

# A Serious Game for Improving Inferencing in the Presence of Foreign Language Unknown Words

Pedro Gabriel Fonteles Furtado\*, Tsukasa Hirashima and Hayashi Yusuke  
Graduate School of Engineering  
Hiroshima University  
Hiroshima Prefecture, Japan

**Abstract**—This study presents the design of a serious game for improving inferencing for foreign language students. The design of the game is grounded in research on reading theory, motivation and game design. The game contains trial-and-error activities in which students create conversations and then watch these conversations play out. Making mistakes results in students receiving feedback and being requested to try again. An evaluation of the system was also conducted, in which participants used both simple text and the game. Post-test scores for using the game were significantly higher than scores when reading the text. User reception to the system was also positive. These results suggest that serious games can be effective for enhancing inferencing when foreign language students face unknown words. Implications for reading comprehension and for incidental vocabulary learning are also discussed.

**Keywords**—*Serious game; foreign language; contextual inference; unknown words*

## I. INTRODUCTION

Inferencing is the process of making connections when trying to interpret a text [1]. Inferencing information from the text is a necessary component of reading. To understand a text, readers must use their previous knowledge as base for the inferences of the new information. It also plays a role when disambiguating the meaning of words and clauses [2]. The inferencing necessary is not extensive, unless the information is too unfamiliar, or the language proficiency of the reader is not sufficient for that specific text [3]. One case of lack of proficiency is lack of vocabulary. Unknown words will often increase the amount of inferencing necessary for understanding the passage. In such cases, the reader will have to rely on the words he does know, the current information he collected from the text and his own background knowledge. In some cases, the reader may also infer the meaning of the unknown word. Children learn thousands of words per year and those words are mostly acquired from context [4], [5]. This also happens for second language (L2) learners [6]–[9].

As such, inference from context in the presence of unknown words plays two roles in reading comprehension:

- Understanding the passage by using the remaining information;
- Inferring the meaning of the unknown word to aid in understanding the passage.

Learning words from context is a form of incidental learning. Incidental learning is the accidental learning of information without intention of remembering that information [10],

[11]. The amount of words the students learn from incidental learning might make one believe that inferring the meaning of a word from context is an easy task. The problem is that students often fail to pick up this contextual information in the presence of unfamiliar vocabulary. Learners may ignore the word and give up on understanding the given passage. There may not be enough information in the context to infer the meaning of the word. It can also be the case that students infer the wrong meaning of a word [10], [12], [13]. Even when using dictionaries, students don't look up the meaning of all the words [14], specially when there are too many new items [15]. Also, new words usually need to be encountered multiple times [4] in order to be learned.

One factor that has been shown to affect incidental learning is task involvement load. Vocabulary enhancing techniques have been used to increase the effectiveness of incidental vocabulary learning [12], [16]–[19]. Past studies also agree that the higher the involvement load, the higher the effectiveness of incidental vocabulary learning. A meta-review on this matter can be seen in [20]. These tasks, however, have limitations. First, time spent in the classroom is limited. This limits the amount of time users will spend interacting with these activities. Users cannot be expected to engage for long periods of time in these tasks on their free time. Yet, as mentioned before, students are able to learn a large amount of words in their school years which cannot be attributed to explicit learning. Voluntary reading, often done outside the classroom, have been associated with better language acquisition [21] and better vocabulary test scores [22]. Voluntary reading works because of the large volume of reading done. Students read in such large volumes that they have multiple encounters with the unknown words. However, the amount of reading necessary and the time it takes to for measurable progress to be made can make extensive reading hard to implement [23].

Current research agrees that reading comprehension is positively affected by motivation [24]–[26]. Students with poor reading skills show more correlation between their motivation and their reading performance [24]. Motivation also affects the total amount of reading done [27]. Interventions to increase reading comprehension have been shown to increase reading performance [28]. Motivation is also a predictor of success in language learning in general. It was shown that motivation had the stronger correlation with language grades and self-evaluation [29]. In that study, motivation outperformed both the attitude towards the learning situation, integrativeness and orientations.

One approach to handling motivation is Digital Game

Based Learning (DBGL). One definition for DBGL is “the innovative learning approach derived from the use of computer games that possess educational value or different kinds of applications that use games for learning and education purposes such as learning support, teaching enhancement, assessment and evaluation of learners” [30]. DBGL includes the use of both commercial games and serious games. Serious games are digital games made for more than entertainment [31]. One core element of games that affects motivation is challenge. Appropriate challenge that matches the skill of the user will greatly affect the experience [32]. If it’s too easy, the player will be bored. If it’s too hard, the player will be discouraged. This fits with the conditions to achieve flow state, a popular construct in entertainment research [33]. It also fits with the need for competence from the Self Determination Theory explained in [34], [35]. As such, proper challenge is one of the factors that influences engagement and motivation in game design. However, game design is not about arbitrarily creating challenge. A game must be both accessible and easy to use while still providing a hard experience for the player [36]. This means that a game’s challenge should not be born from usability issues. It’s necessary to focus on usability in game design. Also cited as an important element is for the player to have freedom to fail and try again, as much as the user needs [32].

On the limitation of using commercial games, [37] conducted a research comparing vocabulary recall in players and watchers of a music game. Players actually interacted with the game and watchers were asked to simply watch the game. Watchers had a much higher score in vocabulary recall. The research reported that players were divided between listening to the words or doing well in the game. This shows that extraneous cognitive load can get in the way of reading, depending on game genre. This shows that while using commercial games is cost-effective, there might be a loss in learning gains compared to a well-designed educational game.

One popular game genre is visual novels. It involves reading a narrative through long periods of time [38]. Visual novels, unlike books, only show one snippet of text at a time. After the player does some sort of interaction with the game (pressing a button, for instance) the game advances to the next snippet of text. This means that players are presented with a limited amount of text at a time, meaning that players don’t have to keep track of their progress in a book’s page, for example. Visual novels also have graphical elements like backgrounds and character artwork. This gives the player a vision of what is going on inside the story. This facilitates reading comprehension [39]. Many visual novels also have some game-play elements between story-line sections, such as [40]. Alternation between story and game-play is a recurring element in game design, and is said to have beneficial elements, such as rewarding the player and improving pacing [41].

The serious game used in this study contains an activity designed to induce students to infer information from context. It locks them into a trial-and-error feedback loop while they attempt to construct a conversation. It combines this activity with a story, similar to a visual novel. In the sections below, the design of the game will be further explained, with a focus put on how it improves inferencing and on its motivational

elements. Then, the experiment will be described, and the results analyzed. The game used in this study has been explored before in the context of extensive reading [42], [43].

This study aims to answer whether or not inferencing improved by using the game when compared to simply reading text. It also aims to present how the design of the game interacts with inferencing.

Section II presents related work in the field, exploring other works that used games for language teaching with a focus on reading. Section III presents the design of the game and details of the experiment. The game’s design portion starts by presenting the challenges of designing a game for improving inferencing, based on the discussion presented in Section I. It then presents the various aspects of the game and ends by summarizing how the aspects of the game answer the design challenges. Then it introduces various aspects of the experiment done. Section IV presents the results obtained through the experiment and discusses their meaning. Section V concludes the study. It presents implications of the results on the field, the shortcomings of the study and possible future research.

## II. RELATED WORK

DBGL has been used successfully for language teaching on various fields, ranging from situated vocabulary learning [44], conversational visual novels [45], commercial games in the classroom [46], [47], relating language gains to gaming habits [48] and so on. The synthesis done in [49] about video games and second language learning concluded that games have a positive impact on learning, specially for vocabulary, with the experimental group surpassing the traditional study control group in some cases [50]–[53]. This shows that it is possible to have gains when reading content in games. The gains from vocabulary measured are due to incidental learning while playing. This shows that inferencing while reading also happens when playing games.

Despite this, there has not been much research that focused specifically on designing a software focused on supporting reading as the main activity in the context of DBGL and foreign languages. The work of [54] attempted to use a augmented reality game to enhance reading comprehension but it failed to show gains in reading comprehension. It did, however, show motivational gains. Some works [55]–[57] focus on first language primary reading skills (among other fields) for young children and showed positive results, but does not focus on reading long texts, focusing instead on more basic skills, such as individual word reading. The study of [58] focuses on first language reading comprehension, but does not evaluate the actual learning gains and does not expose how the design of the game relates to actually attaining those skills. Other works, like [59], address L2 reading but do not go in depth in designing the application to integrate with the reading process.

As far as inferencing during foreign language reading goes, none of the work reviewed addressed the process directly.

## III. METHODOLOGY

### A. Design of the Game

1) *Design Challenges*: Summarizing the material presented in Section I, the challenges in designing an activity for



Fig. 1. Screen-shot of a story sequence in the game.

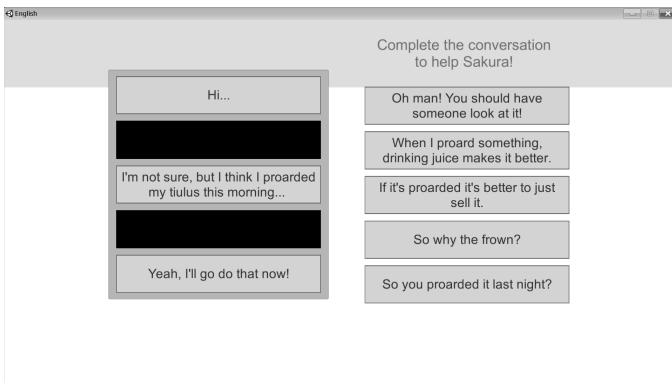


Fig. 2. Screen-shot of the conversation assembling activity.

improving inference from text in the presence of unknown words are:

- Students may ignore passages which contain unknown words;
- Students may infer the wrong meaning of a word;
- Activities would benefit from being intrinsically motivating, which implies:
  - Better performance in reading comprehension and language learning in general;
  - Possibility for the activity to be used in students' free time, thus avoiding time limitations of the classroom;
  - Compatibility with extensive reading.

If the activity is a game, it would also present the following challenges:

- Game elements should not detract from reading or from inferencing;
- The game needs be challenging but not too challenging;
- The activity should allow for students to fail as much as they need in order to progress.

2) *Game Introduction:* The game features a combination of story segments and activity segments. Screen-shots of the segments can be seen in Fig. 1 and 2, respectively.

In the story segment, we can see the characters present in the scene, a box displaying a piece of dialog and a background depicting the current location. Those elements can all be seen in Fig. 1. Upon user input, the story advances. This makes the next line of dialog or narration appear. This dynamic continues until the scene ends. While reading is prevalent, no elements are present to induce or improve contextual inferencing in this segment.

The inducement of contextual inference happens in the activity segments. During activity segments, users attempt to construct a conversation that solves a certain in-story goal. The conversation is constructed by inserting the pieces of the conversation into the empty slots, as shown in Fig. 2. The design of activity segments will be further discussed below.

3) *The Conversation Construction Activity's Design:* This activity consists of constructing a conversation and watching it play out. If the constructed conversation is inappropriate, a new conversation will be formed that will give the user insight into why that conversation is wrong and into how to create the appropriate conversation. From now on we'll refer to the phase of constructing a conversation as the assembling phase and the phase of watching the conversation play out as the result phase. Those two phases will be further developed in the subsections below.

The ideal behavior of the user for this activity can be seen in Fig. 3.

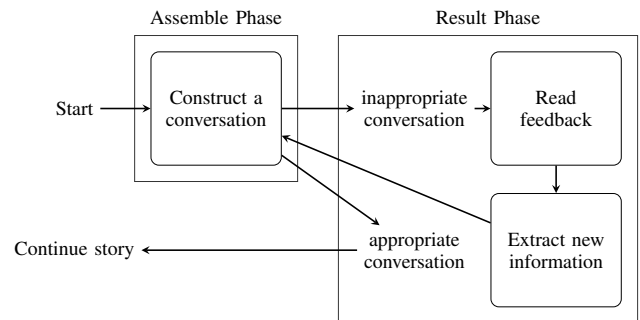


Fig. 3. Ideal user behavior flow for the conversation construction activity.

4) *Conversation Construction's Assembling Phase:* This phase consists of forming a sequential dialog by inserting conversation pieces into a grid, like in Fig. 2. However, the user can only insert the pieces related to what one person says. What the other person says is already fixed on the grid and cannot be moved. This was a deliberate decision to reduce ludo-narrative dissonance. Ludo-narrative dissonance is when game-play and story have a mismatch [60]. If players ask themselves "if I am the main character in the narrative, how come I can control what the other person will say?" immersion would break. Whenever a student fills up all vacant spaces with conversation pieces a button will appear in the interface. Pressing that button will take the student back to the screen of the story segments and the resulting conversation will play out.

Regarding Fig. 3, this refers to the "Construct a conversation" node.

5) *Conversation Construction's Result Phase:* First, the system must check if the conversation is appropriated or

not, by comparing it to the answer. If the conversation is appropriate, it will be shown to the player as it is, and the story will go on. This refers to the “appropriate conversation case” in Fig. 3.

However, if it’s incorrect, the system must logically assemble a new conversation based on the player’s constructed conversation. this is done by using the following steps:

- Find the player’s first mistaken conversation piece in the conversation by comparing the correct conversation with the assembled conversation from top to bottom;
- Discard all conversation pieces below the player’s first mistaken conversation piece;
- Insert the text that has been previously prepared as a reaction to the mistaken conversation piece. This text will show up after the mistaken conversation piece;
- Insert the text that has been previously prepared as a clue for the correct conversation piece that would fit in the position the player made his first mistake. This text will appear after the text of the previous step.

In Fig. 3, this would be the inappropriate conversation case. This new generated conversation is then shown to the player. After the generated conversation ends, the player will go back to the conversation construction screen. This process can be better understood in Fig. 4.

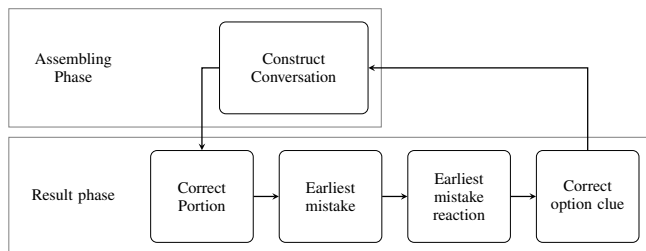


Fig. 4. Flow chart for mistakes during the conversation construction activity.

In the “First mistake reaction”, when the conversation goes into an unexpected flow, the user feedback begins, where users can acquire information on why the card related to the “first mistake” is unappropriated and insight in to what conversation piece would be appropriate for that particular slot. Users who are reading attentively will also be able to clearly point out which conversation piece has been considered inappropriate, since the feedback (the change in the conversation flow) begins at that moment. This feedback is effective because it uses the player’s inappropriate input to generate a conversation, similarly to an error-based simulation (such as the one used in the work of [61]). Instead of simply stating “this conversation is wrong, the correct one is this one”, it allows players to reflect on their input in a more effective way. This refers to the “Read feedback” and “Extract new information” nodes in Fig. 3.

6) “Look for Clues” Functionality: In Section III-A5, it has been stated that feedback starts at the “First mistake reaction”. To further enhance this fact to the player and to support him in relating this feedback to the “First mistake”, a

screen was added, called the “Look for Clues” message, which will appear right before the “First mistake”. The message says the following “This conversation will not go as expected! Read it to find clues!!!”. This happens in the middle of the conversation and it tells the user that:

- The created conversation has a problem.
- Until that point the conversation did not have a problem.
- There is something wrong with the conversation piece that appears right before that message.
- Looking for clues in whatever is coming up next is what the game expects them to do.

This was designed to further induce the ideal behavior in Fig. 3.

7) *Game Design Elements*: This session will describe the game design elements that have been incorporated into the design of the game. Their effects and importance will also be discussed. References for these elements can be seen in Section I.

Challenge and Freedom to Fail: Our approach for challenge has been through natural, emergent difficulty. As we’ve previously shown, reading comprehension for L2 learners can be a fairly difficult task. On extensive reading there is a focus on choosing texts with appropriate difficulty to mitigate this difficulty. The conversation construction task involves extracting information from the text and using that information. As such, it should have a difficulty similar to the reading comprehension process. The difference is that feedback is provided. In our feedback loop, progress will make it simpler for him to solve the activity. This way, every time the user tries to solve the task, he should have more information and the task should become easier.

About freedom to fail, the user is free to fail in our design. Furthermore, he is rewarded with feedback from his failure.

Visual Novel: The game is very similar to a visual novel and could be classified as such. This was not an arbitrary design decision. As discussed before, visual novels have a number of elements that make them an effective reading application.

User Interaction: In our conversation construction activities, drag-and-drop is the main form of interaction used. In [62], drag-and-drop is encouraged and described as an intuitive way to move content through the system. Our conversation construction activity has been designed with this in mind for its intuitiveness and for providing a fast way to construct the conversation. This approach has been used in applications like Monsakun to achieve similar effects and they have been well received [63].

8) *Addressing the Design Challenges*: Students may ignore passages which contain unknown words: A user that displays this behavior is not performing according to the ideal behavior displayed in 3. If the user ignores a passage, he would have trouble building the conversation. Because of this, the chances of the user making a mistake would rise. Upon making a mistake, the user would then be presented with feedback. At

that moment, the “look for clues” functionality described above further points the user to reading the feedback.

Furthermore, the chances of the user solving the activity by luck is 5%, given the default setup of five conversation pieces and two empty slots. It is low enough to make reading the feedback a more suitable strategy than trying to make the correct conversation by luck.

Students may infer the wrong meaning of a word: This would also imply in students making a mistake in the conversation construction activity. The expectancy is that the user will be able to correct his misunderstanding from reading the feedback.

Activities would benefit from being intrinsically motivating: The designing focuses on intrinsic motivation by balancing challenge, offering freedom to fail and by using drag-and-drop for ease of use. As discussed before, these are the elements related to the intrinsic motivation in games.

Game elements should not detract from reading or from inferencing: As seen before, the game has two types of segments, story segments and conversation construction segments. Both segments include reading. There are no actions to be done in-game that don't involve reading in some way. Conversation construction segments would result in multiple readings of the conversation pieces. Students are also expected to be carefully reading the feedback. As such, instead of detracting, the design has a focus on improving inferencing.

The game needs be challenging but not too challenging: One challenge of L2 reading is matching the difficulty of the text to the skills of the user. As such, difficulties in reading are highly content based. The trial-and-error with feedback design mitigates this issue by making the activity progressively easier as the user keeps on reading the feedback. This has been further explored in the subsection Game Design Elements.

The activity should allow for students to fail as much as they need in order to progress: This is a natural part of the trial-and-error design. Further details above in the subsection Game Design Elements.

### B. Game Development

The game has been developed using the C# language and the game engine Unity [64]. A scripting language was created to describe the story sequences and the conversation construction activities. Since the story sequences are similar to visual novels, the commands used are similar in format to the ones found in Ren'py, a visual novel engine [65]. A script interpreter was written to render the scene for the players, while also handling input.

### C. Design of the Experiment

This study used a within-subject design with two conditions for counterbalancing: text-game and game-text. Afterwards, participants took a post-test and some of the users took a user perception survey. Text-game read a text and then played the game. Game-text played the game and then read the text. All measurements were done in the end so that measuring would not affect the behavior of the users. This flow can be seen in Fig. 5. The post-test had three sections:

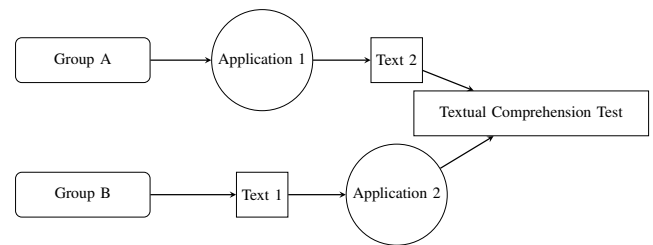


Fig. 5. Experiment flow diagram

remembering section, textual comprehension section and word comprehension questions. Both game and text included dummy words to create a situation where users are reading a material with unknown words.

13 Japanese University students participated in the study and were randomly assigned to each condition.

Two textual contents were used in this study, A and B. Both contents have a game form and a textual form. Thus, we have game A, game B, text A and text B. The text-game group used text A and game B. The game-text group used game A and text B. Content A and B were found to be appropriate or below the difficulty of Grade 2 in the Common Core State Standards [66]. As such, both contents are considered to be accessible and equivalent in difficulty. This was measured using the TextEvaluator tool [67]. Scores given by the tool were found to have high correlation with judgment presented by human experts [68]. Content A has 164 words while content B has 226 words.

The post-test was divided into three sections:

- Remembering section: users were asked to write as much as they could remember with as much detail as possible.
- Textual interpretation section: users were asked questions such as “Did Brian ever get angry in the story? If yes, why did he get angry?”
- Word comprehension section: users were asked to explain the meaning of the dummy words and to translate phrases that used the dummy words.

The textual interpretation questions were designed around the passages that contained dummy words. This means that inferencing information in the presence of unknown words is necessary to correctly answer the questions. Only the last two sections are used to calculate the scores. The first section was included for the possibility of an exploratory analysis, but it is not addressed in this study.

The user perception survey had four questions. Three of them comparing text and game on ease of content understanding, motivation to read and on suitability for studying English. The last question was about the usability of the game.

The experiment was performed in participants individually. They interacted with the game on a computer. The text was read through a PDF file. While interacting with the game, participants were taught that clicking would advance the story. They were also taught how to drag-and-drop to build the conversations. The post-test and the user perception survey were both administered through an online form.

TABLE I. AVERAGE SCORES AND STANDARD DEVIATION FOR THE TWO CONDITIONS AND FOR ALL PARTICIPANTS

	Application M (SD)	Text M (SD)	N
Game-text	0.76 (0.09)	0.58(0.24)	7
Text-game	0.77 (0.18)	0.44(0.24)	6
All participants	0.77 (0.13)	0.52(0.24)	13

TABLE II. SURVEY RESULTS. QUESTION ONE TO QUESTION THREE IS ABOUT COMPARING GAME AND TEXT. QUESTION FOUR IS ABOUT USABILITY

Which one is easier to understand?	
The game is better	28.6%
The game is a bit better	28.6%
They are the same	42.9%
The text is a bit better	0.00%
The text is better	0.00%
Which one makes you want to read it more?	
The game is better	57.1%
The game is a bit better	28.6%
They are the same	0.00%
The text is a bit better	0.00%
The text is better	14.3%
Which one do you think is better for studying English?	
The game is better	14.3%
The game is a bit better	42.9%
They are the same	42.9%
The text is a bit better	0.00%
The text is better	0.0%
Was the application easy to use?	
Easy to use	71.4%
A little bit easy to use	14.3%
Not easy, not hard	0.0%
A little bit hard to use	14.3%
Hard to use	0.0%

#### IV. RESULTS AND DISCUSSION

Table I shows scores obtained by the two conditions and for all participants. Scores have a minimum value of 0 and a maximum value of 1. Game scores had a lower standard deviation than text scores.

A two-way analysis of variance was conducted on the influence of two independent variables (medium, order of use) on the post test scores. Medium includes two levels (game, text) and order of use consisted of two levels (first, second). The only significant effect at the .05 significance level was for the medium factor. The main effect for medium yielded an F ratio of  $F(1, 22) = 11.16, p < .01$ , indicating a significant difference between using the game ( $M = 0.77, SD = 0.13$ ) and reading the text ( $M = 0.52, SD = 0.24$ ). The main effect for order yielded an F ratio of  $F(1, 22) = 1.06, p > .05$ , indicating that the effect for order was not significant, first ( $M = 0.61, SD = 0.24$ ) and second ( $M = 0.67, SD = 0.23$ ). The interaction effect was not significant,  $F(1, 22) = 0.86, p > .05$ .

The above results suggest that using the game results in more information absorbed than using the text. It also suggests that order of use (which one is used first, and which one is used second) does not affect the amount of information absorbed.

As for the user perception survey results, found in Table II, the following trends were found:

- In the area of interest, all users except for one had a positive opinion towards the game, with over half of the users completely favoring the game.
- On perceived comprehensibility and perceived learning, half of the users had a positive opinion while the

other half had a neutral opinion.

- On usability, one user found the game a little bit hard to use, while the clear majority thought the game was easy to use.
- The user who felt the game is a little bit hard to use is the only one user that was unfavorable towards the game in any of the areas. He also favored printed text in the area of interest.

Those trends show that the hypothesis was true. About the one user that was unfavorable towards the game, his scores were checked in order to see if his unfavourability affected his scores. Surprisingly, he was the only user to get a perfect grade related to the content in the game version he used, suggesting that the comprehensibility scores are not affected by dislike of the game. These results fit well with past findings suggesting good affective reception from learners in relation to DBGL, as reported in [69] and in other works ([55], [70], etc.).

Users higher comprehensibility when using the game can be attributed to being able to read the feedback information to solve the conversation construction problems. This suggests that users were performing according to the ideal behavior previously defined, indicating that our efforts to create an activity that can only be practically solved by displaying the needed behavior have been successful. When using the text, users may have been more likely to ignore passages or to make mistakes during inferencing. This gain in performance is reflected not only in reading comprehension but also in incidental vocabulary learning, since the experiment included dummy words. Thus, results suggest that users are able to infer partial meaning of the words better when using the game.

#### V. CONCLUSION

Results suggest that users are able to infer information from context better by using the game. This implies that activity designs based on creating a trial-and-error task with automatic feedback can be useful for improving reading comprehension and improving incidental vocabulary learning. Qualitative results have also been positive.

As for the perception of the game as an English studying tool in comparison to the paper version, around half of the users pointed to them being equally effective. And yet the comprehension scores for the game version have been much higher. This contradiction between user's perceived learning effectiveness and the actual effectiveness has also been reported by [59]. Low perceived learning is also one of the challenges of extensive reading, so making DBGL tools have a higher perceived learning by students should positively impact their performance and studies in that direction are necessary, such as measuring differences in flow and motivation between incidental and explicit learning.

Remaining issues would be performing additional experiments to show more compelling evidence of the increases in reading comprehension, since the number of participants was small. Also, the low perceived learning and the fact that the design relies on the presence of conversations are also limitations. Expansions to this research could focus on making learning more explicit by mixing the narratives with explicit vocabulary teaching, thus making the learning process more

obvious to the student. Another problem is that, currently, producing content for the game is a complex task. Creating a tool to assist this process would allow content to be created by teachers and other content creators.

Another possible next step is adapting the design to be generated based on natural language processing techniques without human input. This would allow for a large amount of game content to be created. This would have implications for improving the performance of extensive reading programs.

#### REFERENCES

- [1] G. Brown and G. Yule, *Discourse analysis*. Cambridge university press, 1983.
- [2] W. Grabe, *Reading in a second language: Moving from theory to practice*. Ernst Klett Sprachen, 2009.
- [3] W. P. Grabe and F. L. Stoller, *Teaching and researching: Reading*. Routledge, 2013.
- [4] W. E. Nagy and P. A. Herman, "Incidental vs. instructional approaches to increasing reading vocabulary." *Educational perspectives*, vol. 23, no. 1, pp. 16–21, 1985.
- [5] W. E. Nagy and R. C. Anderson, "How many words are there in printed school english?" *Reading research quarterly*, pp. 304–330, 1984.
- [6] W. Grabe and F. Stoller, "Reading and vocabulary development in a second language: A case study," *Second language vocabulary acquisition: A rationale for pedagogy*, pp. 98–122, 1997.
- [7] M. Horst, "Learning 12 vocabulary through extensive reading: A measurement study," *Canadian Modern Language Review*, vol. 61, no. 3, pp. 355–382, 2005.
- [8] M. Pigada and N. Schmitt, "Vocabulary acquisition from extensive reading: A case study," *Reading in a foreign language*, vol. 18, no. 1, p. 1, 2006.
- [9] B. Dupuy and S. D. Krashen, "Incidental vocabulary acquisition in french as a foreign language." *Applied Language Learning*, vol. 4, pp. 55–63, 1993.
- [10] J. H. Hulstijn, "Retention of inferred and given word meanings: Experiments in incidental vocabulary learning," in *Vocabulary and applied linguistics*. Springer, 1992, pp. 113–125.
- [11] R. Schmidt, "Deconstructing consciousness in search of useful definitions for applied linguistics," *Consciousness in second language learning*, vol. 11, pp. 237–326, 1994.
- [12] J. H. Hulstijn, M. Hollander, and T. Greidanus, "Incidental vocabulary learning by advanced foreign language students: The influence of marginal glosses, dictionary use, and reoccurrence of unknown words," *The modern language journal*, vol. 80, no. 3, pp. 327–339, 1996.
- [13] B. Laufer and D. D. Sim, "Measuring and explaining the reading threshold needed for english for academic purposes texts," *Foreign language annals*, vol. 18, no. 5, pp. 405–411, 1985.
- [14] J. H. Hulstijn, "When do foreign-language readers look up the meaning of unfamiliar words? the influence of task and learner variables," *The Modern Language Journal*, vol. 77, no. 2, pp. 139–147, 1993.
- [15] F. R. Jones, "Learning an alien lexicon: A teach-yourself case study," *Second Language Research*, vol. 11, no. 2, pp. 95–111, 1995.
- [16] M. Hill and B. Laufer, "Type of task, time-on-task and electronic dictionaries in incidental vocabulary acquisition," *International Review of Applied Linguistics*, vol. 41, no. 2, pp. 87–106, 2003.
- [17] M. Horst, T. Cobb, T. Cobb, and P. Meara, "Beyond a clockwork orange: Acquiring second language vocabulary through reading." *Reading in a foreign language*, vol. 11, no. 2, pp. 207–223, 1998.
- [18] S. Knight, "Dictionary use while reading: The effects on comprehension and vocabulary acquisition for students of different verbal abilities," *The Modern Language Journal*, vol. 78, no. 3, pp. 285–299, 1994.
- [19] Y. Kim and K. McDonough, "The effect of interlocutor proficiency on the collaborative dialogue between korean as a second language learners," *Language Teaching Research*, vol. 12, no. 2, pp. 211–234, 2008.
- [20] S. Huang, V. Willson, and Z. Eslami, "The effects of task involvement load on 12 incidental vocabulary learning: A meta-analytic study," *The Modern Language Journal*, vol. 96, no. 4, pp. 544–557, 2012.
- [21] S. Krashen, "Free voluntary reading: New research, applications, and controversies," *Anthology series-Seameo regional language centre*, vol. 46, no. 1, 2005.
- [22] R. C. Anderson, P. T. Wilson, and L. G. Fielding, "Growth in reading and how children spend their time outside of school," *Reading research quarterly*, pp. 285–303, 1988.
- [23] C. Davis, "Extensive reading: an expensive extravagance?" *ELT journal*, vol. 49, no. 4, pp. 329–336, 1995.
- [24] S. Logan, E. Medford, and N. Hughes, "The importance of intrinsic motivation for high and low ability readers' reading comprehension performance," *Learning and Individual Differences*, vol. 21, no. 1, pp. 124–128, 2011.
- [25] L. Baker and A. Wigfield, "Dimensions of children's motivation for reading and their relations to reading activity and reading achievement," *Reading research quarterly*, vol. 34, no. 4, pp. 452–477, 1999.
- [26] P. L. Morgan and D. Fuchs, "Is there a bidirectional relationship between children's reading skills and reading motivation?" *Exceptional children*, vol. 73, no. 2, pp. 165–183, 2007.
- [27] A. Wigfield and J. T. Guthrie, "Relations of children's motivation for reading to the amount and breadth of their reading." *Journal of educational psychology*, vol. 89, no. 3, p. 420, 1997.
- [28] J. T. Guthrie, A. Wigfield, P. Barbosa, K. C. Perencevich, A. Taboada, M. H. Davis, N. T. Scaffidi, and S. Tonks, "Increasing reading comprehension and engagement through concept-oriented reading instruction." *Journal of educational psychology*, vol. 96, no. 3, p. 403, 2004.
- [29] A.-M. Masgoret and R. C. Gardner, "Attitudes, motivation, and second language learning: a meta-analysis of studies conducted by gardner and associates," *Language learning*, vol. 53, no. 1, pp. 123–163, 2003.
- [30] S. Tang, M. Hanneghan, and A. El Rhalibi, "Introduction to games-based learning," *Games Based Learning Advancements for Multi-Sensory Human Computer Interfaces*. New York: IGI Global, 2009.
- [31] T. Susi, M. Johannesson, and P. Backlund, "Serious games: An overview," 2007.
- [32] S. Deterding, "The lens of intrinsic skill atoms: A method for gameful design," *Human-Computer Interaction*, vol. 30, no. 3-4, pp. 294–335, 2015.
- [33] N. D. Bowman, "A pat on the back: Media flow theory revis (it) ed." *Rocky Mountain Communication Review*, vol. 4, no. 1, 2008.
- [34] E. L. Deci and R. M. Ryan, "Motivation, personality, and development within embedded social contexts: An overview of self-determination theory," *The Oxford handbook of human motivation*, pp. 85–107, 2012.
- [35] A. K. Przybylski, C. S. Rigby, and R. M. Ryan, "A motivational model of video game engagement." *Review of general psychology*, vol. 14, no. 2, p. 154, 2010.
- [36] J. Juul and M. Norton, "Easy to use and incredibly difficult: on the mythical border between interface and gameplay," in *Proceedings of the 4th international conference on foundations of digital Games*. ACM, 2009, pp. 107–112.
- [37] K. Kuwada, "The effect of interactivity with a music video game on second language vocabulary recall," *About Language Learning & Technology*, vol. 74, 2010.
- [38] D. Cavallaro, *Anime and the visual novel: narrative structure, design and play at the crossroads of animation and computer games*. McFarland, 2009.
- [39] J. Liu, "Effects of comic strips on 12 learners' reading comprehension," *TESOL quarterly*, vol. 38, no. 2, pp. 225–243, 2004.
- [40] S. Chunsoft, "Danganronpa: Trigger happy havoc," UMD, 2010.
- [41] H. Hancock, "Better game design through cutscenes," *Gamasutra.[Online]*. Available: <http://www.gamasutra.com/features/20020401/hancock>, vol. 1, 2002.
- [42] P. Furtado, T. Hirashima, and Y. Hayashi, "Transforming foreign language narratives into interactive reading applications designed for comprehensibility and interest," in *International Conference on Artificial Intelligence in Education*. Springer, 2017, pp. 510–513.
- [43] P. G. F. Furtado, T. Hirashima, and Y. Hayashi, "Development and experimental evaluation of an interactive reading application designed for comprehensibility and interest," *Workshop Proc. of ICCE2017*, vol. 25, 2017.

- [44] M. E. C. Santos, T. Taketomi, G. Yamamoto, M. M. T. Rodrigo, C. Sandor, H. Kato *et al.*, "Augmented reality as multimedia: the case for situated vocabulary learning," *Research and Practice in Technology Enhanced Learning*, vol. 11, no. 1, p. 1, 2016.
- [45] I. D. Agusalim, "Developing visual novel game of english conversation for dep eepis," *Journal of Education and Practice*, vol. 6, no. 33, pp. 113–124, 2015.
- [46] H.-J. H. Chen and T.-Y. C. Yang, "The impact of adventure video games on foreign language learning and the perceptions of learners," *Interactive Learning Environments*, vol. 21, no. 2, pp. 129–141, 2013.
- [47] Y. A. Rankin and M. W. Shute, "Re-purposing a recreational video game as a serious game for second language acquisition," *Serious game design and development: Technologies for training and learning*, pp. 178–195, 2010.
- [48] P. Sundqvist and L. K. Sylvén, "World of voccraft: Computer games and swedish learners 12 english vocabulary," in *Digital games in language learning and teaching*. Springer, 2012, pp. 189–208.
- [49] A. Yuditseva, "Synthesis of research on video games for the four second language skills and vocabulary practice," *Open Journal of Social Sciences*, vol. 3, no. 11, p. 81, 2015.
- [50] S. Vahdat and A. R. Behbahani, "The effect of video games on iranian efl learners vocabulary learning," *Reading*, vol. 13, no. 1, 2013.
- [51] C. I. Hitosugi, M. Schmidt, and K. Hayashi, "Digital game-based learning (dgb) in the 12 classroom: The impact of the un's off-the-shelf videogame, food force, on learner affect and vocabulary retention," *CALICO Journal*, vol. 31, no. 1, p. 19, 2014.
- [52] L. Aghlara and N. H. Tamjid, "The effect of digital games on iranian children's vocabulary retention in foreign language acquisition," *Procedia-Social and Behavioral Sciences*, vol. 29, pp. 552–560, 2011.
- [53] S. Suh, S. W. Kim, and N. J. Kim, "Effectiveness of mmorpg-based instruction in elementary english education in korea," *Journal of Computer Assisted Learning*, vol. 26, no. 5, pp. 370–378, 2010.
- [54] H. Tobar-Muñoz, S. Baldiris, and R. Fabregat, "Augmented reality game-based learning: Enriching students experience during reading comprehension activities," *Journal of Educational Computing Research*, vol. 55, no. 7, pp. 901–936, 2017.
- [55] R. Rosas, M. Nussbaum, P. Cumsille, V. Marianov, M. Correa, P. Flores, V. Grau, F. Lagos, X. López, V. López *et al.*, "Beyond nintendo: design and assessment of educational video games for first and second grade students," *Computers & Education*, vol. 40, no. 1, pp. 71–94, 2003.
- [56] M. Ven, L. Leeuw, M. Weerdenburg, and E. Steenbeek-Planting, "Early reading intervention by means of a multicomponent reading game," *Journal of Computer Assisted Learning*, vol. 33, no. 4, pp. 320–333, 2017.
- [57] D. Hooshyar, M. Yousefi, and H. Lim, "A procedural content generation-based framework for educational games: Toward a tailored data-driven game for developing early english reading skills," *Journal of Educational Computing Research*, p. 0735633117706909, 2017.
- [58] L. S. Gaytán-Lugo and S. C. Hernandez-Gallardo, "Towards improving reading comprehension skills in third graders with a serious game," *ICCE 2012*, p. 25, 2012.
- [59] B. E. Shelton, D. Neville, and B. McInnis, "Cybertext redux: using interactive fiction to teach german vocabulary, reading, and culture," in *Proceedings of the 8th international conference on International conference for the learning sciences-Volume 3*. International Society of the Learning Sciences, 2008, pp. 128–129.
- [60] C. Hocking, "Ludonarrative dissonance in bioshock: The problem of what the game is about," in *Well Played 1.0*. ETC Press, 2009, pp. 114–117.
- [61] T. Horiguchi, I. Imai, T. Toumoto, and T. Hirashima, "Error-based simulation for error-awareness in learning mechanics: An evaluation," *Educational Technology & Society*, vol. 17, no. 3, pp. 1–13, 2014.
- [62] I. Apple. (2017) Drag and drop - user interaction - human interface guidelines for macos apps. Apple Inc. [Online]. Available: <https://developer.apple.com/macos/human-interface-guidelines/user-interaction/drag-and-drop/>
- [63] T. Hirashima, T. Yokoyama, M. Okamoto, and A. Takeuchi, "Long-term use of learning environment for problem-posing in arithmetical word problems," in *Proceedings of ICCE*, vol. 2008, 2008, pp. 817–824.
- [64] U. G. Engine, "Unity game engine-official site," *Online*[Cited: February 17, 2018.] <http://unity3d.com>, 2018.
- [65] T. Rothamel, "The ren'py visual novel engine," <https://www.renpy.org/>, 2006, accessed: 2017-02-28.
- [66] C. C. S. S. Initiative *et al.*, *Common Core State Standards for English Language Arts & Literacy In History/Social Studies, Science, and Technical Subjects*. Washington, DC: Council of Chief State School Officers & National Governors Association., 2010.
- [67] K. M. Sheehan, I. Kostin, D. Napolitano, and M. Flor, "The textevaluator tool: Helping teachers and test developers select texts for use in instruction and assessment," *The Elementary School Journal*, vol. 115, no. 2, pp. 184–209, 2014.
- [68] K. M. Sheehan, "A review of evidence presented in support of three key claims in the validity argument for the textevaluator® text analysis tool," *ETS Research Report Series*, vol. 2016, no. 1, pp. 1–15, 2016.
- [69] T. Hainey, T. M. Connolly, E. A. Boyle, A. Wilson, and A. Razak, "A systematic literature review of games-based learning empirical evidence in primary education," *Computers & Education*, vol. 102, pp. 202–223, 2016.
- [70] E. Ojanen, M. Ronimus, T. Ahonen, T. Chansa-Kabali, P. February, J. Jere-Folotiya, K.-P. Kauppinen, R. Ketonen, D. Ngorosho, M. Pitkänen *et al.*, "Graphogame—a catalyst for multi-level promotion of literacy in diverse contexts," *Frontiers in psychology*, vol. 6, 2015.