User experience improvement of japanese language mobile learning application through mental model and A/B testing

Komang Candra Brata, Adam Hendra Brata
Mobile, Game, and Media Research Group, Department of Computer Science, Universitas Brawijaya, Indonesia

ABSTRACT

Advances in smartphone technology have led to the strong emergence of mobile learning (m-learning) on the market to support foreign language learning purposes, especially for the Japanese language. No matter what kind of m-learning application, their goal should help learners to learn the Japanese language independently. However, popular Japanese m-learning applications only accommodate on enhancing reading, vocabulary and writing ability so that user experience issues are still prevalent and may affect the learning outcome. In the context of user experience, usability is one of the essential factors in mobile application development to determine the level of the application’s user experience. In this paper, we advocate for a user experience improvement by using the mental model and A/B testing. The mental model is used to reflect the user’s inner thinking mode. A comparative approach was used to investigate the performance of 20 high-grade students with homogenous backgrounds and coursework. User experience level was measured based on the usability approach on pragmatic quality and hedonic quality like effectiveness (success rate of task completion), efficiency (task completion time) and satisfaction. The results then compared with an existing Japanese m-learning to gather the insight of improvement of our proposed method. Experimental results show that both m-learning versions proved can enhance learner performance in pragmatic attributes. Nevertheless, the study also reveals that an m-learning that employs the conversational mental model in the learning process is more valued by participants in hedonic qualities. Mean that the proposed m-learning which is developed with the mental model consideration and designed using A/B testing is able to provide conversational learning experience intuitively.

Keywords: Mental model Mobile learning A/B testing User experience Usability

1. INTRODUCTION

The high demand in Japanese learning and the limited learning media from Indonesian has led to the strong emergence of m-learning on the market. M-learning systems are being used to improve student capability to learn the Japanese language independently, in addition to facilitating the learning process at any time [1-3]. Researchers are now proposing various adaptive m-learnings which provide different exercise types or learning approaches to the learner based on the current skill level of an individual to assist learners while studying Japanese vocabulary [4]. Recent insight reveals that user experience issues are still prevalent in some of the popular Japanese m-learnings which are available on the market such as “Write it” [5] and “Obenkyo” [6]. The common issues are the lack of conversational style, they only focus on enhancing
reading, vocabulary and writing skills so this condition may affect the learning outcome. Usability dimension is a critical topic for mobile application user experience since it is necessary to develop an application that is not to be hard to use and this has been identified as one of the factors that can determine the quality of an app’s user experience. Usability specifically indicates the degree of user performance whiles using an application to attain a specific goal [7]. In terms of user experience, usability dimension can be categorized into pragmatic and hedonic parameters.

The process of understanding how the system works requires individuals to construct a mental model of the system in their minds. The mental model acquisition is at the heart of meaningful learning. Meaningful learning will be achieved when learners can engage in active cognitive processing. Mayer gives a model for learning in multimedia learning [8]. The main issues in advanced m-learning of Japanese are how the words are pronounced and how to evaluate the correctness of pronunciation. Recent pieces of literature show that speaking is a key indicator of effective communication skills in foreign language learning [9-11]. This puts forward for consideration that there is still a chance to leverage the user experience level that offered by existing m-learning.

Speaking and listening might bring a great impact on the Japanese language learning process. The goal of this paper is to present an approach for enabling improvement in the user experience of m-learning by implement the speaking and listening element as part of the m-learning mental model and utilize A/B testing method to obtain a better user experience. A/B testing is a standard method for evaluating user engagement or satisfaction from a new service, feature, or product, this is also called bucket testing, split testing, or supervised experiment. The objective of this study is to gather mental model form user’s perspective and work with a/b testing in order to present the new concept of m-learning that have better user experience compared with common existing m-learning.

2. RESEARCH METHOD

In this paper, a diary study was used to explore and reflect on the learner learning strategies to learn fundamental Japanese materials. Advances in multimedia contents can extend the experience of m-learning apps. However, it is not a trivial task to successfully deploy various multimedia content in mobile learning environments. This is because the provision of different learning materials with only one smartphone screen requires an effective approach. The steps for constructing mental model for proposed intuitive m-learning, user experience design strategies and system design are discussed below.

2.1. Recording mental models

Mental models are individual inner representations of the external reality that used by individuals to communicate with the environment around them. The concept of the mental model was first proposed by Kenneth Craik, the Scottish psychologist, in 1943 [12]. But what really made this concept popular was Johnson-Laird's with his interpretation that the mental model is a simpler world in people's minds when they understood the objective world [13]. Peter Senge believes that mental model is a series of assumptions, ideas, images or impressions that ingrained in the mind, affecting how we understand the world, and how we take action on [14]. Donald Norman divided the mental model into three interaction related models: design model, user model and system model [15]. To put it simply, a mental model is a unique way of understanding which formed in the process of interacting with the world, and at the same time applies this way of understanding to the next things. It is built by people on their distinctive lives, points of view and world understandings [16]. We are examining personal responses on their understanding of the subject matter of Japanese learning. There can be a different number of participants actually engaged in the field study and testing. Faulkner stipulates that 20 respondents will reveal as many usability problems as more people will do and it is sufficient to offer the usability experimentation a reasonable value-to-cost balance [17].

Based on Faulkner's insight, 20 respondents were involved in this study. They knew m-learning applications and they were passionate about learning Japanese. They are also familiar with reading basic hiragana characters and have basic knowledge about Japanese language. They were categorized by students and teachers. While practicing Japanese using the existing m-learning application, their task achievement was monitored and observed. Based on the gathered information from respondents, we create a user journey map to collect user needs swiftly. The User Journey Map gives a traditional persona a third dimension attribute by simply focusing on a diagram of a user's needs and the product. Recent studies indicate that this strategy is an efficient way to quickly collect user stories to develop an intuitive application [18-19]. We develop a user story depending on user journey perceptions. The user story records the system requirements, the expectations of the system and (optionally) why they are urgently needed. User Story should only capture important requirements of the system. To generate a successful user story, we adopting INVEST standards (Independent, Negotiable, Valuable, Estimable, Small, and Testable) in order to get reliable qualitative
metrics. The simplified format that was used in this study is: “As a <type of user>, I want <goal>, so that <some reason>” [20]. Some ideas and obstacles from ordinary individuals can be seen in Table 1 when they are using m-learning applications for Japanese language exercise. Based on these insights, the current implementation of m-learning lacks an integration of multimedia components and interactive environment to independently improve speaking and listening skills as a portion of the learner’s mental model.

2.2. Constructing m-learning mental models

Mental models are people’s perception of how something should work, based on their past experiences. The mental model is an important concept which is can be taken into account by learning experience designers during the process of requirement analysis and design. Having a deficient mental model may indicate a lack of awareness of the usability risks surrounding the learning activities [21]. From literatures on education, it is clear that dialogue based on the triadic model (question–answer–evaluation) is still widely used and become the mental model in educational practice [22-24].

The mental model can be seen as a way to mine a user needs. Thse user story result in recording mental model phase shows that lack of multimedia contents and interactions are the biggest drawback of the existing m-learning. Mayer’s suggested some multimedia principles in interactive mobile learning implementation, Spatial Contiguity Principle, Generative learning principle, Personalization Principle, and Modality Principle. The Generational Principle indicates that students learn more effective with words and images rather than only with words. This principle allows students to visualize concepts and connect them using the associated illustrations. The Principle of Spatial Contiguity recommends that messages must be presented when the picture is located next to words. Modality Principle obviously shows that when words and images are spoken instead of printed, individuals learn better from the contents. Besides, the Personalization Principle suggests that when words and sentences are conversationally conveyed, like speaking and listening, individuals learn better from a multimedia lesson compared to a formal text style [25-26].

Ideally when processing, the different types of multimedia information (i.e., words, pictures, and sound) are integrated to form a stronger representation than any one by itself. Figure 1 depicts the high level of new information processing in human memory. Once a mental model is built, it is extracted from memories and transmitted to a new situation scenario through analogical reasoning. Analogies are mappings of one mental image to another. If a problem arises, a mental model influences what a person focuses on, what actions and strategies a person takes to solve the problem [27].

2.3. User experience design with A/B testing

In the worlds of application’s user interface design, the user will call an application is intuitive when the conceptual model of the application is a very resemblance to the user’s mental model of how a system should works [28]. This rule also works in m-learning, learners will learn more effectively when their mental model of conversation match with conceptual model of an m-learning. Otherwise, a mismatch may cause frustration and confusion that detracts from the experience of learning. A/B testing was used to determine the user interface and gesture control that appropriate for users concerning provide excellent user experience in proposed m-learning. A/B testing is a standard and commonly used framework for evaluating new concepts and making decisions based on result data. In the application development domain, major platforms conduct tests on a set of customers and evaluate the result to obtain an innovative feature [29-30]. A/B testing recommends dividing the user universe into two or more groups in which each group will use different versions of an application and compare the effect of each version. This approach inspects the overall effectiveness of an application by providing enough insight data to enable better design options and prevent
common pitfalls in user interface and experience designs. The simple experimental setup was conducted to evaluate a factor with two versions of the application, we call it a control (version A) and a treatment (version B). Figure 2 shows the high-level structure of an A/B experiment. In practice, a split number of users by 50% is recommended to provide the experiment with the maximum statistical power [31].

![Diagram of multimedia learning](image1)

**Figure 1.** An information processing model of multimedia learning

![Diagram of A/B testing experiment](image2)

**Figure 2.** The structure of an A/B testing experiment processes

There are two approach when conducting A/B testing. “Go big” approach, the main objective is obtaining large sizes of sample for a small number of experiments to ensure that it can capture even small benefits of a new idea or an intervention of policy. In contrast, in early phase product development “Go lean” approach create many relatively small experiments to explore any innovation insight without outstanding success. The idea is to experiment with many ideas quickly and cheaply, abandon or pivot ideas that don't work and expand ideas that work [32]. In this study, the Go lean approach is very relevant to the experimental process.
The common mental model for learning process can be divided into 2 phases, learning phase and scoring phase. Based on Figure 2, the workflows of testing are included set the goal, develop different product variation, test the products then evaluate the result. In this testing, to get information about which UI flows that have better user experience, two different user interfaces with different gestures in each m-learning version were provided to 10 participants. The scoring phase is one of the crucial requirements of m-learning. In this phase, the combination of multimedia principles needs to be applied to make sure the user can easily understand the question and user interface control to answer. The control version is using simple list-view page navigation to presents all the quiz contents inside the app see in Figure 3 (a). The user also can choose the level of question then the app provide user with each questions in chosen level. While in the treatment version shows in Figure 3 (b) single page design was used to display each question with a title header which is inform the user about question level and category, the sliding page navigation can be used to change the next or previous course object. Both versions of the app will check the correctness of user pronunciation then show the result immediately. Next, the participants were asked to do several common learning task scenarios with both version of proposed m-learning prototype. Figure 3 shows the A/B testing of both version of the proposed learning in learning phase. The control version implemented the list-view user interface which have two buttons in each list item content in order to create the natural conversational style in this m-learning see Figure 4 (a). The treatment version is slightly different in content navigation, instead of display all the contents in list view, it use one page to display each content and user can change content with swipe left or right shows in Figure 4b. Not all features can be A/B tested in this study due to limitations on mobile apps and infrastructure constrain. Detail implementation design and prototype framework can be seen in [33].

![Figure 3. A sample of user interactions of learning phase that compared in the a/b test](image)

To determine the output of A/B test, the same 20 participants was split in two groups. Each group try one version of the proposed app. The evaluation parameters that used in this testing were efficiency and effectiveness. Effectiveness is counted by the number of success in a task, while efficiency is counted by the completion time in particular task. Table 2 shows the A/B test experiment results. Although both version of the apps could easily understand by users with 100% success rate, considering the total time for performing each phase, participants completed each phase faster with control version, for Learning phase. Scoring phase reveals the same result that control version also have better time. The result proves that presenting all course contents in list-view still familiar with the user daily usage. Thus, to provide better user experience, the user interface of proposed m-learning adopt the simple list-view layout to present the course and quiz materials as in control version A.

![Table 2. A/B testing results](image)

| Phase       | Control Version A | Treatment Version B |
|-------------|-------------------|---------------------|
|             | Time (minutes)    | Success Rate (%)    | Time (minutes)    | Success Rate (%) |
| Learning    | 2.90 / 1.0        | 100%                | 6.45 / 1.37       | 100%              |
| Scoring     | 4.91 / 0.86       | 100%                | 7.2 / 0.77        | 100%              |

*User experience improvement of japanese language mobile learning application ... (Komang Candra Brata)*
2.4. User experience comparison

The quality in use of interactive products has become such an important criterion. In the context of user experience, usability evaluation is an essential task in mobile application development. Performing usability inspections during the application development process can bring several benefits such as increasing the quality in use and of the software product before its release [34]. We compare the usability level of our prototype which was developed with mental model and A/B testing with existing m-learning apps. The same 20 participants as in elicitation phase were involved and they were asked to do several task scenarios as described in Table 3 with both of the apps, existing common Japanese m-learning and proposed prototype. While they were performing the evaluation tasks with proposed m-learning, their performance on the tasks was observed and any problems occurred were noted and cultivated.

| Task | Task Description |
|------|------------------|
| 1    | Start the application and explore what they can do inside the application within 15 minutes interaction! |
| 2    | Open the animal name course section and then choose one animal name from learning material, then learn and engage with the chosen materials for 5 minutes! |
| 3    | See the correct pronunciation and review the learning result from application! |
| 4    | Start the quiz then get score to evaluate the proficiency of learning materials. Complete the quiz. |

Experimental conditions in Figure 4 (a) and Figure 4 (b) are showing the same task completion in two different application presentations. Degree of user experience is defined as the value of each pragmatic and hedonic parameter. In software domain, Attractiveness scale is an effective way to record the user’s general impression. Attractiveness is a valence dimension (psychology and emotional reaction on an unspoiled acceptance/rejection response). Attractiveness scale consists of pragmatic and hedonic quality. Pragmatic quality aspects are divided by Perspicuity, Efficiency, and Dependability in which describe interaction qualities that relate to the tasks or goals that the user aims to achieve when using an app. Hedonic quality aspects such as Stimulation and Novelty are not related to tasks and goals, but it can describe related aspects to joy and satisfaction while using an app [35]. Both versions of m-learning are tested to the participant in sequential order (one after the other) and each participant has to fill out a questionnaire concerning user experience. In that circumstance, the number of questionnaire items must be kept to a minimum as possible to make sure the participant not stressed and keep the quality of answers [36]. Based on that fact, the short version of the User Experience Questionnaire (UEQ-S) was used for gathering participants' impressions about each version which is this questionnaire can focus on the only aspects of pragmatic and hedonic. Each item of the UEQ-S consists of a pair of terms with opposite meanings as shown in Figure 5. The first four items in UEQ-S measure the pragmatic aspect and the last four items represent the hedonic quality. Each item has a 7-point Likert scale rate. The answers of the participant are then scaled from -3 (fully agree with the negative term) to +3 (fully agree with the positive term). All items in question form are presented in the same polarity. The left side reflects the negative term and the right side the positive term.

![Figure 4. (a) Screenshots of the common m-learning and (b) the proposed m-learning](image)

![Figure 5. The questionnaire items of UEQ-S](image)
3. RESULTS AND ANALYSIS

Direct comparison was conducted to explicitly compare the two evaluated application experiences in term of pragmatic and hedonic indicators. The UEQ-S results summary of the means and standard deviations with the means comparison of the participants’ user experience factors between typical existing m-learning and proposed m-learning environments are presented in Figure 6 (rated from −3 (lowest) to 3 (highest)). Data variability has become a prevalent issue as UEQ-S has extended to a structured testing. The small value of standard deviation (SD) in each parameter of the UEQ-S questionnaire gives information that the participants have similar experience while using both of the m-learning application. As already mentioned, the evaluation result can determine whether an app is satisfactory for learning task and whether it has better user experience scales. Although both of the apps could help users to learn effectively with similar degree of pragmatic quality (perspicuity, efficiency, and dependability). The pragmatic level gap is quite small which common m-learning presentation have better value (Mean=2.23, SD = 0.36) than the proposed m-learning (Mean=2.15, SD = 0.31). Even though common existing m-learning has a slightly better score in pragmatic quality, it is noteworthy that there were significant deviations between the two interfaces concerning the hedonic quality. The proposed m-learning is more valued by respondents in hedonic quality (Mean=1.45, SD = 0.43) rather than existing app (Mean= -0.44, SD = 0.39). The most interesting result of our evaluation is the differences in overall quality scale which is resulted from the intuitive presentation of proposed m-learning which is working with the mental model and a/b testing. The combination of conversational style of learning with speaking and listening in multimedia content can provide better value in overall UEQ-S score.

![Figure 6](image)

Figure 6. A sample of question item in Hiragana quiz

4. CONCLUSION

In this paper, we explore the learning process mental model and also employ A/B testing to propose the improvement of user experience for Japanese language m-learning. Users with a better mental model perform better in A/B testing scenario tasks. The level of user experience of proposed m-learning then compared with typical existing m-learning with pragmatic and hedonic parameters. In general, participants succeeded in using both version of m-learnings and could complete all learning task scenarios. We conclude that, both versions tied with respect to pragmatic qualities Nevertheless, participants highlighted the hedonic element and the natural interaction of user experience in proposed m-learning are promising for respondents. The combination of speaking combined with listening exercise in multimedia information can provide a better pragmatic and hedonic experience. The findings prove that working with the mental model and A/B testing can improve the user experience, but it should be interpreted with caution to judge that it also could increase the learning outcome due to the learning outcome not evaluated yet during this study.

User experience improvement of japanese language mobile learning application ... (Komang Candra Brata)
REFERENCES

[1] Wang, Minjuan, et al., “The Impact of Mobile Learning on Students’ Learning Behaviours and Performance: Report from a Large Blended Classroom,” British Journal of Educational Technology, 40(4), pp. 673-695, 2009.

[2] Pereira, Orlando RE, and Joel JPC Rodrigues, “Survey and Analysis of Current Mobile Learning Applications and Technologies,” ACM Computing Surveys (CSUR), 46(2), pp. 27, 2013.

[3] Wu, Wen-Hsiung, et al., “Review of Trends from Mobile Learning Studies: A Meta-Analysis,” Computers & Education, 59(2), pp. 817-827, 2012.

[4] Nagella, Uday Bhaskar, and P. Govindarajulu, “Adaptive Approaches to Context Aware Mobile Learning Applications,” International Journal of Computer Science and Security, 2(2), pp. 15-27, 2008.

[5] Jernung, “Write It! Japanese,” in Google playstore, [Online]. Available: https://play.google.com/store/apps/details?id=com.jernung.writeit.jp. [Accessed 2017, Sept 28].

[6] Atomusk, “Obenkyo,” in Google playstore, [Online]. Available: https://play.google.com/store/apps/details?id=com.Obenkyo&hl=en. [Accessed 2017, Sept 28].

[7] Baharuddin, Rosnita, Dalbir Singh, and Rozilawati Razali, “Usability Dimensions for Mobile Applications,” Res. J. Appl. Sci. Eng. Technol, 5(6), pp. 2225-2231, 2013.

[8] Mayer, Richard, and Richard E. Mayer, “The Cambridge handbook of multimedia learning,” Cambridge University Press, 2005.

[9] Savignon, Sandra J., “Communicative competence,” TheTESOL Encyclopedia of English Language Teaching, pp. 1-7, 2018.

[10] Cook, Vivian, “Second Language Learning and language teaching,” Routledge, 2016.

[11] Cohen, Andrew D., “Strategies in Learning and using a Second Language,” Routledge, 2014.

[12] Craik, K. J. W., “The Nature of Explanation,” Cambridge University Press: Cambridge, 1943.

[13] Johnson-Laird, Philip Nicholas, “Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness,” 6, Harvard University Press, 1983.

[14] Senge, Peter M., “The Art and Practice of the Learning Organization,” pp. 464, 1990.

[15] Norman, Donald A., “Design Principles for Cognitive Artifacts,” Research in Engineering Design, 4(1), pp. 43-50, 1992.

[16] Jones, Natalie, et al., “Mental Models: An Interdisciplinary Synthesis of Theory and Methods,” 2011.

[17] Faulkner, Laura, “Beyond the Five-User Assumption: Benefits of Increased Sample Sizes in Usability Testing,” Behavior Research Methods, Instruments, & Computers, 35(3), pp. 379-383, 2003.

[18] Brata, Komang Candra, and Deron Liang, “An Effective Approach to Develop Location-Based Augmented Reality Information Support,” International Journal of Electrical and Computer Engineering, 9(4), pp. 3060-3068, 2009.

[19] Brata, Komang Candra, et al., “Design of Public Transportation Navigation System on Android Wear Device,” Sustainable Information Engineering and Technology (SIET), 2017 International Conference on, IEEE, pp. 385-391, 2017.

[20] Buglione, Luigi, and Alain Abran, “Improving the User Story Agile Technique using the Invest Criteria,” 2013 Joint Conference of the 23rd International Workshop on Software Measurement and the 8th International Conference on Software Process and Product Measurement, IEEE, pp. 49-53, 2013.

[21] Martignoni, Dirk, Anoop Menon, and Nicolaj Siggelkow, “Consequences of Misspecified Mental Models: Contrasting Effects and the Role of Cognitive Fit,” Strategic Management Journal, 37(13), pp. 2545-2568, 2016.

[22] Margutti, Piera, and Paul Drew, “Positive Evaluation of Student Answers in Classroom Instruction,” Language and Education, 28(5), pp. 436-458, 2014.

[23] Kaya, Sibel, Zeynel Kablan, and Diana Rice, “Examining Question Type and the Timing of IRE Pattern in Elementary Science Classrooms,” Journal of Human Sciences, 11(1), pp. 621-641, 2014.

[24] Kaya, Göknan, et al., “Discourse Patterns and Communicative Approaches for Teaching Nature of Science,” Education & Science/Egitim ve Bilim, 41(185), 2016.

[25] Mayer, R. E., “Multimedia learning,” Psychology of Learning and Motivation, 41, pp. 85-139, 2002.

[26] Mayer, R.E., “Principles of Multimedia Learning Based on Social Cues: Personalization, Voice, and Image Principles,” In Mayer, R.E. (Ed.), The Cambridge handbook of multimedia learning, New York: Cambridge University Press, pp. 201–212, 2005.

[27] Halford, Graeme S., “Children’s Understanding: the Development of Mental Models,” Psychology Press, 2014.

[28] Greca, Ieana Maria, and Marco Antonio Moreira, “Mental Models, Conceptual Models, and Modelling,” International Journal of Science Education, 22(1), pp. 1-11, 2000.

[29] Kohavi, Ronny, et al., “Online Experimentation At Microsoft,” Data Mining Case Studies, 11, 2009:39, 2009.

[30] Tang, Diane, et al., “Overlapping Experiment Infrastructure: More, Better, Faster Experimentation,” Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2010.

[31] Kohavi, Ron, and Roger Longbotham, “Online Controlled Experiments and A/B Testing,” Encyclopedia of machine learning and data mining, pp. 922-929, 2017.

[32] Azevedo, Eduardo M., et al., “The A/B Testing Problem,” Proceedings of the 2018 ACM Conference on Economics and Computation, ACM, 2018.

[33] Brata KC, Brata AH., “An Idea of Interactive Japanese Language M-Learning Application to Support Listening and Speaking Exercise,” In 2018 International Conference on Sustainable Information Engineering and Technology (SIET), IEEE, pp. 186-191, 2018.
User experience improvement of Japanese language mobile learning application … (Komang Candra Brata)