

Breanna Jones

SPAN 672

Dr. Greg Thompson

March 24, 2022

Critique of Digital Game-Based Learning (DGBL) in the L2 Classroom: The Impact of the UN's Off-the-Shelf Videogame, Food Force, on Learner Affect and Vocabulary Retention

### **Complete bibliographic reference in APA**

Hitosugi, C., Schmidt, M., & Hayashi, K. (2014). Digital game-based learning (DGBL) in the L2 classroom: The impact of the UN's off-the-shelf videogame, Food Force, on learner affect and vocabulary retention. *CALICO Journal*, 31(1).

<https://journals.equinoxpub.com/CALICO/article/view/22800/18824>

### **Research Questions**

1. How can a videogame be used as sound teaching material for L2 classrooms, and how can we best implement the videogame in an existing curriculum? (Implementation)
2. How do students perceive and respond to the introduction of a videogame in the L2 classroom? (Affect)
3. How does the use of FF impact vocabulary learning and retention compared to textbook reading? (Vocabulary retention)

### **Participants**

The studies consisted of two separate classes of students totaling 20 students, with 9 students in the first study and 11 students in the second study. The average age of participants in the first study was 22.22 years old and the average age was 19.1 years old for the second study. In the first study, there were more male participants (7:2) and in the second study, there were

more female students than males (2:9). While the participants' proficiency levels varied from intermediate high to advanced low, all were pursuing a Japanese degree or certificate in a US university.

### **Research Design/Procedures**

The most important thing to understand about this study is the videogame used in both Study 1 and 2. *Food Force*, developed by the UN World Food Program (UNWFP) in 2005, was described as a “non-violent, educational, humanitarian videogame” (Hitosugi et. al., 2014 p. 23). It was later translated into a Japanese version by Konami and is available for free download with no regional restrictions. The game is set on a fictitious island called *Sheylan* and players take the role of a new member of the UNWFP. The six game “missions” teach players how to fight world hunger (Hitosugi et. al., 2014 p. 23). Only the first two missions were used in the two studies. The first mission, “Air surveillance: Locating hungry citizens in a helicopter,” is a traditional “find-em” game where player man a helicopter and locate as many hungry people as possible in a given amount of time (Hitosugi et. al., 2014 p. 23). The second mission, “Packing balanced food,” is a simulation to create a balanced diet with limited funds in a guessing-game format (Hitosugi et. al., 2014 p. 23).

Students were separated into two groups based on their class. In both groups (labeled as Studies 1 and 2 in the article), students participated in five class sessions with two “missions” and 8 activities. Each group had a pretest and posttest, along with a delayed test. In Study 1, the *FF* unit was used in tandem with the textbook unit on global environmental issues and no vocabulary list was given. There was also no end-of-unit test, nor was their performance in *FF* taken into account for their grade in the class. In Study 2, the *FF* unit was integrated into their food unit along with a story about the Japanese inventor of the cup of noodle. There was a

vocabulary list of 32 words given to Study 2 participants and students took both pop quizzes and an end-of-unit test. The gameplay was factored into their course grade. The comparison between the two groups and their game play helped to answer the third research question. The second research question was addressed in the use of questionnaires which were distributed after the unit for both groups and an interview was administered for group 2. My best guess at why an interview was given for Study 2 and not Study 1 is that the researchers found some trends in Study 2 that they wanted to follow up on, but this is not explicitly stated in the article. The first research question was addressed through their literature review as they determined the best tools to use in implementation. Things they took into consideration included availability of the game, cost of the game, and the use of an authentic game (one that used the target language).

## **Results**

The affect survey results for Study 1, which did not use videogames, had a mean of 6.56/10 with a standard deviation of .73 in regard to whether or not they would like to see more activities using videogames in Japanese class and in Study 2, where this type of activity was implemented, their score for engagement had a mean of 5.36/10 with a standard deviation of 1.12. There was a strong correlation ( $r = .91$ ) between the desire to learn more Japanese and the engagement in games. From these results, they determined that students enjoyed playing the videogames in class, would like more activities similar to it in class, that the language used was just outside of their zone of proximal development, but it did not negatively affect their enjoyment, and that they learned more about world hunger this way.

As for their vocabulary retention, students with just the textbook vocabulary test could remember words just as well as the students taking the videogame test, but that students remembered words from the videogame longer than those who were just given the traditional

instruction. This was determined by a follow-up five weeks later in which they remembered around 12 words, as opposed to the 11 they could remember right after the test (Hitosugi et. al., 2014 p. 32).

Results were analyzed using an ANOVA and a correlation was found (Wilks' Lambda = .20,  $F(2, 7) = 14.41$ ,  $p < .01$  for Study 1; Wilks' Lambda = .10,  $F(2, 9) = 39.42$ ,  $p < .001$  for Study 2).

## **Discussion**

Students who used the videogame to test their vocabulary had the same immediate retention of vocabulary, but their delayed testing showed that they retained the vocabulary longer than their textbook counterparts. The majority of students enjoyed the videogame assessment more than the traditional textbook method, although some did prefer the textbook and this could have been due to some anxiety associated with technology use. They also found that the videogame was more beneficial for students who had higher proficiency due to the more difficult vocabulary included in the videogame. There was no difference in performance between males and females, even though more males expressed excitement about the prospect of playing videogames than females.

## **Critique**

The study included collecting and analyzing data on two different groups. This data was then analyzed, and a correlation was found between the variables examined. While it was a small study (only about 20 people), it constitutes as a significant study because most research in the field of DGBL for Japanese has consisted of mainly ethnographic research conducted on as few as two individuals.

The research questions were clear, and I think that they are important questions to ask as we continue to examine what kinds of activities motivate our students and allow them to interact with the language in ways that enhance their learning. I do think that the first research question was not entirely answered by the study itself, but I do think it was important to include some background as to why they picked the program that they used.

The study is valid as most threats to internal and external validity were controlled for. They mentioned in the study that gender did not play a role in any difference between the two groups, but this could obviously be controlled for better in a future study. The only threat that could have influenced the data that they did not account for was subject bias. They reported that some students felt that they were not good with computers, and this could very well have influenced the students' performance during the activity. In future studies, it would be good to take a survey of participants' feelings towards videogames and computer-based gaming, as it could influence how well students perform during the activity itself.

The study was supposedly repeated twice, as they themselves did this experiment twice, but it is still unclear to me as to whether or not the students in Study 1 all participated in the textbook version or if they were randomly assigned to be in a videogame group or not. I think it could be repeated and I would recommend it be repeated with randomly assigned groups this time.

Game-based learning is important to the field of language acquisition as it increases motivation and the multi-modality of learning. This study was valuable then, especially to the Japanese language classroom and the second language classroom in general as it analyzes what happened in a classroom setting as opposed to a controlled lab environment. I would suggest that

it should be done on a larger scale in the future in order to make it generalizable. As it stands, the study is not widely generalizable and thus loses some of its value.

## Critique of Mobile-Assisted Language Learning: A Duolingo Case Study

### Complete bibliographic reference in APA

Loewen, S., Crowther, D., Isbell, D.R., Kim, K.M., Maloney, J., Miller, Z.F. & Rawal, H. (2019). Mobile-assisted language learning: A Duolingo case study. *ReCALL* 31(3): 293-311. <https://doi.org/10.1017/S0958344019000065>

### Research Questions

1. How effective is Duolingo in developing L2 knowledge in ab initio learners of Turkish?
2. What are the experiences of ab initio learners of Turkish using Duolingo to study an L2?

### Participants

The study's participant pool consisted of nine individuals from Michigan State University, three female and six male. Of these, eight were graduate students and one was a professor. Their backgrounds all varied in L1 and L2 learning experience. The native languages of participants were English, Chinese, Korean, and Nepali and all were at least bilingual while others were multilingual. The learners served not only as participants, but also as the researchers for the study, the study task itself being a class assignment. They acknowledge that this affects the generalizability of their data and the study's validity. The nine participants all participated in the first two phases of the study, while half then participated in phases three and four.

### Research Design/Procedures

The study specifically examined Duolingo, a popular language learning app that many people use when trying to learn a new language. Duolingo was chosen due to its widespread popularity and ease of access without a subscription fee. The app uses gamification strategies in that you gain experience points as you attain certain skills and are encouraged daily to keep up

your language learning streak. The points accumulated can be turned in for lingots which allow users to upgrade their learner's avatar and have a truly individualized experience.

Participants kept journals in which they reflected on their time spent in the app, the platform they used (mobile versus computer), levels completed and use of review, successes achieved/difficulties encountered, and connections they found between the app and ISLA theory. Each participant recorded weekly reflections. Students also took a Duolingo progress test which gives a score from 1-5 of their overall proficiency with no additional feedback. After this, participants took a Turkish 151 Test which would normally be given to students in Turkish 151 at the end of the semester.

The study was split into four phases alternating between Turkish study/journaling and language assessment. In the first phase, participants agreed to study Turkish on Duolingo for at least one hour a week for twelve weeks and were only allowed to use resources available on Duolingo, but were given complete autonomy in how they used said resources. At the end of this phase, they took the Duolingo progress test and the Turkish 151 exam. In phase three, half of participants then participated in additional Turkish study on Duolingo in order to reach 34 hours of time spent on Duolingo. These participants then took another progress test and the Turkish 151 exam again after they completed the 34-hour study target.

## **Results**

The mean time studied was 29 hours with a standard deviation of nine and a range of 23 (12-35). The experience points received had a mean of 2,897 with a standard deviation of 1,465 and a range of 3,617 (730-4347). The Duolingo progress test had a mean of 0.63 out of 5 with a standard deviation of 0.48 out of 5 and a range of 1.43 (0.35-1.78). The Turkish Test scores had a mean of 48 with a standard deviation of 19 and a range of 53 (23-76). Using the chart provided,



this would put the average student below the lowest level on the grading scale, meaning that they are not ready for the next level, with the highest scoring student at somewhat weak/somewhat poor meaning that they would struggle at the next level. While we are provided some way to interpret the Turkish exam scores, there was no way to interpret the Duolingo scores. There were moderately strong correlations between Duolingo experience points and the Turkish 151 Test scores and subscores. While the degree to which participants' participation varied, all but one participant failed to successfully achieve the 70% criterion for mastery in the university's Turkish 151 course.

In response to research question two regarding their experience using Duolingo, students reported enjoying the flexibility of use, though it sometimes led them to practice less frequently for longer periods of time. They also commented on their wavering motivation, which was speculated to be due to their lack of investment in the language or the culture itself. Students were fairly skeptical about their perceived progress in comparison with the progress that Duolingo stated they had made.

## **Discussion**

This study points out that there were other factors that could have influenced student motivation and progress, but that overall, even after 34 hours, only one student would have been allowed to move on past Turkish 151, which could speak against the claims that Duolingo makes. Additionally, the progress students were being told they had made did not reflect well on their perceived improvement or the improvement they saw in the Turkish exam.

## **Critique**

The study does involve collecting and analyzing data, but the data they collect is on themselves. Additionally, the research questions are clear and important, but due to their small sample size, I believe they begin to lose their importance.

This study had several threats to internal validity, which is unfortunate because I think something like this would be interesting to track a little more with a larger group. While the study mentions quite a few limitations, I'd like to point out that they allowed some students to retake the test, which could have led to an instrumentation threat due to having a pre and post test. Additionally, students did not have autonomy in their choice of language, which could have led them to lack motivation to finish the course, at least intrinsic motivation. All students were extrinsically motivated because their participation in the study was part of their grade, but this brings up some questions for me about the way these researchers went about getting participants. I think it would have been beneficial to have brand new students log time spent in Duolingo and compare the scores of students who did Duolingo each day with those who did not. Even with the help of the classroom practice, I think it would be good to see the results. If they really want to focus on Duolingo-only, they could have asked students at Michigan State University to enroll in Duolingo in a language course they had never taken and then record their results on challenge exams.

The final threat to internal validity is the Turkish 151 exam. We are not told exactly what this test covers or how it is a good measure of proficiency. Seeing as it specifically pertains to a course on campus, I would wager to guess that it focuses on things specifically discussed in that course and is not necessarily a great test of proficiency overall. I would love to understand more about the Duolingo progress test, especially because the students did not fair well on it. I wonder at what point that test should be taken and what evidence they have that 34 hours is sufficient.

Despite all this, the study is repeatable and I think it would be a good study to replicate. On a larger scale, this study would be value to the field of gamification and to Duolingo as well as language educators around the world.

Critique of Gamifying Portuguese Language Learning: A Case Study Examining a Quest-Based Website to Prompt Oral Production and Interaction in Learners of Portuguese L2

**Complete bibliographic reference in APA**

Xavier, C. (2020). Gamifying Portuguese language learning: A case study examining a quest-based website to prompt oral production and interaction in learners of Portuguese L2. *Revista Brasileira de Linguística Aplicada*, 20(4), 733-760. <https://doi.org/10.1590/1984-6398202016396>

**Research Questions**

1. Will the proposed activity intrinsically and extrinsically motivate the learners to communicate in their second language?

**Participants**

The participants consisted of ten English native speakers who were interested in learning Portuguese as a second language. They were equivalent to the B1 level (which is taken from the Common European Framework) and spoke at least one other foreign language. The mean age of the participants was 20-years-old. The participants were members of the researcher's class, but the inclusion of their results mentioned in the study was voluntary.

**Research Design/Procedures**

The study was carried out through class activities and was modeled on the Educational Gamification Five Step Model adapted from Huang and Soman (2013). This particular five-step-model is based on 1) understanding the target audience and the context 2) defining learning objectives 3) structuring the experience 4) identifying resources and 5) applying gamification elements. Using these elements, the researchers created a scenario in which participants went through a simulation of applying for a job in the target language. The basic idea was that students

would first create a CV and a cover letter for the position. They would submit this material and if it passed, they received a badge. Those who received a badge were then invited to role-play an interview for the position. Once this was complete, students completed a questionnaire in which they reflected on the efficiency of the activity and how relevant they felt it was along with how much they enjoyed it. This provided the researchers with a useful map about how influential the activity was in their intrinsic and extrinsic motivation.

### **Results/Discussion**

This study did not have a results section, which was weird to me. However, they claimed that the badges boosted students' motivation and gave them a sense of control and ownership over the activity. They also claimed that the novelty of the activity likely boosted intrinsic motivation and curiosity. The researchers reported that none of the students felt overly pressured or demotivated if they were held back a stage, but did not report how many students were held back or how many actually got the positions they were applying for.

### **Critique**

This article was probably the worst of those that I reported on concerning its validity, importance, and clarity. To start, it doesn't involve collecting and analyzing data. It was important and valuable, but in my mind, does not qualify as research. It was also confusing to read, thinking that it was a research study. Thus, their research questions were not clear and while creating a game is important, the fact that they didn't report on the findings, made their questions less valuable.

As for the study's validity, it had too many variables at play to be valid with too few participants. I think that they were attempting to account for too many things. In reality, this was closer to qualitative research than quantitative research, but they also didn't report their findings

in a way that was very well done for qualitative research. This study could be repeated and I would like to think it would produce the same positive effects on intrinsic and extrinsic motivation, but I think that a lot would need to be changed in order to include more data collection. I do believe that this topic is valuable and I would love to see the difference between two classes, one using gamification and the other not. If I were to redesign this study, I would have two classes charged with applying to an internship/job in a foreign language. For one class, I would use the gamification strategies outlined in this paper. For the other class, I would have students peer review and submit their drafts to me. For either class, I would treat this as an IPA and look at how well they conducted themselves in the mock interview and how well their CV/cover letter depicted their language proficiency. I would then have them complete a survey asking questions much like the ones described in this article. From these quantitative and qualitative data points, I would attempt to draw conclusions about the intrinsic and extrinsic motivation experienced in the different activities. This would be a great mixed-methods study, especially if you created some randomization and treated the traditional group as the control group.