chat partners was small (only 8),

Although none of the interviewees had found chat partners, we do find evidence that the players liked the social of others in the game. *I occasionally ran in to other people, but we didn't talk. It was encouraging to see what other people were doing. ... It would be cool to share the words I learned with other people.* One of the interviewees also speculated on the feature: *it seems like most players in the game have learned some Japanese already so I feel that I can always ask them for help.* We also believe that the shared virtual space enhanced social interactions. Players who saw other players' avatars were far more likely to chat. Furthermore the shared worlds provided grounding for some of the chat: *can't you see the isu here?*. Other chat referred to the in-game jobs: *Gonna try work for the Janitor so I can get some cash.*

Evidence that players prefer roleplaying

We find evidence in the chat logs and survey that players the game reframed activities to have more meaning for players. In the game, rather than simply reading text, the player is “having a conversation” and rather than choosing items from a list, the player is “doing a job”. We believe that our players felt that this was qualitatively different from other learning experiences. One interviewee said: *you have quests that make you go talk to people and have a short dialogue with them. I think that's really useful and am glad that the game wasn't just about learning vocabulary like others I've seen in the past. It feels immersing; I really do enjoy it.* Note that although the some of the activities may be similar to to activities in other language learning applications, the language that players use to describe these actions are different. The player says that he was “talking to people”, rather than than taking a quiz.

There is also evidence that situating learning using the game world makes learning more meaningful to participants. One of the interviewees described the experience like this: *As a person learning Japanese in a relative vacuum, the illusion of interacting with people who speak the language is valuable, and offering the opportunity to suspend disbelief (that is, I'm not simply filling in the blanks, I'm talking to some dude who has lost some other dude) is useful.* Some participants compared the game to Rosetta Stone. One said: *I tried Rosetta Stone for a while, and I memorized a lot, but I didn't learn anything, you know? and then games have a way of showing you something in an interesting way.*

Evidence that simulating situations was motivating

We have evidence that the questing structure of the game encouraged learning. Feedback that we got from participants in the Job prototype study after adding additional reward mechanics was entirely positive. Some participants specifically mentioned that they felt motivated to earn money for new clothes.

In the version of the game released to Reddit, we found instances of players doing jobs in order to earn money in the chat log:

maya: are you playing the janitor thing?
Breathless: to get the new vocab from him (Breathless mistakenly answers the question “why are you playing the janitor job”)
Breathless: and money to buy clothes
...
username: clothes arent very kawaii desu
Dan: Gonna try work for the Janitor so I can get some cash.

There was also feedback in the Reddit thread and interviews that drew parallels between *Crystallize* and the highly engaging online roleplaying game *World of Warcraft: props for the idea, even if the yellow quest markers kind of give me bad flashbacks to being addicted to World of Warcraft. ;)* One interviewee said: *it reminds me of WoW, I like doing quests.*

The rewards that we added only had meaning within the virtual world. The fact that players were motivated to earn these

rewards suggests the players were engaged in the imagined world and scenarios.

Evidence for improved social situational learning

By examining the chat logs, we found that there were instances of of players first learning phrases through quests and then using them in conversations with other real players. This finding provides additional insight into R4. For example, Breathless first used the Japanese phrase for “How old are you?” and “what’s your hobby?” in quests and then practiced the newly acquired skill in a social context with a real other player :

Breathless: anata ha nan sai desu ka? (how old are you?)
maya: watashi? (me?)
maya: himitsu desu! (it's a secret!)
...
Breathless: anata no kyoumi (your hobby...)
Breathless: ha nan desu ka? (...what is it?)
maya: e wo kaku (painting pictures)

There are around 20 instances of people using the input system in the game (meaning they first collected the words using the game mechanics) to produce Japanese greetings or farewells in social contexts with other real people. One of the interviewees also noted that they used some of the words that they learned in the game in conversations with their real-world friends.

LIMITATIONS

We designed the game to engage learners of Japanese who are tech-savvy and interested in using online tools and games to improve their language skills. This target population is certainly not representative of all language learners, but the Reddit audience reflects this target population very well. Of survey respondents, 48% agreed that learning a second language is best using an application (when compared to classroom or learning abroad) and 41% were neutral between learning through an application and other methods.

Although the return rate of 13% may seem low to readers, it is not unusual in the world of free online video games. In fact, a study of *Hello Worlds!*, a non-educational video game that has been played 1.5 million times, reported a return rate of about 20% [11]. Of course, we would still like to see this improve in future iterations.

Although the system was designed to support long term engagement, in this particular study, we lack sufficient data to answer the question of whether the long-term engagement mechanisms are effective or not. Future work will examine this question in more detail.

CONCLUSIONS AND FUTURE WORK

The problem we set out to address was to build a game that promotes social situational language fluency. We highlighted three major challenges that such a game should address: 1) how to design and maximize the impact of long-term engagement mechanisms, 2) how to represent real-world situations precisely enough that players feel immersed, and 3) how to use game mechanics and player interactions to enable social

engagement. To address these challenges, we presented a language learning game that combines traditional learning approaches with a situated learning paradigm by integrating a spaced-repetition system within a language learning roleplaying game. The game features a new quest paradigm to facilitate long-term engagement with the game through “jobs” that allow a small amount of design effort to generate a large set of highly-scaffolded tasks that grow iteratively. Finally, a large-scale evaluation of the language learning game “in the wild” with a diverse set of 186 people revealed that the game was highly engaging for some players. By examining interviews, surveys and chatlogs we found that players were engaged by the roleplaying aspects of the game (R1). Players learned an average of 8.7 new words, despite most players beginning with extensive prior knowledge (R2). Interviews revealed that players felt this method of learning was more useful than more abstract methods because players imagined they were participating in real-world activities (R3). Finally, chatlog analysis and the interviews revealed that only a limited number of players felt socially engaged, but that subset of the players benefited greatly from the presence of these features (R4). The study also revealed three insights that give promising directions for future work.

Supporting community outside of the game

Releasing the game through Reddit allowed us to engage with the community and troubleshoot issues during the study. It is possible that this engagement with the community encouraged participants to provide more depth in their comments. After helping one Reddit user with an issue, he commended our active participation in the community: *Thank you very much, haven't seen a whole lot of your game yet (but I am sure I will) but you have a damn fine "customer" support so far, lol. ;)* We can also see evidence of our community presence in the chat log:

maya: Did you find this from a reddit thread?
maya: they seem to be working hard to mend problems
...
maya: Maybe report the chat box thing
Breathless: I'll do

The Reddit thread also made it possible to avoid an issue that may have alienated many users. Many of the learners who wanted to try the game already had experience with Japanese and therefore already knew some combination of Hiragana, Katakana and Kanji. Although we allowed Hiragana and Katakana to be unlocked through the game, these learners felt that using the English was too much of step backward. The top comment in thread was *Is it only in romaji? That seems to be a bit of a damper on the “immersion” aspect....* We caught this comment early on and added an option to change the game into Kana initially.

Designing for meta players

When asked about why they came to the game, some interviewees said that they were actually more interested in exploring the ideas and design of the game than using the game for learning. For example, one interviewee when asked about why he played the game said: *I think much of my interest*

came from curiosity toward how you were going to implement educational ideas. There was also an interviewee and one poster that mentioned that they had similar ideas for a game and were really excited to see the game being implemented. Many comments we received acknowledged the underlying rationale behind the game mechanics. For example, one interviewee mentioned an annoyance with the limited inventory: *I understand that this is a gating mechanic for language acquisition, but an alternate method needs to be developed to avoid having to pop in and out after every conversation.* This suggests that in designing the game, we can also consider engaging these expert learners by allowing them to contribute to the game design for example through game modifications or mods.

Designing for learning ecosystems

We found that many of the learners did not rely on a single tool or approach for language learning by rather on a complex ecosystem of tools and social engagements such as the sub-Reddit we released our game in. All 6 of the interviewees had tried to learn using more than 5 different approaches (classes, software, Influent, Anki, speaking with native speakers, lang8, pen-pals, to name a few) and some mentioned more than 10 throughout the interview. This finding is in sync with the findings of Townsend et al. [38] who describes the shift in education towards learning from many sources.

Furthermore, interviewees specifically suggested making *Crystallize* more integrated with these larger learning ecosystems. Three separate interviewees recommended that we integrate the game with *Anki*. By integrating the review system in the game with something more portable (such as *Anki*), as one of the interviewees pointed out, it becomes possible to manage the reviews while on vacation using a mobile phone. Another interviewee noted that this would make doing *Anki* reviews much more fun because they would be linked to progress in the game.

The embeddedness of learning tools into an ecosystem also has important implications for how we assess language learning. Learning from many sources means that learners each have qualitatively unique knowledge which means personalization requires more nuance than a simple placement test. For example, one interviewee previously learned Kanji (Chinese characters in the Japanese language) using *Remembering the Kanji* [26] which teaches the meanings of Kanji without teaching pronunciation. There was a similar scenario in the job-driven version of the game: many participants came from China and could understand the meaning of a character without being able to pronounce it.

We believe that this presents an interesting design challenge of designing not just for learning of the language, but also design for integration into larger learning strategies. The many requests for integration with *Anki* highlight this need. Furthermore one interviewee suggested creating challenges in the game that would require the player to search outside the game in order to solve it.

REFERENCES

1. 2015. Anki. www.ankisrs.net. (2015).

2. 2015. Duolingo. www.duolingo.com. (2015).
3. 2015. FluentU. www.fluentu.com/japanese/. (2015).
4. 2015. italki. www.italki.com. (2015).
5. 2015. LiveMocha. <http://livemocha.com/>. (2015).
6. 2015. Rikai Chan. www.polarcloud.com/getrcx/. (2015).
7. 2015. Super Memory. supermemo.com/. (2015).
8. 2015. WaniKani. waniKANJI.com. (2015).
9. Erik Andersen, Yun-En Liu, Richard Snider, Roy Szeto, Seth Cooper, and Zoran Popović. 2011. On the Harmfulness of Secondary Game Objectives. In *FDG '11: Proceedings of the Sixth International Conference on the Foundations of Digital Games*. ACM, New York, NY, USA.
10. Erik Andersen, Yun-En Liu, Richard Snider, Roy Szeto, and Zoran Popović. Placing a Value on Aesthetics in Online Casual Games. In *CHI 2011*. ACM Press.
11. Erik Andersen, Eleanor O'Rourke, Yun-En Liu, Richard Snider, Jeff Lowdermilk, David Truong, Seth Cooper, and Zoran Popović. The Impact of Tutorials on Games of Varying Complexity. In *CHI 2012*. ACM Press.
12. Richard Bailey and Jesse Davey. 2011. Internet-based Spaced Repetition Learning In and Out of the Classroom: Implementation and Student Perception. *CELE Journal* 20 (2011), 39–50.
13. Lawrence W Barsalou. 1999. Language comprehension: Archival memory or preparation for situated action? (1999).
14. Eric Butler, Erik Andersen, Adam M Smith, Sumit Gulwani, Zoran Popovic, and WA Redmond. 2015. Automatic Game Progression Design through Analysis of Solution Features. In *Proc. of the SIGCHI Conf. on Human Factors in Computing (CHI2015)*.
15. Gabriel Culbertson, Erik Andersen, Walker White, Daniel Zhang, and Malte Jung. Crystallize: An Immersive, Collaborative Game for Second Language Learning. In *CSCW 2016*.
16. Michele D Dickey. 2007. Game design and learning: A conjectural analysis of how massively multiple online role-playing games (MMORPGs) foster intrinsic motivation. *Educational Technology Research and Development* 55, 3 (2007), 253–273.
17. Nicolas Ducheneaut and Robert J Moore. The social side of gaming: a study of interaction patterns in a massively multiplayer online game. In *CSCW 2004*. ACM Press.
18. Hermann Ebbinghaus. 1913. *Memory: A contribution to experimental psychology*. Number 3. Teacher's College, Columbia University, New York City. Originally published in 1885.
19. Darren Edge, Elly Searle, Kevin Chiu, Jing Zhao, and James A Landay. MicroMandarin: mobile language learning in context. In *CHI 2011*. ACM Press.
20. Uschi Felix. 2005. E-learning pedagogy in the third millennium: the need for combining social and cognitive constructivist approaches. *ReCALL* 17, 01 (2005), 85–100.
21. James Paul Gee. 2003. What video games have to teach us about learning and literacy. *Computers in Entertainment (CIE)* 1, 1 (2003), 20–20.
22. James Paul Gee. 2007. Why are video games good for learning. *Unpublished manuscript*. Retrieved 23 (2007).
23. James Paul Gee. 2008. Learning and games. *The ecology of games: Connecting youth, games, and learning* 3 (2008), 21–40.
24. Arthur M Glenberg, Tiana Gutierrez, Joel R Levin, Sandra Japuntich, and Michael P Kaschak. 2004. Activity and Imagined Activity Can Enhance Young Children's Reading Comprehension. *Journal of Educational Psychology* 96, 3 (2004), 424.
25. Ward H. Goodenough. 1957. Oceania and the Problem of Controls in the Study of Cultural and Human Evolution. *The Journal of the Polynesian Society* 66, 2 (1957), pp. 146–155.
26. James W Heisig. 2007. *Remembering the Kanji, Volume 1: A Complete Course on How not to Forget the Meaning and Writing of Japanese Characters*. Vol. 1. University of Hawaii Press.
27. Dorothy Holland and Naomi Quinn. 1987. *Cultural models in language and thought*. Cambridge University Press.
28. Robert Howland, Sachi Urano, and Junichi Hoshino. 2012. SanjigenJiten: computer assisted language learning system within a 3d game environment. In *Advances in Computer Entertainment*. Springer, 262–273.
29. Jean Lave and Etienne Wenger. 1991. *Situated learning: Legitimate peripheral participation*. Cambridge university press.
30. Derek Lomas, Kishan Patel, Jodi L. Forlizzi, and Kenneth R. Koedinger. Optimizing challenge in an educational game using large-scale design experiments. In *CHI 2013*.
31. Jane McGonigal. 2003. A real little game: The performance of belief in pervasive play. *Proceedings of DiGRA 2003* (2003).
32. Gisli Palsson. 1994. Enskilment at sea. *Man* (1994), 901–927.
33. Mark Peterson. 2010. Massively multiplayer online role-playing games as arenas for second language learning. *Computer Assisted Language Learning* 23, 5 (2010), 429–439.
34. Paul Pimsleur. 1967. A memory schedule. *Modern Language Journal* (1967), 73–75.

35. Yolanda A Rankin, McKenzie McNeal, Marcus W Shute, and Bruce Gooch. 2008. User centered game design: evaluating massive multiplayer online role playing games for second language acquisition. In *Proceedings of the 2008 ACM SIGGRAPH symposium on Video games*. ACM, 43–49.
36. Joan Rubin. 1975. What the “good language learner” can teach us. *TESOL quarterly* (1975), 41–51.
37. Guoquan Sha. 2009. AI-based chatterbots and spoken English teaching: a critical analysis. *Computer Assisted Language Learning* 22, 3 (2009), 269–281.
38. T Townsend, P Clarke, and M Ainscow. 1999. Third Millennium Schools: prospects and problems for school effectiveness and school improvement. *T. Townsend, T. Clarke & M. Ainscow, Third Millennium Schools: A world of difference in effectiveness and improvement. Lisse: Swets & Zeitlinger* (1999).
39. Kathleen Tuite, Timothy Pavlik, Sandra B Fan, Tyler Robison, Alexander Jaffe, Yun-En Liu, Erik Andersen, and Steven Tanimoto. 2012. Picard: A creative and social online flashcard learning game. In *Proceedings of the International Conference on the Foundations of Digital Games*. ACM, 231–234.
40. Luis von Ahn. 2013. Duolingo: learn a language for free while helping to translate the web. In *IUI’13*. ACM, 1–2.