Published in Mindtrek '18 Proceedings of the 22nd International Academic Mindtrek Conference pp. 236-243. New York: ACM.ISBN: 978-1-4503-6589-5. https://doi.org/10.1145/3275116.3275150.

Finnish Upper Secondary Students User Expectations and Experiences Using MALL System

Jaakko Vuorio[1], Jussi Okkonen[1] & Jarmo Viteli[1]

[1]University of Tampere, Faculty of Communication Sciences, Tampere, Finland

Full Paper

ABSTRACT

The skills to master foreign languages (FL) are increasingly acquired by using interactive mobile technology delivering time and context free learning activities. As mobile learning is claiming space in the formal education and schools are increasingly adopting novel technology to their pedagogical approaches, there is still much to be discovered on how learners perceive technology-enhanced language learning in a mobile learning context. A study was carried out to discover subjective user expectations and user experiences and their differences among Finnish upper secondary students (n=15) using a Mobile-Assisted Language Learning (MALL) system. Subjective Metrics with User Experiments (SUXES) and System Usability Scale were used to study user experiences and usability aspects of the system under assessment. Content analysis was included to deliver implications for both developers and educators to examine the outcomes of using these systems for educational purposes. Results indicate that the MALL system was not thoroughly meeting subjective user expectations raising questions on usability and user experience aspects. In addition, detailed information is delivered for further development and implementation in educational context.

1 INTRODUCTION

Language learning is increasingly maneuvered with and through technology by the cause of advanced internet availability and quality as well as robust hardware such as smart phones. Alongside with formal education, mobile technologies offer everyday opportunities to access learning resources [1]. As mobile technology continues to develop further adding more performances and feasibilities to use for educational purposes, language learning, among other disciplines, has proven to be a promising context for the use of mobile technology. At their best, foreign language (FL) learning enhanced with mobile technology may leverage learner interest, motivation, increased access to learning material, and provide new ways for interaction and feedback. In addition, innovative technologies may allow instructors to interact with numerous students and arrange the learning content in an efficient manner. At their worst, new technologies may result in dubious input, flat interaction and defective feedback as well as frustration with both software and hardware leading to diversion from the original learning task and overall over-emphasis on delivery modality over learning goals. [2] With both opportunities and challenges, there is a need to study these mobile technologies from the end-user perspective. Especially, how learners perceive using these systems and what are the further implications for software developers to deliver context aware services. In addition, educators desire evidence of using such technologies to deliver their students the best means available for both personal skills and academic success. This study focuses on Mobile-Assisted Language Learning (MALL) system used among Finnish upper secondary students to study respondents' user expectations and user experiences and their differences.

1.1 Mobile Learning

The concept of mobile learning recognized in 2005, falls under the umbrella of learner-centered pedagogies [3]. Mobile learning or "m-learning" according to Crompton [4] is "learning across multiple contexts, through social and content interactions, using personal electric devices" whereas Sharples, Taylor and Vavoula [5] define it as "the processes of coming to know through conversations across multiple contexts amongst people and personal interactive technologies". These quite different explanations are to illustrate the variety of mobile learning definitions, and there are many other. We take the latter and proceed with defining certain well-known principles of what makes learning "mobile".

Mobile learning context is considered free of space and time thus delivering essential skills anytime and everywhere on-demand. This means learning occurs across space to retrieve ideas and resources, and adapting them in another context, and learning across time by reviewing knowledge obtained earlier in a different context. [5] Mobile devices mediate mobile learning since the technology affords the 24/7 opportunities to access the internet powerhouses of learning resources [6]. However, to highlight the intertwined nature of formal and informal learning, mobile learning can occur inside the classroom (or outside) for example participating in a formal lesson on a mobile device [4]. Often misunderstood, mobile learning is not solely about learning with using mobile phones yet it encompasses the mobile nature of learning but in conjunction has a strong underlining for technology offering the mobile possibilities. Sharples, Taylor and Vavoula [5] address the learner being constantly on the move and the reality that people learn increasingly in informal settings that are not necessarily separated from formal education environments thus placing the interest towards the mobility of learning itself. Although mobile learning indeed occurs in formal learning environments, formal conventional learning is typically distinguished as countering the possibilities of space and time having objects such as school buildings, walls, classrooms as well as timetables and schedules [7]. In recent years, steps has been taken to tackle the formal/informal boundaries by introducing flipped learning and Bring Your Own Device (BYOD) movements, just few to mention, which take place in both formal and informal learning environments and implement the basic idea of mobile learning.

Mobile learning might be a challenge for the more conventional ways of teaching and learning but it may also offer new prospects [8]. To mention examples of why mobile learning in language learning context might prove its worth (See 1.2.), we urge the importance of a much broader social-cultural meaning of mobile learning for education. Mobile lifestyle is an integral part of nowadays careers and working life [9], and students are already living in the digital world with daily access to mobile technology thus increasing expectations towards formal education to follow this pattern [10]. To continue, as video games and interactive entertainment are the cultural motor, students are expecting high level of engagement in their learning activities as well [11].

1.2 Mobile-Assisted Language Learning

Deriving from the mobile learning context, Mobile-Assisted Language Learning (MALL) is "the use of mobile technologies in language learning, especially in situations where device portability offers specific advantages" [8]. Briefly, Computer-Assisted Language Learning (CALL) includes similarities with the aforementioned MALL definition and certainly, they share a common ground. However, Kukulska-Hulme and Shield [12] draw a line between CALL and MALL where the latter is distinctive in its use of personal, portable devices enabling new ways of learning, addressing continuity or spontaneity of access and interaction across different contexts.

Mobile-assisted language learning requires trained teachers with advanced understanding of how to implement new technologies into their content teaching. Chinnery [13] addresses that technology or "instructional tools" in language learning context requires deep knowledge of second language learning pedagogy. Therefore, teachers' pedagogical expertise will continue to play a vital role yet it might need to be re-examined to take into account the special characters of mobile learning [1]. Although students are increasingly using different MALL systems (i.e. applications) to master foreign languages, there is still much to be done in in-service teacher training to train sufficient technological and pedagogical content knowledge when it comes to mobile learning context (See [14]).

A study by Viberg and Grönlund [15] found that for second and foreign language learning, mobile devices are highly appreciated among higher education students for they enable individualization, collaboration and authenticity. In addition, integrating mobile devices to language learning environments does not only require instructors' willingness but a sound pedagogical basis as well. A study by Lai and Gu [16] examined university students self-regulated out-of-class language learning reporting that students are strategic users of technology and use technology to regulate various aspects of their language learning thus proposing that students need metacognitive knowledge about technology-enhanced language learning i.e. how to learn a language by using technology. According to Golonka et al., [2], very few empirical studies support the notion that technology-enhanced foreign language learning would lead to improved learning processes, although several studies indicate an increased motivation and enjoyment of learning activities. Another study by Dashestani [17] reported that implementing mobile learning to English as a Foreign Language (EFL) context was perceived highly among students as mobile learning enabled ubiquitous learning. To end with, Burston [18] conducted a meta-analysis for the past twenty years of MALL publications (n = 291) concluding, that studies having statistically reliable measures of learning outcomes (n = 35) 80 percent of the cases were unquestionably positive when implementing a MALL system. However, Burston mentions that the whole MALL research field lacks robust and replicable research settings and suffers from technocentric positioning.

1.3 System Used in the Study

In our study, a specific MALL system was under assessment to examine participants' user expectations and user experiences. The MALL system works both in desktop internet browser and in smart phone environments (i.e. iOS and Android). The MALL system provides 10 different languages to choose from, including Finnish, Swedish and English among other.

The basic logic for the end-user in the MALL system is to listen, read and write, choose correct answers and complete tasks in order to receive points and to proceed further to more advanced levels with the language learning material (See Figure 1).

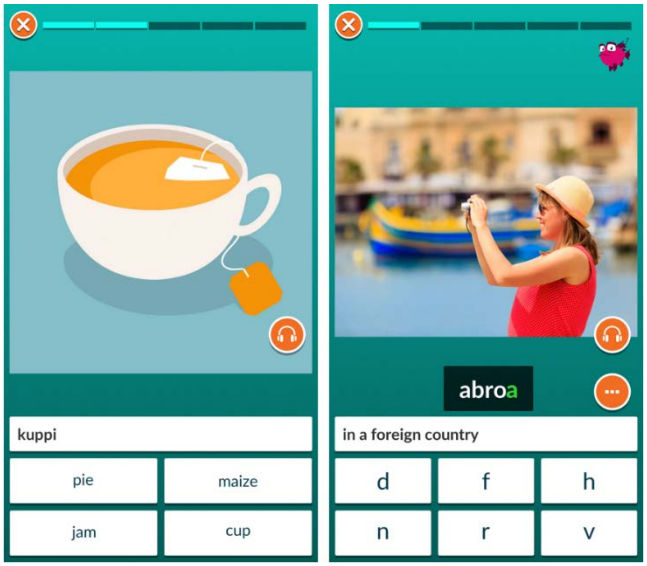

Figure 1. Screenshots of MALL systems` user interface

Among basic language skills modules, there are context specific language modules provided such as vocabulary for travelling. In addition, MALL system provides end-users interactive scoring and to a certain level, feedback based on end-user success. Rather than just a technological system to provide language skills, there are game performances to be found such as scoreboard, avatars and jingly sounds indicating the correct (or incorrect) answer.

The MALL system offers mainly vocabulary training in its different language modules, and offers more tailored FL learning packages for preparing to summative language tests in the formal education system.

2 METHODOLOGY

2.1 SUXES Method

Subjective Metrics with User Experiments (SUXES) is an evaluation method developed by Turunen et al., (2009) to collect subjective user expectations and experiences [19]. The Method originates from the SERVQUAL framework introduced by Zeithaml, Parasuraman and Berry (1990) for measuring service quality metrics [20]. Ever since, SUXES has been used to assess interactive systems measuring subjective user expectations and experiences. For instance, the SUXES has been incorporated to assess user experience of mobile dictation application with speech recognition [21], multimodal media center interface [22], and home media controller using a mobile phone [23]. The evaluation process is straightforward and relies on pre- and post- test setting where user expectations are collected with a questionnaire before the planned use of a particular interactive system, and after the use, the users fill in the same questionnaire capturing their experiences. SUXES involves 11 statements regarding the user experience dimensions in a seven step Likert scale ranging from “Strongly disagree” (1) to “Strongly agree” (7). By comparing pre- and post-test median values and their differences, it is possible to see whether subjective expectations towards the system were met, exceeded or was the system underperforming. In order to elicit user’s perceptions, qualitative material is also being collected to deliver insights. As SUXES is concerned on assessing user experience, we utilized System Usability Scale (SUS) to capture usability properties as well. Applying both SUXES and SUS to assess an interactive system has been used before by Keskinen et al., (2013) [21], and Vuorio, Okkonen & Viteli (2017) [24].

2.2 System Usability Scale (SUS)

Developed by John Brooke (1996), the System Usability Scale (SUS) is a robust ten-item scale to measure subjective usability of a system [25]. In a retrospective published in 2013, Brooke describes objectives of SUS as “To provide us with a measure of people’s subjective perceptions of the usability of a system and, to allow us to do so in the very short time available to us during an evaluation session” [26]. Five-step scale ranging from “Strongly disagree” (1) to “Strongly agree” (5), the SUS scales main analysis is to grade responses in statements leading to an overall sum between 0 - 100 representing the overall usability of a system. The final SUS score is not a percentage of how usable a particular system is, and this has led to misunderstandings with interpretation [26]. To add more clarity and feasibility, Bangor, Kortum & Miller (2008) created adjective ratings that highly correlates with SUS scores [27]. Being a robust scale, it is possible to get reliable results with a sample size of 8 to 12 participants [28].

2.3. Content Analysis

By Neuendorff (2002), content analysis is “summarizing, quantitative analysis of messages that relies on scientific method (including attention to objectivity-intersubjectivity, a priori design, reliability, validity, generalizability, replicability, and hypothesis testing) and is not limited as to the types of variables that may be measured or the context in which the messages are created or presented” [29]. For content itself (open-ended data), we used inductive content analysis to organize single quotations provided by participants. We used open coding to determine a prompt characterization, heading or a notion for a single quotation or a sentence while reviewing the content. After iterative reviews, codes were put to thematic categories where single codes were representing the same dimensions. After categories were determined, we again grouped the codes with similar categories to reduce the number of common nominators [30].Categories reassemble patterns or themes that derive from the data by using analysis [31]. Coding and categorization was performed by using Atlas.ti data analysis software.

2.4 Sample and Procedure

Conducted in 2017, Finnish upper secondary students were recruited from the same school locating in Southern Finland resulting to 157 students willing to participate to our study. Students received pre-test questionnaire in their e-mails before they started to use the MALL system. Out of 157 participants, 60 entered the digital questionnaire platform from which 59 filled the questionnaire successfully. This research did not include interventions of any kind for teachers or students and there were no instructions or obligations for the use.

After the MALL system was available for the participants to use for 11 weeks, participants received the post-test questionnaire in their e-mails. Accessed by 21 participants, 19 filled the identical post-questionnaire properly. To collect a sample with the same participants with complete records from both pre- and post-test questionnaires to use at the SUXES and SUS analysis, we resulted to 15 (n=15) participants from which 10 were female and 5 were male. Ages varied between 16 and 19. In content analysis, we used the complete 19 respondents from post-test sample that included open-ended data.

Table 1. Participants' age and gender distribution.

		Age				
		16	17	18	19	Total
Gender	Female	5	1	2	2	10
	Male	3	2	0	0	5
Total		8	3	2	2	15

Wilcoxon signed-rank test was carried out for SUXES variable results measuring the median differences between two repeated measurements with the same participants (Field, 2013). To match the SUS procedure based on five-point scale, SUS results were coded from seven-point scale to five-point scale as followed: 1=1; 2, 3=2; 4=3; 5, 6=4; 7=5. Recoding is based on validated notion that end of the scales are more precise that points between the extremes and therefore the respondent can be considered somewhat indifferent between 2 and 3 or 5 and 6. Thus those can be recoded to single point in order to make it possible to compare different measures (cf eg.. Krossnick and Pressesr 2010 or Wright and MacRae 2007).

3 RESULTS

3.1 SUXES Results

The null hypothesis is that there is no significant difference ($p < .05$) between pre- and post-test scores. To explore the statistical difference, Wilcoxon signed-rank test was conducted. In addition, to deliver a complete picture of the differences between expectations and experiences, mean (M), median (Mdn) and their differences are also presented. The results are listed in Table 2.

Table 2. SUXES pre- and post-test analysis results.

Variables	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11
	Using the application is fast.	Using the application is pleasant.	Using the application is clear.	Using the application correctly is effortless.	The application functions error-freely.	Using the application is easy to learn.	Using the application is natural.	The application is useful.	I would use the application in the future.	Using the application is fun.	The application looks pleasant.
n	15	15	15	15	15	15	15	15	15	15	15
M (pre)	5,47	5,07	5,27	5,13	5,07	5,53	5,40	5,27	4,93	5,13	5,20
Mdn	6	5	5	5	5	6	6	5	5	5	5
SD	1,407	1,100	1,163	0,990	1,438	1,457	1,242	1,223	1,580	0,915	1,146
M (post)	4,47	3,93	4,40	4,27	4,20	5,20	4,13	4,13	3,93	3,73	3,93
Mdn	5	4	5	5	4	6	5	5	4	4	4
SD	1,807	1,751	1,502	1,668	1,821	1,897	1,807	1,807	1,831	1,870	1,870
M Diff.	-1,00	-1,14	-0,87	-0,86	-0,87	-0,33	-1,27	-1,14	-1,00	-1,40	-1,27
z-score	-1,873	-1,9	-1,796	-1,706	-1,253	-0,423	-1,868	-1,621	-1,675	-2,157	-1,746
Sig p <.05	0,061	0,057	0,072	0,088	0,21	0,672	0,062	0,105	0,094	0,031	0,081

Results of user experience scores did not outperform the scores of user expectation scores. First, mean (M) values of pre-test are higher throughout the variables comparing to post-test values meaning participants had higher expectations than their actual experience was. Wilcoxon signed-rank test was carried to reveal any significances between two samples. Particularly, the system was not as fun as expected resulting in statistically significant value ($p = .031$, $z = -2{,}157$). In addition, marginal significant values were found across the results, respectively. As pre-test results outperform post-test results when examining mean score differences, we proceed comparing the median values of pre- and post-test results and their differences for further implications. Median comparisons of the two samples are presented in Figure 2.

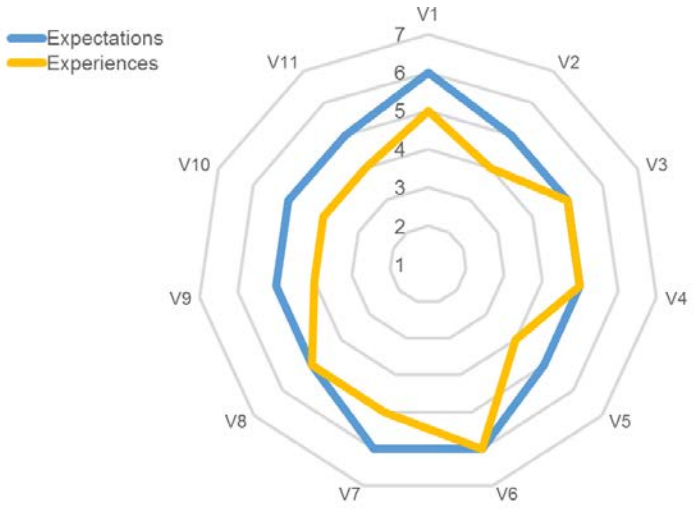

Figure 2. SUXES pre- and post-test median scores in a radar chart.

Several pre-test median scores were higher than post-test scores: fast use, pleasant use, error-free, natural use, future intention of use, is using the system fun, and does the system look pleasant. Variables with no differences between pre- and post-test samples are clear use, is using the application correctly effortless, is it easy to learn, and is the system useful. Concluding, user experience scores were lower or the same when comparing to user expectations median values. As the user experience median results are not strongly indicating system performances are satisfyingly established, the scores are falling to median values of four and five suggesting the performances of the system are somewhere in the marginal in a seven step scale. To proceed elaborating the results and draw a more concrete overall assumptions of this drop between user experience and user expectations, the System Usability Scale results may provide further implications.

3.2 SUS Scores

Proceeding, the pre-test and post-test scores of the system usability scale are presented in Figure 3. Results include the same participants between two samples ($n = 15$) presented earlier. By examining not just the actual SUS score based on experiences but also the expectations score, we may draw a more sophisticated overall view of how user perceptions have changed.

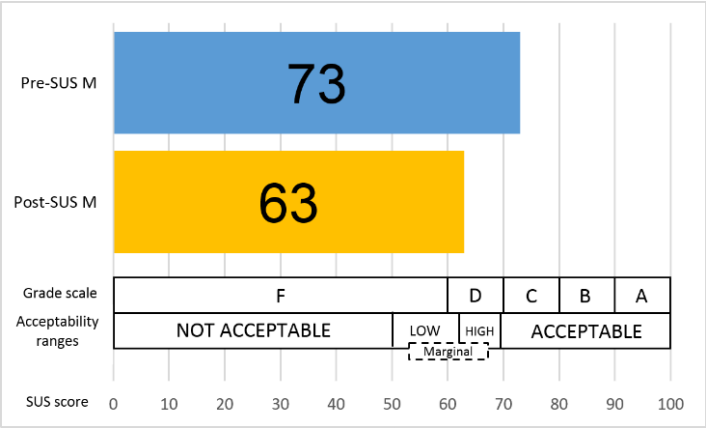

Figure 3. SUS pre- and post-test results with grade scale and acceptability ranges.

User expectations resulted to a mean (M) SUS score of 73, and user experiences to a mean score of 63. Thus, user expectation scores are higher than experience scores as mentioned in aforementioned results. Acceptability ranges indicate that the MALL system was indeed in an acceptable range due to expectation scores, whereas after using the system, scores fell to higher marginal. Grade scales show that scores fell from C to D. Based on SUXES and SUS results: (1) the MALL systems user experience scores are generally lower or the same when comparing to user expectation scores, and (2) overall user experience scores fell in the marginal indicating the system has unknown user experience disputes. To reveal these hidden problems with user experience, a content analysis was carried out to elicit where the problems are locating.

3.3 Results of Content Analysis

A preliminary open-ended question "What are your expectations toward using the application [MALL system]?" was asked from participants. 54 participants (n=54) filled in their expectations from which one answer was deleted as irrelevant as others were taken for analysis. Single quotations and strings were coded and categories were created to determine overall classifications for single codes, and using interpretation, what they might represent about user expectations in using the MALL system (See Table 3). To interpret following results, we remind that participants used the MALL system mostly for acquiring English skills, but also for Swedish and German. In addition, following results are isolated from pre-test questionnaire and due to a smaller post-test sample size, it is ambivalent to make comparisons between the two samples.

Table 3. Content Analysis Categories of User Expectations.

Categories	*n of quotations*	*%*
Vocabulary & Grammar	25	26,0
Learning Experience	22	22,9
Enhance & Support	19	19,8
Performances	19	19,8
Reasoning	6	6,3
Informal Learning	5	5,2
Total	96	100

Six (6) thematic categories were noticed. Vocabulary and Grammar includes 25 quotations, sentences, words and/or phrases. Users expected to learn more words, to learn a wider personal vocabulary and new phrases. Moreover, users wanted to learn words fast and expected to learn grammar as well.

Learning Experience, including 22 quotations refers to delivering a profound holistic experience for the learner. This means users expect to learn meta-skills of language learning i.e. learning better ways to learn a language (learning to learn), to learn more thoroughly, to learn how to speak a specific language more natural, to learn efficiently, and also users expected the system would motivate them to learn languages and it would also be fun.

Enhance and Support includes 19 quotation strings. Users expected the system to support them to learn language skills (in general), and that the user can enhance their skills that need improving or recalling.

Performances include quotations about expectations toward user interface (UI) and user experience (UX) but also to what sort of content users are expecting to see when using the system. First, user expected to use a reliable, clear, easy to use and fast to learn system. They expected a smooth user experience from using the system that is well constructed and interacts with other information platforms as well. In addition, gamified environment and Quizlet like interface were desired.

Reasoning includes string quotations from users reporting reasons why they wanted to use the system. Users reported they wanted to learn, learn in better ways, they wanted to perform better in school (as in grades), and they wanted to prepare for their matriculation examination.

Finally, Informal Learning includes five quotations about user expectations that refer to using the system for learning in informal surroundings (e.g. in a bus) and as a complementary for learning from the book thus extending the learning beyond formal language learning environments.

After using the system, 19 participants (n = 19) filled in the post-test questionnaire having the possibility to deliver more exact information about their experiences. 24 separate quotations were included to analysis and labeled with codes representing a broader nominator. This led to 64 (n =64) individual quotations that were placed under categories in Table 4.

Table 4. Content Analysis Categories of User Experiences.

Categories	*n of quotations*	*%*
Complications	21	32,8
Unfulfilled Expectations	12	18,8
Ease of Use	10	15,6
Positive Experience	8	12,5
Reflection	7	10,9
Simple	6	9,4
Total	64	100

First, Complications category includes 21 quotations relating to users experiencing problems with the system interface, did not fully understood navigation of the system and experienced the system as dull. Some users did not understand how the scores are generating in the system whereas other users did not see the benefit for their learning.

The "word bubbles" moved slowly, all of the answers did not appear immediately and poor visualization made it even worse (…) I feel there´s no benefit for me especially when I can´t choose to learn a specific vocabulary, and when you can use Quizlet instead. (P9)

Unfulfilled expectations are suggestions or critical assessments about what users were expecting from the systems technical and content performances. Total of 12 quotations includes what users expected or desired from the system. Users desired more theory (i.e. grammar) and variety.

Applications language learning methods were too narrow, there should´ve been more variety. Also, theory was not taught and as new words emerged [to the screen] it was hard to remember them. (P2)

Ease of Use contains 10 quotations relating to systems straightforwardness, fast use, free of effort and easy to learn dimensions.

I think using the application is easy for it is very clear and one can learn how to use it very quickly. (P18)

Positive Experience category include users reporting overall positive experiences albeit we do not know what exactly were the foundations for positive experiences. Category includes relatively prompt quotations such as “Nice”, “Handy” and “I liked it”.

Reflection, including seven (7) quotations include participants’ self-reflection and meta-level thinking about the possible outcomes of using the system. Participants report they would probably learn if they would use the system constantly or that they have not yet learned new skills and the system would probably benefit those with only basic language skills.

The final category labeled as Simple with six quotations contains descriptions about systems simple nature such as “It is rather simple”, “There was not too many functions” and “For a basic internet user, it was enough”. It seems these answers are two-folded as participants experience simple as both positive and negative outcome of the system.

Indeed, using the MALL system generated a variety of expectations and experiences, which were analyzed using both quantitative and qualitative methods. Next, more condensed notes and implications are provided.

4 CONCLUSIONS

A Mobile Assisted Language Learning System was used among Finnish upper secondary students to study user expectations and experiences and their differences by using SUXES and SUS methodologies as well as content analysis.

As results argue, user experience scores did not outperform the scores of user expectation in any of the variables when examining the mean differences. These differences between two repeated samples is relatively strong evidence that there are disputes with the systems user experience albeit that Wilcoxon signed-rank test showed only one variable resulting to a significant p-value. In addition, also other variable results were in the grasp of falling under .05. Recognizing the limitations with interpreting p-values (See Cohen, 1994; Lambdin, 2012) [32][33], there is a strong pattern indicating, that (1) user expectation scores were higher across the variable results and therefore we may assume users expectations were not met, and (2) differences between two samples are relatively high but not quite significant, respectively. Although witnessing the difference between the two samples, more conjunct elaboration was needed to deliver more incisive results of where these problems are laying and why end-users responded the way they did.

As SUXES method is particularly interested about changes in the median values, we proceeded to compare each variables median score in Figure 1. Seven (7) variables failed to meet the end-user expectations with a negative

median score drop of one (1): fast, pleasant use, error-free, natural, future use, fun, and pleasant look. Four (4) variables with no difference between the two samples were clear, effortless, easy to learn, and useful. Variables with a negative drop seem to represent more of the participants' personality, subjective feelings (i.e. what feels and looks good) and hedonistic needs (fun) where we move beyond technical/functional properties of the system whereas variables with no different between the samples represent more of the learning side, straightforwardness and usefulness of the system. It seems end-users were perhaps expecting more hedonistic and natural experience, cool looking interface, and more enjoyment with using the system instead of focusing solely on the learning and usefulness side.

System Usability Scale was added to capture usability aspects and to deliver more information about the systems overall performance. Grade scale and acceptability range were included to add more interpretation for the results. Pre-test (expectation) resulted to a SUS score of 73 (grade C) and is inside the acceptable range. According to Bangor et al (2008), at least 70 is the acceptable limit for a product although better products score high 70s to upper 80s whereas outstanding and superior products score above 90. Grade C can be regarded as university grade in countries where letter grades are commonly used (e.g. UK & US) and so C stands in the middle of 5-step grade scale (A, B, C, D, F).

In the post-test (experiences), we witness a drop from the pre-test score of 73 to post-test score of 63 (grade D). There are several notions what this means. Anything less than 70 have usability issues that are cause for concern in the overall usability of the product [27]. Moreover, the drop from C to D confirms that indeed the system falls to slippery area where it no longer reaches the average or mean score. Finally, as we look at the acceptability rating, the system falls barely to higher marginal. Low and high marginal separate the "Unacceptable" and "Acceptable" ranges and was created to separate two ranges where anything below 50 would result to unacceptable and above 70 acceptable. First quartile in Bangor et al., (2008) studies of SUS scoring empirical evaluation (n=206) ends at score 62 meaning that in their empirical review of SUS studies one fourth of mean SUS scores reach the first quartile.

Concluding, the system fell from pre-test score "Acceptable" to a marginal status where the system is no longer in an acceptable range and there are unknown disputes with the systems usability, respectively. These findings are aligned with previous implications with the SUXES results where post-test scores dropped notably compared to pre-test scores, and the median scores fell from scores 6 and 5 to 5 and 4 thus supporting the evidence that the system is indeed performing somewhere in the middle of acceptable and unacceptable ranges.

To deliver more evidence where exactly were the possible disputes with the systems user experience and usability aspects, the content analysis was conducted for participants' open-ended inputs. First, participants expected to learn a wide array of vocabulary (both new and recalling the old) and grammar. According to Schmitt (2008), second language learners need to be exposed to a large directory of vocabulary and it is a complex and gradual process [34]. It seems the demands for the system are high in both theoretical point (how people learn a new language) as from the end-user aspects (I expect state of the art experience). The expectations to learn new words and grammar were not fully met since participants reported they did not find meaningful contents that would suit their specific needs nor did they find more theoretical (e.g. grammar) learning content.

Participants also expected to receive sophisticated learning experience such as how to learn a language efficiently and fast, and they expected the learning to be motivational among other aspects. It is plausible that the MALL system could suffer from lack of engaging user interface, usability malfunctions or logic errors and personalized user experience. It is likely that the system has disputes with its navigation structure and task interface since participants reported several complications with finding preferences or they struggled with filling words correctly to blank boxes. More UX and usability studies are needed in order to determine where these end-user problems are in the system using more specific methods (e.g. focus groups).

Unfulfilled expectations were derived from the content analysis data as something participants felt they missed or desired from the system. Gamified experience among other were reported but also, again, users felt the system did not deliver high-end content for their needs. Whereas there is a distinctive group of participants yearning for more, we also witnessed a more satisfied group of users experiencing the system as delivering what they were expecting. This result with finding two strong voices from the data may suggest that the system is suitable and fit for users with basic level language skills that are satisfied with a more common content. On the other hand we have users with high demands who struggled with finding more uncommon words and content that would challenge their language knowledge thus enabling learning. What was felt as “simple” or “easiness” was negative for other users whereas for others it was felt a good character of the system. Users also delivered answers where they self-reflected that the system would probably suit for those with basic language skills.

Finally, assessing the MALL system differences between user expectations and experiences resulted to findings where the results left the system to a marginal where some implications can be provided for further studies and development of mobile assisted language learning systems. First, the context is challenging as the theory of learning a language combined with relatively high demanding end-users result in a mixture of multiple struggle areas. In our study, the system under assessment failed to meet the end-user expectations in multiple levels thus falling to a marginal range where the system needs serious reshaping of its usability structures. To what extent the content challenges advanced language learners is also a worthy taking. Moreover, the system seems to suffer from lack of personalized learning experience and engagement of the learner. Engagement could be triggered by applying more game elements, fixes, revises of the usability aspects of the whole system, and update the user interface elements to further attract end-users.

Limitations apply as we studied the system solely depending on two samples before and after using the system with no instructions nor spectating and therefore more studies are to be made to provide concrete solutions for the product development. Moreover, it is desired to conduct more studies to what extent these systems deliver language learning in measurable metrics and to what degree these systems reach basic and advanced learners. Also further analysis on UX would require auxiliary data on grades on foreign languages, mobile device user profile, background with digital games and serious games, and maybe even teacher assessment on certain participant.

REFERENCES

[1]

KUKULSKA-HULME, A. 2009. Will mobile learning change language learning? In ReCALL (Vol. 21, pp. 157–165). Cambridge University Press. https://doi.org/10.1017/S0958344009000202

[2]

GOLONKA, E. M., BOWLES, A. R., FRANK, V. M., RICHARDSON, D. L., & FREYNIK, S. 2013. Technologies for Foreign Language Learning: A Review of Technology Types and Their Effectiveness. Computer Assisted Language Learning, 27(1), 70–105. https://doi.org/10.1080/09588221.2012.700315org/10.1080/09588221.2012.700315

[3]

BERGE, Z. L., & MUILENBURG, L. 2013. Handbook of mobile education. Retrieved from https://ebookcentral.proquest.com

[4]

CROMPTON, H. 2013. A Historical Overview of M-Learning: Toward Learner-Centered Education. On Berge Zane & Muilenburg Lin (Ed.) Handbook of mobile education. Retrieved from https://ebookcentral.proquest.com. 80-107

[5]

SHARPLES, M., TAYLOR, J., & VAVOULA, G. 2005. Towards a theory of mobile learning. In Proceedings of mLearn (Vol. 1, No. 1, pp. 1-9).

[6]

TERRAS, M. M., & RAMSAY, J. 2012. The five central psychological challenges facing effective mobile learning. British Journal of Educational Technology, 43(5), 820–832. https://doi.org/10.1111/j.1467-8535.2012.01362.x

[7]

KEARNEY, M., SCHUCK, S., BURDEN, K., & AUBUSSON, P. 2012. Viewing mobile learning from a pedagogical perspective. Research in Learning Technology, 20(1). https://doi.org/10.3402/rlt.v20i0/14406

[8]

KUKULSKA-HULME, A. 2013. Mobile-Assisted Language Learning. The Encyclopedia of Applied Linguistics, 1(1). https://doi.org/10.1002/9781405198431.wbeal0768

[9]

BECKMANN, E. A. 2010. Learners on the move: Mobile modalities in development studies. Distance Education, 31(2), 159–173. https://doi.org/10.1080/01587919.2010.498081

[10]

KUKULSKA-HULME, A., & DE LOS ARCOS, B. Researching emergent practice among mobile language learners. On mLearn 2011, Beijing, China.

[11]

MARTIN, F., & ERTZBERGER, J. 2013. Here and now mobile learning: An experimental study on the use of mobile technology. Computers & Education, 68, 76–85. https://doi.org/10.1016/j.compedu.2013.04.021

[12]

KUKULSKA-HULME, A., & SHIELD, L. 2008. An overview of mobile assisted language learning: From content delivery to supported collaboration and interaction. ReCALL, 20(03), 271–289. https://doi.org/10.1017/S0958344008000335

[13]

CHINNERY, G. M. 2006. Going to the MALL : Mobile Assisted Language Learning. Language Learning & Technology, 10(1), 9–16. https://doi.org/ISSN 1094-3501

[14]

BARAN, E. 2014. A Review of Research on Mobile Learning in Teacher Education. Educational Technology & Society, 17(4), 17–32. https://doi.org/10.1007/s10639-011-9182-8

[15]

VIBERG, O., & GRÖNLUND, Å. 2013. Cross-cultural analysis of users' attitudes toward the use of mobile devices in second and foreign language learning in higher education: A case from Sweden and China. Computers & Education, 69, 169–180. https://doi.org/10.1016/j.compedu.2013.07.014

[16]

LAI, C., & GU, M. 2011. Computer Assisted Language Learning Self-regulated out-of-class language learning with technology. Computer Assisted Language Learning, 24(4), 317–335. https://doi.org/10.1080/09588221.2011.568417

[17]

DASHTESTANI, R. 2016. Moving bravely towards mobile learning: Iranian students' use of mobile devices for learning English as a foreign language. Computer Assisted Language Learning, 29(4), 815–832. https://doi.org/10.1080/09588221.2015.1069360

[18]

BURSTON, J. 2015. Twenty years of MALL project implementation: A meta-analysis of learning outcomes. ReCALL, 27(1), 4–20. https://doi.org/10.1017/S0958344014000159

[19]

TURUNEN, M., HAKULINEN, J., MELTO, A. & HEIMONEN, T. 2009. SUXES - User Experience Evaluation Method for Spoken and Multimodal Interaction. Interspeech 2009: 10Th Annual Conference of the International Speech Communication Association 2009, Vol 1-5, (January), 2551–2554.

[20]

ZEITHAML, V. A., PARASURAMAN, A., & BERRY, L. L. 1990. Delivering quality service: Balancing customer perceptions and expectations. Simon and Schuster.

[21]

KESKINEN, T., MELTO, A., HAKULINEN, J., TURUNEN, M., SAARINEN, S., PALLOS, T., … SALANTERÄ, S. 2013. Mobile dictation for healthcare professionals. In Proceedings of the 12th International Conference on Mobile and Ubiquitous Multimedia - MUM '13 (pp. 1–4). New York, New York, USA: ACM Press. https://doi.org/10.1145/2541831.2541880

[22]

TURUNEN, M., KALLINEN, A., SÁNCHEZ, I., RIEKKI, J., HELLA, J., OLSSON, T., … RAISAMO, R. 2009. Multimodal Interaction with Speech and Physical Touch Interface in a Media Center Application. ACE '09 Proceedings of the International Conference on Advances in Computer Enterntainment Technology, 19–26. https://doi.org/10.1145/1690388.1690392

[23]

TURUNEN, M., MELTO, A., HELLA, J., HEIMONEN, T., HAKULINEN, J., MÄKINEN, E., … SORONEN, H. 2009. User expectations and user experience with different modalities in a mobile phone controlled home entertainment system. In Proceedings of the 11th International Conference on Human-Computer Interaction with

Mobile Devices and Services - MobileHCI ’09 (p. 1). New York, New York, USA: ACM Press. https://doi.org/10.1145/1613858.1613898

[24]

VUORIO, J., OKKONEN, J., & VITELI, J. 2017. Enhancing user value of educational technology by three layer assessment. Proceedings of the 21st International Academic Mindtrek Conference on - AcademicMindtrek ’17, 220–226. https://doi.org/10.1145/3131085.3131105

[25]

BROOKE, J. 1996. SUS-A quick and dirty usability scale. Usability evaluation in industry, 189(194), 4-7.

[26]

BROOKE, J. 2013. SUS : A Retrospective. Journal of Usability Studies, 8(2), 29–40. https://doi.org/10.1074/jbc.R115.675280

[27]

BANGOR, A., KORTUM, P. T., & MILLER, J. T. 2008. An empirical evaluation of the system usability scale. International Journal of Human-Computer Interaction, 24(6), 574–594. https://doi.org/10.1080/10447310802205776

[28]

TULLIS, T. S., & STETSON, J. N. 2004. A Comparison of Questionnaires for Assessing Website Usability. Usability Professional Association Conference, 1–12. Retrieved from https://www.researchgate.net/publication/228609327

[29]

NEUENDORF, K. A. 2002. The content analysis guidebook. Sage Publications. Retrieved from https://tamcat.finna.fi/Record/tamcat.343883

[30]

ELO, S., & KYNGÄS, H. 2008. The qualitative content analysis process. Journal of Advanced Nursing, 62(1), 107–115. https://doi.org/10.1111/j.1365-2648.2007.04569.x

[31]

HSIEH, H.-F., & SHANNON, S. E. 2005. Three Approaches to Qualitative Content Analysis. Qualitative Health Research, 15(9), 1277–1288. https://doi.org/10.1177/1049732305276687

[32] COHEN, J. 1994. The earth is round ($p < .05$). American Psychologist, 49(12), 997–1003. https://doi.org/10.1037/0003-066X.49.12.997

[33] LAMBDIN, C. 2012. Significance tests as sorcery: Science is empirical—significance tests are not. Theory & Psychology, 22(1), 67–90. https://doi.org/10.1177/0959354311429854

[34] SCHMITT, N. 2008. Review article: Instructed second language vocabulary learning. Language Teaching Research. https://doi.org/10.1177/1362168808089921