Evaluating Augmented Reality for Situated Vocabulary Learning

Marc Ericson C. SANTOS[a]*, Arno in Wolde LUEBKE[a], Takafumi TAKETOMI[a], Goshiro YAMAMOTO[a], Ma. Mercedes T. RODRIGO[b], Christian SANDOR[a], & Hirokazu KATO[a]
[a]*Interactive Media Design Laboratory, Nara Institute of Science and Technology, Japan*
[b]*Ateneo Laboratory for the Learning Sciences, Ateneo de Manila University, Philippines*
*chavez-s@is.naist.jp

Abstract: Augmented reality (AR) is an emerging technology for communicating learning contents. Several AR systems are designed for learning. However, studies that have investigated instructional strategies for applying AR are few. This investigation requires the implementation of prototypes that use state-of-the-art technology and sound learning theory.

In this work, we implemented two prototypes for learning Filipino and German words by first developing a handheld AR platform. These prototypes demonstrate situated vocabulary learning. Using our AR system, students can learn words related to their current environment. We assessed the quality of these prototypes by conducting usability evaluations. For the theoretical grounding, we leveraged on multimedia learning theory to design the content.

Through our handheld AR platform, we evaluated situated vocabulary learning by comparing our prototypes to a flash cards application. In the first evaluation, students scored significantly lower when using AR in an immediate post-test. However, this difference disappeared after taking into account the variability in usability scores via analysis of covariance. Taking account usability is fairer when comparing an emerging technology to traditional technology. Test scores were also not significantly different in a delayed post-test. In the second evaluation, although the post-test score and answering time of students did not differ, our results showed that they feel more satisfied and can keep their attention better when using AR. For the first time, we demonstrated situated vocabulary learning by using AR. Moreover, our preliminary study confirms the intuition that students can achieve the same score using AR, but with benefits such as ease in maintaining attention and increased satisfaction.

Keywords: augmented reality, mobile learning, situated cognition, vocabulary learning

1. Introduction

Augmented reality (AR) is an emerging technology in the field of computers in education (Wu et al., 2013). We provide a useful summary of AR prototypes applied to learning in our review (Santos et al., 2014). Researchers do not usually take advantage of the most important feature of augmented reality: showing the explicit relationship of the virtual content to objects found in the real world. As such, we implemented and evaluated this display interaction for situated vocabulary learning (Figure 1).
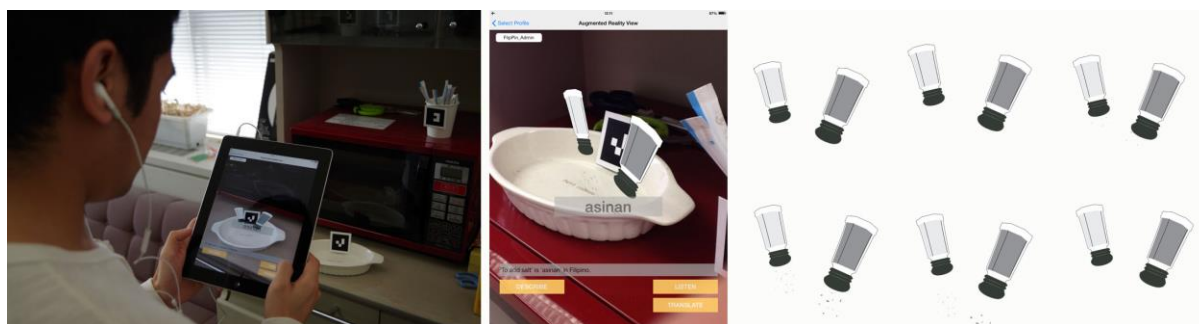
Figure 1. Situated Vocabulary Learning. Student (left) is learning the word "asinan," the Filipino word for adding salt, on a dish inside a kitchen-like environment. Our interface (middle) integrates sprite animations (right) to explicitly illustrate the virtual action on a real object found in the environment.

In situated vocabulary learning, the physical environment is the context of the vocabulary. In this paper, we did a preliminary comparison of our approach (Figure 2) against a flash cards application which doesn't contextualize the vocabulary to the environment. For the first time, we evaluated usability, learning, and motivation in AR-based situated vocabulary learning.
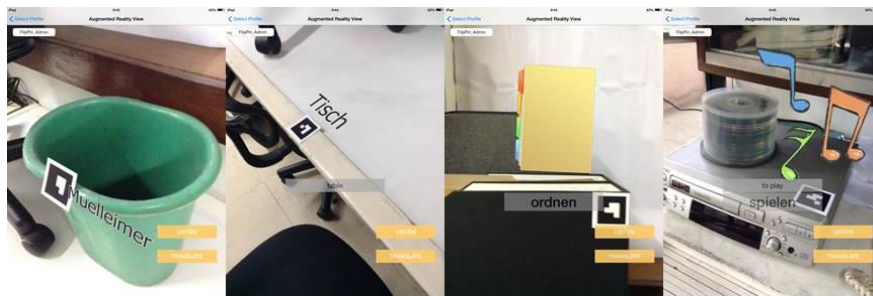

Figure 2. Nouns are displayed as labels, whereas verbs are shown as animations on real objects.

2. Related Works

People are familiar with AR because of the many AR browsers used to locate interesting places in the world (Grubert et al., 2011). People use AR browsers to see virtual labels and symbols integrated with a live video feed of the real environment. Thus, understanding location-related information, such as names of buildings, distances of restaurants, arrows for navigation, and so on, becomes easier. In our AR application, instead of displaying names and direction, our system displays the objects of nouns and illustrates the action of verbs.

Several AR systems have been developed for educational settings (Santos et al., 2014). One important work is Construct3D (Kaufmann et al., 2000; Kaufmann, 2002), which uses AR to teach students mathematics and geometry concepts. AR is suitable because students can interact naturally with three-dimensional shapes without the use of a mouse and keyboard. While wearing a head-mounted display, students can move around virtual shapes and perform operations on them. Moreover, the students see the same shape, thereby allowing them to work together on the target virtual object. Although Construct3D and other works take advantage of embodied cognition and collaborative learning, previous prototypes do not use the main feature of AR. AR, where "3-D virtual objects are integrated into a 3-D real environment in real time" (Azuma, 1997), displays the relationship of the virtual object to the real environment. In the present paper, we teach vocabulary by displaying the relationship between virtual objects and the real environment.

We can apply handheld AR to show the relationship of the educational content to the real environment. Handheld AR has gained attention in the field of educational technology because of its benefits such as ubiquitous learning (Dede, 2011), situated cognition (Specht et al. 2011), and collaboration (Li et al., 2011). Billinghurst and Duenser (2012) argue that handheld AR technology is mature for this application. AR software can already run on mobile phones equipped with fast processors, big display screens, data connection, built-in camera, and other sensors. Billinghurst and Duenser (2012) call for more interdisciplinary research to ground AR applications in learning theories. For our experiments, we designed the content of our AR prototype by applying the principles of multimedia learning theory (Mayer, 2009) and its related research.

2.1 Vocabulary Learning Systems Applying Various Contexts

People learn new words in meaningful contexts. Words are interpreted as a part of a passage of text. As such, Chen et al. (2013) proposed hypertext annotations for supporting quick definitions when reading electronic text. To support vocabulary learning, many other interfaces leverage on constructing various other contexts such as word games (Lin et al., 2008), virtual environments (Pala et al., 2011), collaboration (Joseph et al., 2005), and interaction with robots (Wu et al., 2008).

A natural context for learning vocabulary would be the physical environment. As such, Edge et al. (2011) and Dearman & Truong (2012) propose to present words that are related to the student's environment. Their systems involve the use of handheld devices and GPS positioning to detect the

student's current environment and then present words that are related to that environment. In both systems, the students browse the words on the device screen. We extended these works by using AR to display content onto the physical environment. In our AR system, the vocabularies are either animations or labels on real objects as shown in Figure 2.

2.2 Systems for Situated Vocabulary Learning

Liu's (2009) HELLO system demonstrated how handheld devices can be designed to promote significantly better learning outcomes. The HELLO system uses the campus network to deliver content for an English language learning system. It can detect the location of the user through QR codes spread around the school. At each location, students practice conversations with a virtual tutor on the device. In the user testing phase involving 64 students, Liu reported that students who used their system scored higher compared with those who used printed materials and audio recordings. This effect is attributed to practicing English in situations that could really happen in specific locations.

Beaudin et al. (2007) employed a different strategy to teach Spanish. For their research, they built a smart home learning environment that can detect user movements and intentions by using various sensors. Equipped with a mobile device, the smart home can identify who the user is and present relevant information. For example, they implemented a feature wherein voice-overs of Spanish words or phrases are triggered when users touch specific objects. This interaction makes an explicit connection between the Spanish content and the objects found in the learner's environment, thereby promoting situated cognition. Both the works of Liu and Beaudin et al. take advantage of near-transfer, that is, applying knowledge learned in a particular situation to another situation that is almost similar in context (Dunleavy & Dede, 2014). Combining ideas from these prototypes, we implemented a handheld AR platform (Section 3) to present contextual texts, images, and sounds in the real-world environment.

2.3 Multimedia Learning Applied to Vocabulary Learning

In multimedia learning theory, multimedia refers to pictures and words (both written and spoken). This theory assumes three things, namely, dual-channels, limited capacity, and active processing (Mayer, 2009). First, multimedia learning takes advantage of the two separate channels for perceiving visual and auditory information. Second, it recognizes that individuals have a limited capacity of information that they can attend to. Lastly, learning only takes place if the learner actively processes incoming information by connecting them to prior knowledge.

Given that individuals have a limited capacity of information that they can attend to, Lin and Yu (2012) investigated the cognitive load induced by different types of media presentations on a mobile phone. In their study with 32 eight graders, they investigated the use of four multimedia modes, namely, text, text with audio, text with picture, and text with audio and picture. They discovered that the multimedia mode does not have a significant effect on vocabulary gain and retention. However, the learners rated the combined text-audio-picture as the mode that induced the least cognitive load.

Lin and Wu (2013) investigated the use of these four multimedia modes in a succeeding study with 423 junior high school students. They did not find any significant differences in vocabulary recognition nor in any interaction between multimedia mode and learning style preferences of the student. However, the participants who used text with audio and picture performed best in listening tests followed by the text with sound group. This result confirmed the intuition that audio annotations contribute to the construction of phonological knowledge of words and then applying this knowledge in listening to sentences. More importantly, they reported that the learning effects of the audio were maintained for two weeks with minimal attrition. Based on these works, we implemented features in our AR platform (Figure 3) to allow users to access text, audio, and pictures during the learning scenario.

In a separate study with 121 senior high school students, Lin and Hsiao (2011) studied the effects of the use of still images against simple animations in vocabulary learning. Their results showed that the animation group performed significantly better in learning Chinese and English vocabularies compared with the image group. They recommended the use of animations to illustrate dynamic words and processes. Thus, to facilitate better understanding of vocabulary in our handheld AR platform, we included a feature where sprite sheet animations can be used. We found this feature to be a simple solution to illustrate verbs in our learning scenario.

3. Implementation

We created two AR applications for learning Filipino and German words in a real environment. We achieved this objective by first creating a handheld AR platform that can display any situated multimedia – images, animations, sound, and text displayed on a real environment (Figure 3). We then filled the platform with content for the situated vocabulary learning of Filipino and German words.

3.1 Handheld Augmented Reality Platform

Figure 3 shows the package diagram of our platform and the sample interface enabled by our platform. The main part of the platform is the Controller, which has access to learning contents, sensor (camera), and user inputs. The Controller receives the marker ID and camera view matrix from the Tracker and uses these information to specify the behavior of the on-screen display. The Tracker was built using ARToolkit, and the Renderer was built on OpenGL ES 2.04.
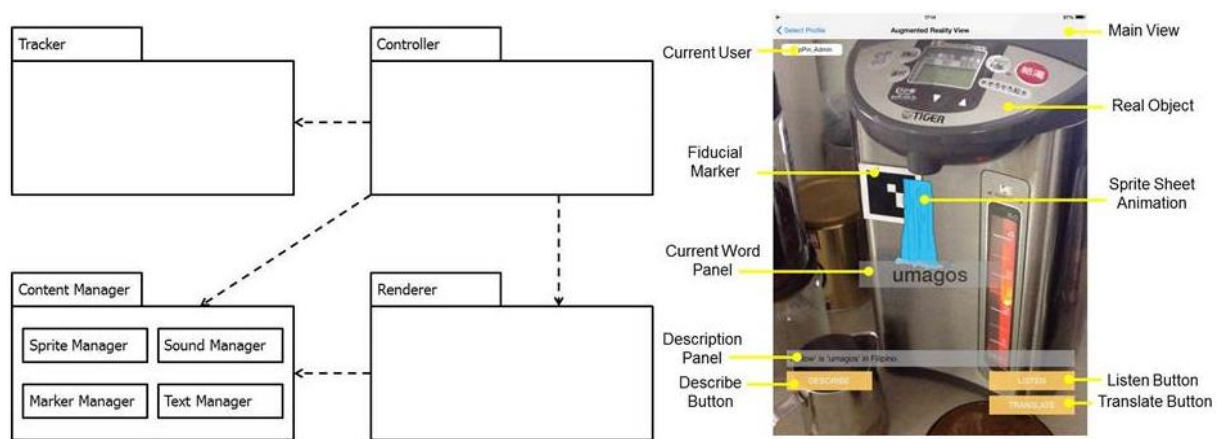

Figure 3. Package Diagram of Our Handheld Augmented Reality Platform (left); Sample Interface for Situated Vocabulary Learning (right)

We used the ARToolkit (Kato & Billinghurst, 1999) to measure the camera pose with respect to the target object. Markers in the video feed were located using the ARToolkit, which also outputs the marker's ID and the matrix representing the current view of the camera. The image was transformed to the correct view using the matrix, and then it was rendered accordingly using OpenGL ES 2.04.

The platform runs entirely on iPad tablets. For our experiments, we used the iPad 2 (dual-core A5, 512MB DDR2 RAM, 32GB, 601 g, 9.7 in display, 1024-by-768 at 132 ppi), and the iPad mini (64-bit A7, 512MB DDR2 RAM, 16GB, 331 g, 7.9 in display, 1024-by-768 at 163 ppi). The platform works with fiducial markers (Figure 3) to determine the target object and the viewing angle of the tablet's back camera. We used the back camera set to 640x480 pixels at 30 fps to sense the marker and to provide a video feed. After identifying the marker, the platform loads the corresponding audio, text, and image. Audio and text can be accessed using buttons (LISTEN, TRANSLATE, DESCRIBE). The images can either be still images or sprite sheet animations (Figure 3; Figure 1). The images are transformed depending on the camera view and are inserted in the video feed to suggest 3-D registration, that is, to give an impression that the graphics co-exist with the real objects.

3.2 Situated Vocabulary Learning Content

We used the platform to construct two situated vocabulary learning systems: one for 30 Filipino words and the other for 10 German words. We based the design of the content from previous works (Lin & Hsiao, 2011; Lin & Yu, 2012; Lin & Wu, 2013) by using a combination of text, audio, images, and animations as content. The text data are the vocabulary, its translation in English, and the description of the scene (only for the Filipino version). The audio data is the proper pronunciation of the vocabulary as spoken by a native speaker. The image data are text labels, images, or labels, as shown in Figure 2.

4. Experiments

We explored the strengths of our AR applications for situated vocabulary learning over its non-situated counterpart (Figure 4) in two preliminary experiments. Through these experiments, we aim to evaluate the use of AR for viewing vocabulary content that is situated in the real environment. We compared the AR applications to a non-situated version which is a tablet application that mimics flash card interaction. Our comparison does not employ any kind of special instructional design such as game mechanics and collaborative learning. As summarized in Table 1, users simply point the tablet PC to objects found in their environment when using our AR application. On the other hand, the flash cards application allows the user to flip through contents by pressing either next or previous.
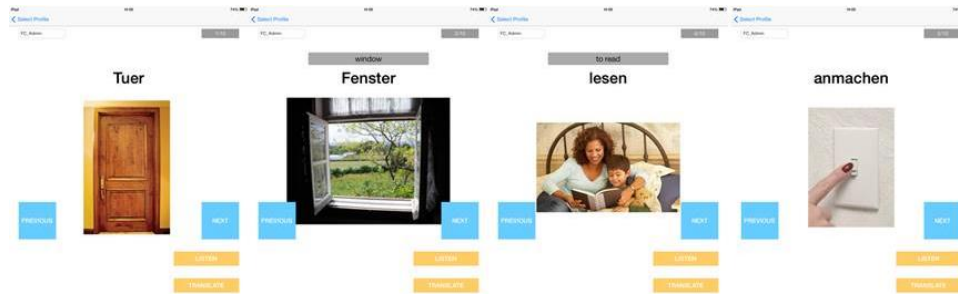

Figure 4. Non-situated version of the AR applications.

Table 1. Comparison of Two Interfaces for Vocabulary Learning

	Situated (AR app)	Non-situated (Flash cards app)
Interaction	Users find an object with a marker. They then point the tablet PC to the marker to reveal the content.	Users press "next" or "previous" to switch between contents.
Inherent Feature	Users can see the markers in their environment even when they are not studying.	Users can quickly go through all the material because they are arranged in a series.
Visual Display	Images and animations are displayed on the real environment.	Static illustrations are shown on a white background.
Place and Time	Users can only use it inside their laboratory at any time.	

We considered inherent features of the interaction as part of the treatment. Thus, interventions were not done to control it. For example, one advantage of an AR learning system is that the students see the objects in their surroundings even when they are not studying. We imagine this feature to trigger unintended rehearsal of the vocabulary, thereby improving learning. This unintended rehearsal is part of AR learning; thus, we did not control this aspect. We did not forbid the students in the situated treatment from visiting the study place when they are not studying.

Another inherent feature is that students tend to cover all the vocabularies several times in one study session when flash cards are used. The flash cards are serially arranged, and students try to go through all the content two to three times in one sitting. Even if this is the case, interventions were not made because it is an inherent feature of the use of flash cards. Moreover, advising the students who use the AR application to view all the content several times will interrupt their natural learning style.

For our experiments, we controlled both location and time constraints. All of our students were only allowed to use the applications inside their respective laboratories. However, the applications are available to them at any time they want to study on that day. Given these features, we had seven hypotheses which we tested for significance in the 0.05 level via student's t-test and analysis of covariance (ANCOVA). The hypotheses are as follows:

H1. Students will perform better in an immediate post-test with non-situated vocabulary learning.
H2. Students will perform better in a delayed post-test with situated vocabulary learning.
H3. Students will rate situated vocabulary learning as a more motivating instructional material.
H4. Students will maintain their attention better with situated vocabulary learning.

H5. Students will find the contents of situated vocabulary learning as more relevant to them.
H6. Students will feel more confident with non-situated vocabulary learning.
H7. Students will feel more satisfied with situated vocabulary learning.

4.1 Experiment 1: Learning Thirty Filipino Words in Five Days

We adapted a between-groups approach with 31 participants (26 male, 5 female, aged 23–42, information science graduate students) to test our application for studying Filipino words. The first languages of the participants are Japanese (13), Chinese (5), Portuguese (3), German, English, Turkish, Bosnian, Indonesian, Finnish, Arabic, Spanish, Nepali, and Wolof. In our experiments, we divided the people into the treatment groups with consideration to the distribution balance of their first languages.

Eighteen participants were recruited from one laboratory. We set up our system inside their laboratory (Figure 5) so that they can learn words related to their refreshment area. All of them have experienced using an AR application before. As such, AR is not a novel technology for them.

Twelve participants from three laboratories were asked to use the non-situated version. Similar to the situated group, the non-situated group have used AR before and they are familiar with other novel interfaces. We distributed tablet computers to them with the flash cards application installed.

Figure 5. Refreshment area with markers (left), Learner using situated vocabulary learning (middle), Learner using non- situated vocabulary learning (right)

The participants used the assigned application for a recommended duration of 10–15 min per day for five days. The situated version was used inside a refreshment area with a maximum of four people using the application at the same time (Figure 5). On the other hand, the learners used the non-situated version wherever they went inside their laboratory office.

In this experiment, we evaluated the participants' learning outcomes and the usability of the application. On the fifth day, the participants answered the System Usability Scale (SUS) to measure the perceived usability of the applications (Lewis & Sauro, 2009). They then immediately took a post-test. After 12–14 days, they took a delayed post-test. The immediate post-test (27 items) and delayed post-test (24 items) consists of questions on recognizing the word in a multiple choice question, recalling the translation of the word, and guessing which word fits in different contexts.

4.2 User Testing 2: Learning German Words

We adapted a within-subjects design with 14 participants (8 male, 6 female, aged 17–20, science majors) to test the application for learning 20 German words (10 situated and 10 non-situated). Each participant used the situated and non-situated versions for a maximum of 8 min. Seven used the situated version first, whereas the other seven used the non-situated version first to balance any effect of the ordering of the treatment. For the situated version, the learners viewed the content on a small area around a laboratory technician's desk. The markers are placed near each other in a small area to minimize the time spent transferring from one object to another. This is important because we wanted to observe the study time of the students. For the non-situated version, they used the application while sitting inside the same room.

The students are then asked to answer 10 multiple choice questions that test their skill to recognize a word. Aside from logging the answer, we also logged the time it took for the learner to

answer the question. After taking the quiz, the participants also answered a subset of the Instructional Materials Motivation Questionnaire or IMMS. We picked 30 questions that are applicable to our system out of the 36 questions listed in the work of Huang et al. (2006). IMMS models the extent of motivation one gets from an instructional material by using the ARCS model (Attention, Relevance, Confidence, and Satisfaction). This model has been applied to AR instructional materials by Di Serio et al. (2013).

5. Results and Discussion

Our experiments involved a small sample size, thus the results should be interpreted with caution. These should be replicated with a bigger sample size. Nevertheless, these results can guide future design of AR applications and experiments in situated vocabulary learning with AR. In our experiments, no significant differences were observed in learning outcomes between situated and non-situated vocabulary learning. However, students report better attention and satisfaction in using our system. We found evidences that support hypotheses H4 and H7 but not H1–3 and H5–6.

5.1 Experiment 1: No significant difference in usability and learnability

We computed the SUS score and its factors from the participant responses in Experiment 1. The results in Table 2 show that the AR application has an SUS score of 74%, which is close to its flash cards application counterpart with 80%. According to Sauro (2011), both interfaces are above average (68%); thus, they are both good interfaces. Moreover, the results in Table 3 show that our participants did not have difficulty in learning these new interfaces.

      We did not find a significant difference between the two interfaces. As such, using these interfaces to compare situated and non-situated vocabulary learning is reasonable. We achieved a good usability score because we applied previous research in multimedia learning. Furthermore, our current interface features are minimal, and the study task is simple.

Table 2. System Usability Scale Scores for Situated and Non-Situated Vocabulary Learning

	Application	N	Mean	Standard Deviation	T value
SUS Score	AR	18	74%	12%	1.64
	Flash cards	13	80%	6%	

Table 3. Factors of the System Usability Scale Scores

Factor	Application	N	Mean	Standard Deviation	T value
Usability	AR	18	70%	14%	1.50
	Flash cards	13	76%	7%	
Learnability	AR	18	90%	13%	1.53
	Flash cards	13	96%	5%	

5.2 Experiment 1: Significantly higher score with non-situated for immediate post-test but not for the delayed post-test

Table 4 is a summary of the results of the immediate and delayed post-tests in Experiment 1. In the immediate post-test, the non-situated group scored significantly higher with a moderate effect (d = 0.75) than the situated group. The breakdown in Table 5 shows that the situated group scored lower than the non-situated group in all types of questions. This result is indicative of an overall inferior mastery of content rather than a weakness in a particular question type.

      In most practical cases, people do not apply their learning immediately after studying. Rather, they would use their knowledge after a few days, either for a test or to apply it to a new lesson. As such, the delayed post-test is a more important point of comparison for learning than the immediate post-test. After 12–14 days, the significant difference in learning disappeared (Table 4).

Table 4. Total Scores in Immediate and Delayed Post-tests

Post-test	Group	N	Mean	Standard Deviation	T value
Immediate	Situated	18	71%	20%	2.14*
	non-situated	13	86%	20%	
Delayed	Situated	12	68%	23%	0.31
	non-situated	13	70%	18%	

*$p < 0.05$

Table 5. Immediate Post-test Scores for Each Question Type

Type	Group	N	Mean	Standard Deviation	T value
With illustrations	situated	18	87%	12%	0.99
	non-situated	13	92%	20%	
Recognizing Filipino with choices	situated	18	80%	15%	2.54**
	non-situated	13	94%	15%	
Recognizing Filipino without choices	situated	18	64%	30%	1.95*
	non-situated	13	83%	24%	
Translating from English to Filipino	situated	18	55%	31%	2.54**
	non-situated	13	81%	23%	
Transfer of word usage with choices	situated	18	75%	19%	2.40*
	non-situated	13	91%	16%	

*$p < 0.05$, **$p < 0.01$

5.3 Experiment 1: No Significant Differences in Immediate Post-test Scores After Considering Usability as Covariant in ANCOVA

If both AR and flash card applications have the same SUS score, then we could do the fairest comparison possible. However, despite our best efforts, a small difference of six SUS points was still observed between the two groups. We conducted ANCOVA to take into account this difference in quality. We assume that the quality of the implementation of the interface affects the students' scores.

ANCOVA was conducted because the difference in SUS score is not significant. We also checked the homogeneity of variance using the Levene's test. The results of the Levene's test showed that no significant differences ($p > 0.05$) were observed. Thus, our data have homogenous variances. The ANCOVA results (Table 6) show that no significant differences were observed in the test scores of situated and non-situated group for both immediate and delayed post-tests. We guess that if we can improve our AR application to the same level as the flash card application, then students can perform equally well with a novel interface.

Table 6. Analysis of Covariance of Post-Test Scores with System Usability Scale Score as Covariant

Test	Group	N	Mean	Standard Deviation	Adjusted Mean	F value	p value
Immediate	situated	18	71%	20%	72%	3.02	0.09
	non-situated	13	86%	20%	85%		
Delayed	situated	12	68%	20%	69%	0.00	1.00
	non-situated	13	70%	16%	69%		

5.4 Experiment 2: No significant difference in Post-test and Motivation, but significantly better Attention and Satisfaction with Situated Vocabulary Learning

No significant differences were observed in the immediate post-test between situated (m = 94%, sd = 8%) and non-situated (m = 95%, sd = 8%) vocabulary learning. On the average, the non-situated group answered our multiple questions faster (m = 2.28 s, sd = 0.92 s) than the situated group (m = 2.60 s, sd = 1.03 s) for each question. However, this difference was not significant.

Experiment 2 focuses on evaluating motivation by using the ARCS model. Although two interfaces can arrive at the same learning result, performance in tests should not be the only measure of success in creating interfaces. User experience is another important consideration. As such, we also evaluated the interfaces in terms of its ability to motivate students to learn.

Overall, no significant difference was observed in the IMMS rating of situated and non-situated vocabulary learning (Table 7). However, looking at the factors of the IMMS (Table 8), significant differences were observed in the attention and satisfaction factors. The students report that the AR application catches and holds their attention more than the flash cards. Moreover, they report higher satisfaction with their learning experience. The learners were slightly more confident to use flash cards probably because it is a more familiar interface. The learners rated AR to be higher in relevance by five points, which is attributed to the implicit connection between learning contents and real environment. However, no statistical significance was observed for the relevance and confidence factors.

Table 7. Instructional Material Motivation Survey Scores for Situated and Non-situated Scenarios

	Treatment	N	Mean	Standard Deviation	T value
Motivation Score	situated	14	76%	12%	1.34
	non-situated	14	71%	11%	

Table 8. Factors of the Instructional Material Motivation Survey Scores

Factor	Treatment	N	Mean	Standard Deviation	T value
Attention	situated	14	75%	14%	1.84*
	non-situated	14	65%	14%	
Relevance	situated	14	74%	14%	0.97
	non-situated	14	69%	13%	
Confidence	situated	14	80%	12%	0.74
	non-situated	14	83%	8%	
Satisfaction	situated	14	77%	16%	1.71*
	non-situated	14	66%	18%	

*$p < 0.05$

6. Conclusion

We are the first to use AR for explicitly displaying the relationship between vocabulary learning contents and real world environment for situated vocabulary learning. This preliminary study supports our intuition that AR can enable the same knowledge acquisition with added benefits of better attention and satisfaction. We did not employ special instructional strategies in our experiments such as game mechanics or collaboration between students. As such, the differences in the learning experience are attributed to the inherent advantages or disadvantages of the interfaces: augmented reality and flash cards representing situated and non-situated vocabulary learning, respectively.

Our system can be improved by applying other learning theories and instructional strategies that are not possible for traditional interfaces. Currently, we applied multimedia learning theory because AR is essentially a presentation medium. In the future, we can apply insights on location-based games and collaborative learning to create better augmented reality learning experiences.

References

Azuma, R. T. (1997). A survey of augmented reality. *Presence, 6*(4), 355-385.

Beaudin, J. S., Intille, S. S., Tapia, E. M., Rockinson, R., & Morris, M. E. (2007). Context-sensitive microlearning of foreign language vocabulary on a mobile device. In *Ambient Intelligence* (pp. 55-72). Springer.

Billinghurst, M., & Duenser, A. (2012, July). Augmented Reality in the Classroom. *Computer, 45*(7), 56-63.

Chen, I., Yen, J. C., & others. (2013). Hypertext annotation: Effects of presentation formats and learner proficiency on reading comprehension and vocabulary learning in foreign languages. *Computers & Education, 63*, 416-423.

Dearman, D., & Truong, K. (2012). Evaluating the implicit acquisition of second language vocabulary using a live wallpaper. *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, (pp. 1391-1400).

Dede, C. (2011). Emerging technologies, ubiquitous learning, and educational transformation. In *Towards Ubiquitous Learning* (pp. 1-8). Springer.

Di Serio, A., Ibanez, M. B., & Kloos, C. D. (2013). Impact of an augmented reality system on students' motivation for a visual art course. *Computers & Education, 68*, 586-596.

Dunleavy, M., & Dede, C. (2014). Augmented reality teaching and learning. In *Handbook of research on educational communications and technology* (pp. 735-745). Springer.

Edge, D., Searle, E., Chiu, K., Zhao, J., & Landay, J. A. (2011). MicroMandarin: mobile language learning in context. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (pp. 3169-3178).

Grubert, J., Langlotz, T., & Grasset, R. (2011). *Augmented reality browser survey.* Tech. rep., Institute for Computer Graphics and Vision, University of Technology Graz.

Huang, W., Huang, W., Diefes Dux, H., & Imbrie, P. K. (2006). A preliminary validation of Attention, Relevance, Confidence and Satisfaction model-based Instructional Material Motivational Survey in a computer-based tutorial setting. *British Journal of Educational Technology, 37*(2), 243-259.

Joseph, S., Binsted, K., & Suthers, D. (2005). PhotoStudy: Vocabulary learning and collaboration on fixed & mobile devices. *Wireless and Mobile Technologies in Education, 2005. WMTE 2005. IEEE International Workshop on*, (pp. 5--pp).

Kato, H., & Billinghurst, M. (1999). Marker tracking and hmd calibration for a video-based augmented reality conferencing system. *Augmented Reality, 1999.(IWAR'99) Proceedings. 2nd IEEE and ACM International Workshop on*, (pp. 85-94).

Kaufmann, H., Schmalstieg, D., & Wagner, M. (2000). Construct3D: A Virtual Reality Application for Mathematics and Geometry Education. *Education and Information Technologies, 5*(4), 263-276. Retrieved from http://dx.doi.org/10.1023/A%3A1012049406877

Lewis, J. R., & Sauro, J. (2009). The factor structure of the system usability scale. In *Human Centered Design* (pp. 94-103). Springer.

Li, N., Chang, L., Gu, Y. X., & Duh, H. B. (2011, July). Influences of AR-Supported Simulation on Learning Effectiveness in Face-to-face Collaborative Learning for Physics. *Advanced Learning Technologies (ICALT), 2011 11th IEEE International Conference on*, (pp. 320-322).

Lin, C. C., & Hsiao, H. S. (2011). The Effects of Multimedia Annotations via PDA on EFL Learners' Vocabulary Learning. *Proceedings of the 19th International Conference on Computers in Education.*

Lin, C. C., & Wu, Y. C. (2013). The Effects of Different Presentation Modes of Multimedia Annocations on Sentential Listening Comprehension. *Proceedings of the 21th International Conference on Computers in Education.*

Lin, C. C., & Yu, Y. C. (2012). EFL Learners' Cognitive Load of Learning Vocabulary on Mobile Phones. *Proceedings of the 20th International Conference on Computers in Education.*

Lin, C. P., Young, S. C., & Hung, H. C. (2008). The game-based constructive learning environment to increase English vocabulary acquisition: Implementing a wireless crossword Fan-Tan game (WiCFG) as an example. *Wireless, Mobile, and Ubiquitous Technology in Education, 2008. WMUTE 2008. Fifth IEEE International Conference on*, (pp. 205-207).

Liu, T. Y. (2009). A context-aware ubiquitous learning environment for language listening and speaking. *Journal of Computer Assisted Learning, 25*(6), 515-527.

Mayer, R. E. (2009). *Multimedia learning.* Cambridge university press.

Pala, K., Singh, A. K., & Gangashetty, S. V. (2011). Games for Academic Vocabulary Learning through a Virtual Environment. *Asian Language Processing (IALP), 2011 International Conference on*, (pp. 295-298).

Santos, M., Chen, A., Taketomi, T., Yamamoto, G., Miyazaki, J., & Kato, H. (2014, March). Augmented Reality Learning Experiences:Survey of Prototype Design and Evaluation. *Learning Technologies, IEEE Transactions on, 7*(1), 38-56.

Sauro, J. (2011). Measuring usability with the system usability scale (SUS). *Measuring usability with the system usability scale (SUS)*. Retrieved from http://www.measuringusability.com/sus.php

Specht, M., Ternier, S., & Greller, W. (2011). Mobile augmented reality for learning: A case study. *Journal of the Research Center for Educational Technology, 7*(1), 117-127.

Wu, C. C., Chang, C. W., Liu, B. J., & Chen, G. D. (2008). Improving vocabulary acquisition by designing a storytelling robot. *Advanced Learning Technologies, 2008. ICALT'08. Eighth IEEE International Conference on*, (pp. 498-500).

Wu, H. K., Lee, S. W., Chang, H. Y., & Liang, J. C. (2013). Current status, opportunities and challenges of augmented reality in education. *Computers & Education, 62*, 41-49.