Ogata, H. et al. (Eds.) (2015). Proceedings of the 23rd International Conference on Computers in Education. China: Asia-Pacific Society for Computers in Education

The Effect of Human Factors on User Usability of a Customized Augmented Reality English Learning System

Gwo-Haur HWANG[1*], Beyin CHEN[2] & Heng-Lin HUANG[2]
[1]*Department of Information Networking and System Administration, Ling Tung University, Taiwan*
[2]*Department of Information Management, Ling Tung University, Taiwan*
*ghhwang@teamail.ltu.edu.tw

Abstract: Many scholars indicated that augmented reality (AR) can effectively enhance learners' learning motivation and effectiveness. In addition, many researches showed that human factors will affect learning, such as prior knowledge, cognitive style and even gender. So, a customized AR English learning system was developed in this study. Also, we made a usability assessment analysis for this system. According to the results of the analysis, there are some significant differences between gender, education background and department when using the customized AR English learning system.

Keywords: Human factors, user usability, customized, augmented reality, English learning system

1. Introduction

Taiwan is an EFL (English as Foreign Language) English learning environment. In Taiwan, a formal exposure to English is until entering elementary schools (Lu, 2012). In the traditional teaching environments, students learned English through explaining the rigid text of the textbooks by teachers. The English learning then can be conducted. But it is passive learning (Savignon, 1988). Looi et al. (2010) have used a context-aware ubiquitous learning system to record the situation of learners. It allowed teachers to rapidly understand students' learning process and individual differences and to develop a better teaching method. By this, students can get the needed information in appropriate timing (Chen, Lien, & Lu, 2009). According to the "Spatial Contiguity Principle" and "Temporal Contiguity Principle" of 12 multimedia design principles proposed by Mayer (2009), the learning effectiveness of learners will be increased if the corresponding and associated information is able to be immediately generated beside the objects at the time of scanning real objects.

The augmented reality (AR) is a kind of technology that can combine the virtual information and the real image (Azuma, 1997). The technology allows that the learning process can more meet the above principles proposed by Mayer (2009). In addition, Chen and Macredie (2010) have proposed that human factors will affect learning, such as prior knowledge, cognitive style and even gender. So, this study has developed a customized English learning system based on AR technology, and then made a usability assessment analysis with it. According to the results of the analysis, we have found out there are some significant differences among different kinds of human factors. Based on these findings, we can design these types of AR learning systems for the different kinds of users in the future and let the system more matching the characteristics of each person.

2. Literature Review

2.1 The problems of English learning

In the past, second/foreign language learning relied upon teachers' lecturing to explain the learning materials of textbooks (Savignon, 1988). In other words, because the teaching time is limited,

traditional teaching is considered unable to enhance learning motivation and interest (Brown, Collins, & Duguid, 1989). Looi et al. (2010) have indicated that learning will be able to be conducted at any time and any place when technologies of context-aware and ubiquitous learning are imported. Thus, teachers can guide students to learn actively and attract their attention, and learners' ability of observation of the real world and the ability to actually solve problems can be enhanced (Chen, Lien, & Lu, 2009).

Among the technologies of context-awareness, RFID and QR code technologies have the disadvantages of information discontinuous problems. In this study, we use AR technology to implement an English learning system, and make a usability assessment with it. It is hoped that the learning motivation and learning effectiveness of learners can be enhanced. Thus, according to the results of the analysis, we will improve the AR English learning system.

2.2 The influence of human factors in learning

Chen and Macredie (2010) have proposed that human factors will affect learning, such as prior knowledge, cognitive style and even gender which may cause different effects. The main focus of personalized learning is to understand learners' individual factors, such as their extent and learning experience. Then, according to the learners' individual differences, the learning system can adjust the learning materials (Cho, Kim, & Kim, 2003). According to the previous studies, we added personalized content into the AR English learning system; it can allow users adjust the learning environment according to their preferences. In this study, we want to explore whether there are significantly differences in system usability assessment and using attitudes between different kinds of human factors, such as gender, knowledge background, using experience, etc.

2.3 Customized AR English learning system

A customized AR English learning system was proposed and designed by Hwang, Chen and Huang (2014). For customized design, the system adds the functions of related words and 3 learning scopes, i.e., phrases, sentences and related words. In addition, an easy mode and an advanced mode are provided. User can also choose the boy's or girl's pronunciation in the settings menu with their preference at any time.

When learners successfully scan an object, the screen will show the main teaching material and the related material. After the learners click the object that want to learn, the corresponding function buttons will appear at the top of screen. If learners click the function button, the system will read out the materials, and the Chinese meaning will be explained in Chinese voice. In addition, all users' behavior operation process will be recorded in a portfolio database by the behavior-code. The screenshots of the customized AR English learning system is as shown in Figure 1.

Figure 1. The screenshots of Customized AR English learning system.

2.4 Nielsen usability assessment

When a system is developed, it must be tested and evaluated by users. Then, developers can improve the system according to users' needs. Virvou and Katsionis (2008) indicated that system usability will affect the learners' performance and their perception. Greenberg, Fitzpatrick, Gutwin, and Kaplan (2000) pointed out that heuristic evaluation is the most rapid, cheap, and effective way to identify usability problems. In this study, Nielsen's heuristic approach is selected because it is the most commonly used and can be effectively applied by both novices and experts (Nielsen, 1994a; Nielsen & Mack, 1994). Nielsen (1994b) assessment proposed 10 heuristics, such as interface design, system flexibility and efficiency, the system interactivity, etc. The details of the revised set of 10 heuristics (H) are as shown in Table 1. So, in this study, we hope to analyze users' opinion through the Nielsen assessment questionnaire and improve the customized AR English learning system in the future.

Table 1: Nielsen's 10 heuristics (Nielsen, 1994b).

Heuristics	Explanations
H1: Visibility of system status	The system should always keep user informed about what is going on by providing appropriate feedback within reasonable time
H2: Match between system and the real world	The system should speak the user's language, with words, phrases and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order
H3: User control and freedom	Users should be free to develop their own strategies, select and sequence tasks, and undo and redo activities that they have done, rather than having the system do these for them
H4: Consistency and standards	Users should not have to wonder whether different words, situations, or actions mean the same thing and the system should follow platform conventions
H5: Error prevention	Even better than good error messages is a careful design, which prevents a problem from occurring in the first place
H6: Recognition rather than recall	Make objects, actions, and options visible. The users should not have to remember information from one part of the dialogue to another. Instructions for use of the system should be visible or easily retrievable whenever appropriate
H7: Flexibility and efficiency of use	Allow users to tailor frequent actions. Provide alternative means of access and operation for users who differ from the "average" user (e.g., physical or cognitive ability, culture, language, etc.)
H8: Aesthetic and minimalist design	Dialogues should not contain information that is irrelevant or rarely needed. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility
H9: Help users recognise, diagnose and recover from errors	Error messages should precisely indicate the problem and constructively suggest a solution. They should be expressed in plain language
H10: Help and documentation	Even though it is better if the system can be used without documentation, it may be necessary to provide help and documentation. Any such information should be easy to search, focused on the user's task, list concrete steps to be carried out, and not be too large

3. Methodology

3.1 Research design

First of all, we collected relevant literatures. Then, we confirmed the specifications of the system and started to develop the system. When the system was developed successfully, we invited about 70 students including college students and Masters of the information college to do a trial of the system, and let them fill out the Nielsen's assessment questionnaire. After the questionnaires being collected, the statistical software of SPSS 19 was used to analyze and explore the possible effect of different kinds of human factors on user usability of this system. Finally, in this paper, we will propose the amendments of system according to the analysis results and improve the customized AR English system in the future. The flow chart of our research process is as shown in Figure 2.

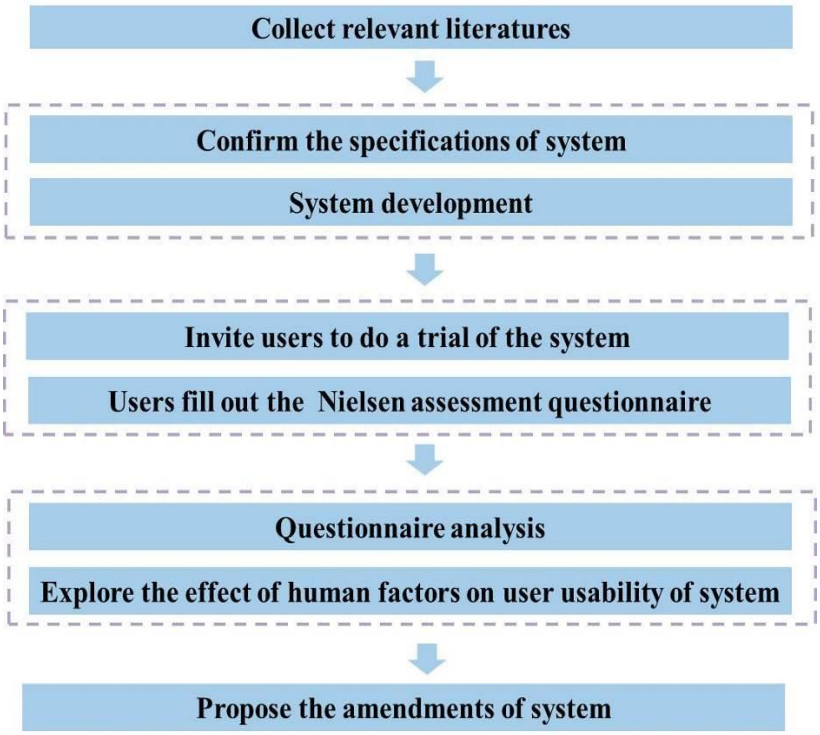

Figure 2. The research process.

3.2 Questionnaire design

In this study, the statistical software of SPSS 19 is used. And then the questionnaire is designed with Nielsen's 10 heuristics and 5-point Likert scale (Likert, 1932) where 1 is strongly disagree, 2 is disagree, 3 is neither agree nor disagree, 4 is agree, and 5 is strongly agree. There are total 60 questions in the questionnaire, which include 5 positive questions and one reverse question for each Heuristic (H).

After collecting questionnaires, in order to confirm the reliability of the questionnaire, we conducted a questionnaire reliability analysis first. As shown in Table 2, the result shows that the Cronbach's Alpha values of all heuristics (H) are all higher than 0.8. It means the reliability of this questionnaire is good enough to do the following analysis. Then, to explore whether there are significantly differences between different kinds of human factors or not, the *t*-test and the analysis of variance (one-way ANOVA) are used.

Table 2: The reliability of the questionnaire for each heuristic (H).

Heuristics (H)	Cronbach's Alpha Value
H1	0.85
H2	0.80
H3	0.84
H4	0.81
H5	0.92
H6	0.83
H7	0.89
H8	0.93
H9	0.88
H10	0.92

3.3 Population and sample

The valid participants of the experiment are 65 information college students from a university in central Taiwan, including 35 students in department of Information Networking and System Administration (INSA), 16 students in Information Management (IM) and 14 students in Information Technology (IT). Among 65 students, there are 51 male users and 14 female users. In addition, there are 46 college students and 19 master students. The reason why we chose these students is trying to find out whether there exists significantly difference in usability and using attitudes among INSA, IM and IT departments in this customized AR English learning system.

3.4 Data collection procedures

The process of the experiment is listed as shown in Figure 3. At first, all participants have to watch the system operating instruction video. Afterwards, the participants login the customized AR English learning system with their accounts and passwords to use the system, as shown in Figure 4. The participants have to fill out Nielsen's Heuristic evaluation questionnaires. Finally, for each heuristic, we interview with particular users with low mean points after the questionnaire being analyzed.

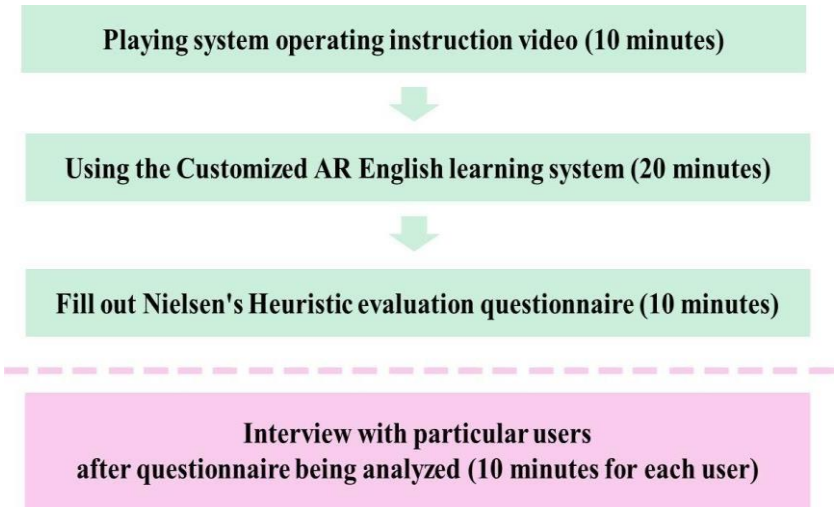

Figure 3. The process of the experiment.

Figure 4. The situation during the experiment.

4. Data analysis and Results

4.1 The 10 heuristics (H) differences between genders

First of all, in order to verify the differences of the usability evaluation between genders, *t*-test is used. According to the analysis, it is found that the recognition mean values of male users had significantly higher than those of female users on H2 and H9; with $t = 2.41$ ($p = 0.019 < 0.05$) and $t = 2.23$ ($p = 0.029 < 0.05$), and medium effect sizes $d = 0.67$ and $d = 0.62$, respectively. The recognition mean values of H2 and H9 in male users (4.60 and 4.51) was higher while those of the genders in female users (4.26 and 4.18) were relatively lower, as shown in Table 3.

Table 3: The *t*-test results for genders.

	Gender	N	Mean	S.D.	*t*	*d*
H1	male	51	4.61	0.44	1.58	
	female	14	4.38	0.58		
H2	male	51	4.60	0.43	2.41*	0.67
	female	14	4.26	0.56		
H3	male	51	4.53	0.46	1.58	
	female	14	4.30	0.61		
H4	male	51	4.36	0.53	1.67	
	female	14	4.10	0.51		
H5	male	51	4.71	0.44	1.56	
	female	14	4.42	0.65		
H6	male	51	4.45	0.48	0.92	
	female	14	4.31	0.60		
H7	male	51	4.48	0.47	0.89	
	female	14	4.31	0.66		
H8	male	51	4.25	0.59	0.43	
	female	14	4.16	0.92		
H9	male	51	4.51	0.46	2.23*	0.62
	female	14	4.18	0.60		
H10	male	51	4.54	0.51	1.87	
	female	14	4.23	0.71		

*$p<.05$

4.2 The 10 heuristics (H) differences between educational backgrounds

Then, in order to verify the differences of the usability evaluation between different educational backgrounds, *t*-test is used. According to the result of analysis, it is found that the recognition mean values of college students had significantly lower than those of Masters on H1, H3 and H5; with $t = -2.91$ ($p = 0.005 < 0.05$), $t = -2.80$ ($p = 0.007 < 0.05$) and $t = -2.27$ ($p = 0.027 < 0.05$) , and medium effect sizes $d = -0.73$, $d = -0.66$ and $d = -0.56$, respectively. The recognition mean values of H1, H3 and H5 of educational backgrounds in Masters (4.78, 4.69 and 4.83) were higher while those of educational backgrounds in college students (4.47, 4.40 and 4.58) were relatively lower, as shown in Table 4.

Table 4: The *t*-test results for different educational backgrounds.

	Educational background	N	Mean	S.D.	*t*	*d*
H1	college student	46	4.47	0.51	-2.91**	-0.73
	Master	19	4.78	0.32		
H2	college student	46	4.48	0.52	-1.50	
	Master	19	4.64	0.33		

H3	college student	46	4.40	0.55	-2.80**	-0.66
	Master	19	4.69	0.29		
H4	college student	46	4.26	0.56	-1.21	
	Master	19	4.43	0.44		
H5	college student	46	4.58	0.54	-2.27*	-0.56
	Master	19	4.83	0.32		
H6	college student	46	4.43	0.53	0.14	
	Master	19	4.41	0.46		
H7	college student	46	4.40	0.54	-1.13	
	Master	19	4.56	0.44		
H8	college student	46	4.31	0.68	1.49	
	Master	19	4.04	0.61		
H9	college student	46	4.40	0.54	-1.25	
	Master	19	4.57	0.42		
H10	college student	46	4.41	0.61	-1.71	
	Master	19	4.64	0.44		

*p<.05; **p <.01

4.3 The effect of 10 heuristics (H) among different departments

Next, in order to verify the differences of the usability evaluation among 3 departments (INSA, IM and IT) in the information college, analysis of variance (one- way ANOVA) is used as shown in Table 5. The results showed that the recognition mean values of H8 (aesthetic and minimalist design) among 3 different departments has significantly difference in multiple comparisons (LSD) ($p=0.41 < 0.05$, $F = 3.36$). The recognition mean value of H8 of departments in INSA is 4.31, while that of departments in IM is 3.88 and that of departments in IT is 4.29. It means that, for H8, the level of acceptance of students of INSA is higher than that of IM. And the level of acceptance of students of IT is higher than that of IM. But there is no significantly difference between INSA and IT.

Table 5: The results of ANOVA for different departments on H8.

	Department	N	Mean	S.D.	*F*	Multiple comparisons (LSD)
H8	(a) INSA	35	4.31	0.67	3.36*	(a>b)
	(b) IM	16	3.88	0.70		(c>b)
	(c) IT	14	4.29	0.53		
	Total	65	4.23	0.67		

*p<.05

Finally, the results showed that H1 (visibility of system status), H3 (user control and freedom) and H10 (help and documentation) have significantly differences among 3 different departments in Post-Hoc test (Games-Howell), as shown in Table 6. The recognition mean (S.D., standard deviation) values of departments in INSA on H1, H3 and H10 were 4.48 (0.46), 4.36 (0.51) and 4.45 (0.52), while those of departments in IM were 4.48 (0.59), 4.51 (0.58) and 4.26 (0.74), and while those of departments in IT were 4.85 (0.23), 4.75 (0.24) and 4.80 (0.31). It means that, for H1 and H3, the level of acceptance of the students of IT is higher than that of INSA. For H10, the level of acceptance of the students of IT is higher than those of INSA and IM. But there are no significantly differences between INSA and IM.

Table 6: The results of ANOVA for different departments on H1, H3 and H10.

	Department	N	Mean	S.D.	*F*	Post Hoc (Games-Howell)
H1	(a) INSA	35	4.48	0.46	3.63*	(c>a)

	(b)	IM	16	4.48	0.59		
	(c)	IT	14	4.85	0.23		
		Total	65	4.56	0.48		
	(a)	INSA	35	4.36	0.51		(c>a)
H3	(b)	IM	16	4.51	0.58	3.28*	
	(c)	IT	14	4.75	0.24		
		Total	65	4.48	0.50		
	(a)	INSA	35	4.45	0.52		(c>a)
H10	(b)	IM	16	4.26	0.74	3.71*	(c>b)
	(c)	IT	14	4.80	0.31		
		Total	65	4.48	0.57		

*p<.05

5. Discussions

Firstly, we found that the recognition mean value of male users (4.60) is higher than that of female users (4.26) on H2. It means that male users feel the system is more in line with real-life logic than female users. In addition, the recognition mean value of male users (4.51) is also higher than that of female users (4.18) on H9. After interviewing with 2 male and 2 female users, the 2 female users pointed out that system operating instructions button can add some captions to make the explanation clearer. The 2 male users thought that the buttons and menu contents shown in the system can be used in common sense easily. This result is coincided with theory proposed by Misu (2001), i.e., the male users often have more positive attitude than female users when using computer technology for learning and female users need more detailed documentation when using the computer technology. So, we can added more detailed explanations in this kind of customized AR English learning system for the female users in future to enhance the usability and satisfaction of the system.

Secondly, we found that the recognition mean values of college students and Master users were 4.47 and 4.78 on H1, while those of them on H3 were 4.40 and 4.69, and while those of them on H5 were 4.58 and 4.83. For all these three mean values of heuristics, Masters have higher values than college students. After interviewing with 8 college students and 8 Master users, for H1 (visibility of system status), all college students thought that it is difficult to recognize the “real learning object” and “virtual related learning object” when learning with this system. Conversely, the Master users thought it can easily recognize which “object” they were learning with the Chinese text materials. For H3 (user control and freedom), the respondents of college students said that they had no particular views on the choice of pronunciation style, but whether the pronunciation can be heard clearly or not is most important when learning with this system. However, the respondents of Master users thought that since the functions of this system are developed mainly for 5-6 grade elementary school students, this kind of design can make these students focus on system operating more intently due to the personalization setting functions. For H5 (error prevention), the respondents of college students suggested that the system can provide more text prompts on the inputted screen of the accounts and passwords before login the system. But, the Master users thought that the documentation for instructions has provided detailed explanation in the “system explanations interface”. From the above, the Master users have deeper observation and analysis than college students. The reason may be that the system development experiences and logical reasoning ability of the Master users in the information college are stronger than college students. This result is coincided with theories proposed by Chen and Macredie (2010) and Virvou and Katsionis (2008), i.e., the prior knowledge will affect the views of the system usability to this system.

Thirdly, we found that the recognition mean value of departments in IM (3.88) is lower than in INSA (4.31) and IT (4.23) on H8 (aesthetic and minimalist design). There were 3 respondents for each department. According to the results of interviews, all respondents of IM thought that the system's graphic design are not perfect and can be more refined. However, the respondents of INSA and IT thought that it is interesting of learning with AR technology while graphic design is secondary consideration. The reason for this situation may be the students of IM take art-related courses more than

those of INSA and IT. In addition, the students of IM were enrolled to school in commercial category more than those of INSA and IT.

Finally, we also found that the recognition mean values of H1, H3 and H10 on departments of IT were 4.85, 4.75 and 4.80 which were higher than those in departments of INSA and IM. The reason for this situation may be the students in the departments of IT take more programing courses than the students on the department of INSA and IM. Overall, most of users thought that the usability of this customized AR English learning is very well and interesting.

6. Conclusions & Future Works

This study has developed a customized augmented reality English learning system and completed a usability assessment of it. According to the results of the analysis, we found that some human factors (gender, educational background and departments) will certainly cause significant difference when using this type of AR learning system, and we will improve the current system from the results.

In the near future, we will invite 2 teachers in an elementary school in central Taiwan to provide 100 English vocabularies suitable for grade 5 students and to conduct experimental teaching. The experimental subjects are 100 students of 4 classes, and the time for experiment is about total 8 weeks for 2 hours per week. All the operation process will be recorded in database so that further personalized preference analysis can be conducted in the future.

Acknowledgements

This study is supported in part by the Ministry of Science and Technology of the Republic of China under Contract No. MOST 103-2511-S-275-002-MY2.

References

Azuma, R. (1997). A survey of augmented reality. *Presence: Teleoperators and Virtual Environments, 6*, 355-385.

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, (18)*, 32-42.

Chen, S. Y. & Macredie, R. (2010). Web-based interaction: A review of three important human factors. *International Journal of Information Management, 31*(6), 1-9.

Chen, Y. H., Lien, C. J. & Lu, Y. L. (2009). *Influences of different learning preferences on learning achievement in E-learning*. Paper presented at the 13th Global Chinese Conference on Computers in Education (GCCCE 2009). Taipei, Taiwan.

Cho, Y. H., Kim, J. K., & Kim S. H. (2013). A personalized recommender system based on web usage mining and decision tree induction. *Expert Systems with Applications, 23*(3), 329-342.

Greenberg, G., Fitzpatrick, G., Gutwin, C., & Kaplan, S. (2000). Adapting the locales framework for heuristic evaluation of groupware. *Australian Journal of Information Systems, 7*, 102-108.

Hwang, G. H., Chen, B., & Huang, H. L. (2014). *Development of a customized English learning system based on augmented reality technology.* Proceedings of the 22nd International Conference on Computers in Education (ICCE 2014). Japan, Nara.

Likert, R. (1932). A Technique for the measurement of attitudes. *Archives of Psychology, 22*, 1-55.

Looi, C. K., Seow, P., Zhang, B.M, So, H. J., Chen, W., & Wong, L. H. (2010). Leveraging mobile technology for sustainable seamless learning: A research agenda. *British Journal of Educational Technology, 41*(2), 154-169. doi:10.1111/j.1467-8535.2008.00912.x.

Lu, C. M. (2012). *Application of simple quasi-markless augmented reality: A case study of English conversation.* National Taipei University of Education, Taipei City.

Mayer, R. E. (2009). *Multimedia learning, second edition.* Cambridge University Press.

Misu, N. A. (2001). The cultural construction of the computer as a masculine technology: an analysis of computer advertisements in Korea. *Asian Journal of Women's Studies, 7*(3), 93-114.

Nielsen, J. (1994a). Usability inspection methods. *Conference Companion on Human Factors in Computing Systems*, 413-414.

Nielsen, J. (1994b). *Heuristic evaluation. Usability inspection methods*. New York: John.

Nielsen, J., & Mack, R. L. (Eds). (1994). *Usability inspection methods*. New York, NY: John Wiley & Sons.
Savignon, S. J. (1988). In second language acquisition / foreign language learning, nothing is more practical than a good theory. *Issue and Developments in English and Applied Linguistics, 3*, 84-98.
Virvou, M., & Katsionis, G. (2008). On the usability and likeability of virtual reality games for education: The case of VR-ENGAGE. *Computers & Education, 50*(1), 154-178.