

A MOBILE APPLICATION FOR USER REGULATED SELF-ASSESSMENTS

Fotis Lazarinis¹, Vassilios S. Verykios¹ and Chris Panagiotakopoulos²

¹*School of Science & Technology, Hellenic Open University, Patras, Greece*

²*Department of Primary Education, University of Patras, Greece*

ABSTRACT

In this paper we present a mobile application for self-assessment. The work describes the main features of the application and focuses on its acceptance by students and the increase on their learning, through its usage in real testing settings. The application supports the retrieval of questions based on a number of criteria and it was evaluated with the aid of students who self-assessed their knowledge prior to in-class pencil and paper tests. An improvement in the performance of students who actively engaged with the system has been observed.

KEYWORDS

Assessment, mobile learning, self-assessment, adaptation.

1. INTRODUCTION

Mobile devices have significantly changed in recent years. While their primary purpose was to enable users to communicate through voice, their functions have considerably expanded, to simulate a resourceful personal computer. New applications are constantly being developed, giving the user personalized access to various data. Education is a field which could greatly benefit via this technology. Collaboration, personalized learning, peer-assessment, learning in context and all these novel ideas which have been proposed in the literature could more easily be realized by the utilization of a personal mobile device.

In this work, we focus on self-assessment through the utilization of a mobile application. Students are able to adjust the assessment data to prepare for the actual exams. They can define the topics and the difficulty level of the testing material and they get detailed explanations based on their answers to True/False, multiple choice and fill-in-the gap questions. The application has been implemented as an android application and utilized in real classes. The evaluation showed that the application is useful and students advanced their understanding on the selected topics. The main focus of the work is to discuss the effect in learning by using such a mobile application for exam preparation.

The paper is organized as follows: a literature review is provided in section 2, then in section 3 the system is described and in section 4 the evaluation results are presented. Conclusions and recommendations for future expansions are made in the final section.

2. LITERATURE REVIEW

Assessment plays an important part of every educational activity (Biggs, 1998; Gipps, 1999; Harlen, 2007). Several forms of assessment have been proposed. Fetterman et al. (1996) discuss the importance of self-assessment in knowledge. Self-assessment techniques have been applied in large scale in MOOCs and has been proved beneficial (Kulkarni et al., 2013). Formative assessment refers to assessment that is specifically intended to generate feedback on performance to improve and accelerate learning (Sadler, 1998). Tools for self-regulating the assessment process have been already proposed (Nicol & Macfarlane-Dick,

2006). Castle and McGuire (2010) conducted a study concerning student self-assessment of learning at an online university and suggested a positive impact on self-assessment to student learning. Self-assessment in secondary education students showed an improvement in performance and also improved their ability to self-assess their performance over time (Butler and Lee, 2010). With the improvement of technology and the ability to carry a personal computer anywhere and whenever, a number of systems have been proposed in order to take advantage of these technologies in assessment.

Hwang et al. (2011) proposed a formative assessment-based approach for improving the learning achievements of students in a mobile learning environment. Their experiments showed that their proposed approach promotes students' learning interest and attitude and improves their learning achievement. A formative assessment-based approach for improving the learning achievements of students in a mobile learning environment has been implemented. The researchers combined digital learning resources and real-world learning contexts to improve the understanding of students.

Lalos et al. (2009) developed a version of the "snakes and ladders" game for portable devices. They used a combination of location detection techniques, m-learning, m-assessment and learning standards to support the learning process of young children. Zhu et al. (2010) used a novel mobile telephone food application that will provide an accurate account of daily food and nutrient intake. They collected and evaluated dietary information reducing in that way the burden of more classical approaches for dietary assessment. This is an example of one of many practical aspects of mobile phones in various forms of assessment in a broader sense.

The effects of using vocabulary learning programs in mobile phones on students' English vocabulary learning have been investigated in (Başoğlu and Akdemir, 2010). The results of this study indicated that using mobile phones as a vocabulary learning tool is more effective than one of the traditional vocabulary learning tools.

In a more recent study, Bogdanović et al. (2014) showed that the integration of the mobile quiz application into Moodle improves students' results and increases satisfaction and motivation for using mobile devices in their learning process. Jacobs et al. (2015) developed an app which provides an individual student with summary feedback from tutors and provisional/final marks for all their assignments. The app is available on desktop and mobile devices. The aim of their work was to make feedback more accessible, and thus to engage students more actively in the process and this will have a beneficial impact on their learning.

Our work focuses on self-regulating formative assessments and we aim to make the process of self-assessment more accessible and adaptable to the students' goals. The system described below provides a summary of the process to make individual students aware of their true knowledge.

3. DESCRIPTION OF THE SYSTEM

mSAT (mobile Self-Assessment Tool) is an android app with the aim to help users to prepare for their exams through the execution of preparatory tests. This work is a new implementation for mobile devices of one of our older works presented in (Lazarinis et al., 2015). As in our older work, the main goal of the current work is to provide a flexible environment for self-assessment, where the test participants can regulate the testing process based on their current goals. However, the new design grants the users the freedom to test their knowledge anywhere and whenever they need to. Further, the system is enriched with more focused feedback per item and suggestions for further studying based on the performance of a student.

Each mobile app has a layered architecture, which implements the user experience, the business logic and the data. The data play a crucial point in this application as the testing material should be readily accessible once they are created or updated. Limitations arising from the features of the mobile device running the application (e.g., the characteristics of memory, storage, processor speed etc.) should not challenge the features of the application.

The application was developed in Android Studio with the SDK Android, using different libraries in order to extend the services with features such as cloud database. The selected cloud platform was the Parse Server at Parse.com in which a database was created. The main entities of our application are:

- Users: where the details of the registered users of the application are stored. It is used for the login and the authentication process.

- Topics: the details of the topics of the testing items.
- Question Items: contains the text, the type, the choices and the correct answer(s) of the testing items. These are managed centrally by a single administrator in the first version of the system, but adding more users with administrative rights is a straightforward approach.
- User Results: where the results for each attempt are stored.

Users can use the system either as registered users or as guests (see Figure 1 – images have been edited to present the messages in English). This flexibility was granted in order to increase the utilization of the application, as students could be discouraged to use the application if more information on their behalf is needed. The older experiences with the Web based version of the system, showed that many students were reluctant to register, due to lack of technical ability or techno anxiety or due to unwillingness to be engaged in such a process. Therefore, we preferred to offer both options with the goal to gradually persuade them to register through their involvement with the tool in order to benefit from the additional features, such as comparison to older attempts, etc.

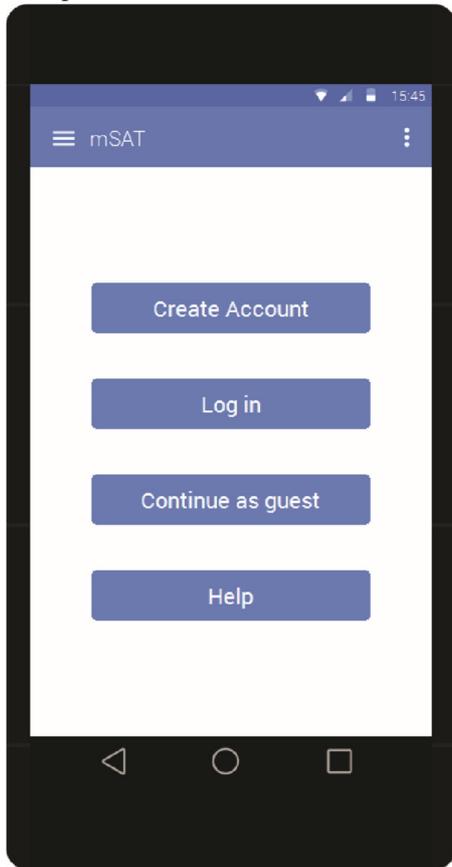


Figure 1. Initial Screen.

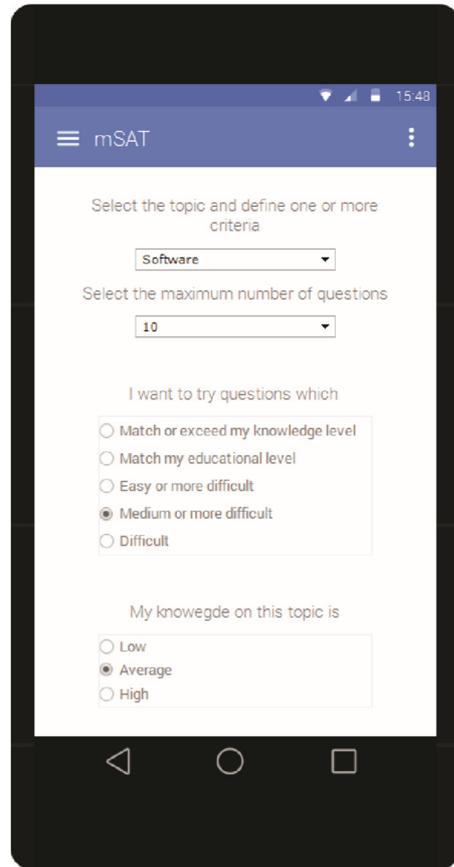


Figure 2. Self-assessment criteria.

The system is designed to be a flexible self-assessment tool, where users are able to regulate the testing process by defining their knowledge and their goals (see Figure 2):

- Selecting the assessment topics to be tested on.
- Describing their knowledge level and/or the level of completed education on the selected topics.
- Defining the characteristics (difficulty, type, number etc.) of the testing items they wish to attempt.

Simply put, the application logic is that administrators define and update the topic hierarchy; educators enrich the item bank with questions of various topics, difficulty and education level and students fulfill their assessment goals by selecting the most appropriate items to be tested on.

Test participants select testing items by applying one of the following strategies:

- i. The level of difficulty and/or the appropriate educational level of the testing items: Questions are classified as easy, medium or difficult, and learners can select assessment items based on their difficulty. They can select questions that match or pass or are below a specific level of difficulty, e.g., “show questions that are above average difficulty”.
- ii. The learners’ knowledge level: students can select questions that match or are above or below their knowledge level. Easy, medium and difficult questions correspond to low, good and high knowledge level. So students who, for example, have a “good” knowledge of a topic, they can define rules like “show questions greater or equal to my knowledge level”. In that case, the tool retrieves testing items with a high difficulty.
- iii. The learners’ educational level: this possibility supports the selection of questions based on the educational level they correspond to. Users are able to instruct the system “retrieve testing item that match their level of education”.

Alternatively students may ask the tool to decide on the arrangement of testing items based on the user provided data of registered users. The application retrieves items that correspond to the student’s educational level, on a specific topic. The questions are grouped by subtopic and by their level of complexity.

In the next step, the list of questions matching the criteria is presented and students can select anyone of the list to answer. The retrieved testing items are clustered by the subtopic they relate to. If a student has defined the maximum number of questions to attempt, and there are remaining questions in a subtopic, these are grouped at the end of the list of questions under a title called “More questions”.

Administrators and educators log in through a Web interface and update the question database. They define the text of the question, the possible answers, the correct one(s), the difficulty level and the educational level. They match each question to a broad topic or a narrow subtopic. In any case the question is automatically assigned to the topics in the higher level of the hierarchy. This supports more clustering alternatives which could eventually improve the self-testing process.

Once a student completes the self-evaluation of her/his knowledge, the tool estimates the knowledge level and presents statistics per topic. For registered users, these statistics are stored in the database to be utilized in future assessments for this user. The new level of knowledge is based on the average score on each topic. Scores are normalized based on their question weights. A mark below 50% results to “low” understanding while good knowledge means a score between 50% and 75%. A result greater than these limits, leads to high knowledge level. Further, statistics per topic and subtopic are presented to students who are informed about their weaknesses with the encouragement to study the relevant materials for the topics with low performance.

4. EVALUATION

mSAT is designed to be a flexible tool which can be used even between classes to help students assess their understanding using a multicriteria question retrieval approach. The main purpose of the current study is to discuss the results of an evaluation experiment with the aim to understand whether the tool has an impact to the learning of students. An evaluation experiment was carried out with the aid of 87 students of the first grade of a Greek senior high school.

The research goals of the experiment were:

- (i) to measure the potential improvement on the performance of the learners in regular in-class assessments, *and*
- (ii) to understand if the mobile version of the tool prompts users to be more engaged in the process of evaluating their understanding.

The trials run between September and December 2016. Students (aged 15 to 16) were randomly divided into two groups of 44 (Group A) and 43 students (Group B). During this period three paper & pencil in-class tests have been planned for both students groups. The in-class tests concerned basic algorithmic skills and Scratch programming questions. Group A (experimental group) had the ability to utilize the android application for one day prior to each test, while Group B (control group) did not have access to the application. The questions loaded in the application concerned the same topics but they were not same as those asked in the summative tests. Moreover, we only included a few representative questions of varying difficulty for each subtopic. We primarily needed to make the students aware of their true knowledge per subtopic and topic they

worked on and not to simply give them the opportunity to practice extensively on questions which resemble the actual testing items.

Prior to the experimental setting, we administered the same paper & pencil test to both groups to understand whether there are significant differences in the distribution of knowledgeable students. The t-test run with the scores of the two groups showed no significant difference in the results. The two groups of students had statistically equivalent abilities before learning the subject unit. This outcome was expected as both groups were randomly created and consisted of an enlarged number of students.

Table 1 shows the results of the first in-class paper & pencil test. Group A had an average of approximately 64.55% and Group B an average of 62.69%. The distribution of the results are comparatively similar and again the t-test showed no significant difference. Admittedly, these first results were discouraging. However, we looked closer to the individual results and we compared them to the results of the pre-experiment test and also to the utilization of the application among students. Through this comparison, it was made obvious that only 12 students of Group A used the application, which is about 25% of the student population of the experimental group. Furthermore, only 4 of them actually retrieved questions with a difficulty level higher than their knowledge level, as it was provided to the tool by them prior to the testing. These students received some alarming statistics at the end of the evaluation and as they informed us in the focused personal interviews, they had to re-study for the exam. Thus they had actually an improvement on the actual summative test. The other 8 students, either tried easy questions or quitted the application early or had already a high knowledge level, so they received no worrying results.

Table 1. Scores of the student groups in the first in-class paper & pencil test.

Score	Group A	Group B
0-50%	9	8
50%-75%	27	30
75%-100%	8	5

This qualitative analysis of the individual results showed some improvement to some students and these participants actually “spread the word” to their peers. So, in the second phase of the experiment, 33 students used the mobile application and we measured an improvement in the scores. The average score of Group A increased to almost 67%, while the average score of the Group B increased by approximately 1%, i.e. to 63.4%. The t-test showed that the two averages are different (t-value=2.56677 at $p < .05$). The greatest difference was in the middle class where the averages score was 68.24% for Group A and 63.4% for Group B. All the students of Group A of this class have used the application to practice and to study again some of the learning materials.

Table 2. Scores of the student groups in the second in-class paper & pencil test.

Score	Group A	Group B
0-50%	8	5
50%-75%	25	30
75%-100%	11	8

The last in-class test, showed again an improvement to the performance of the students of Group A, who used the application. The marks of Group A in the individual classes have been increased to almost reach the upper limit of each class which was not the case in Group B. By this trial, several students of Group B wanted to try this application as well, especially students who performed worse than expected in the previous two tests.

Table 3. Scores of the student groups in the third in-class paper & pencil test.

Score	Group A	Group B
0-50%	7	6
50%-75%	24	27
75%-100%	13	10

Based on the statistics, the comparison of the individual cases, and the opinions of the students during the focused interviews, we have strong indications that the application helped the students to get a better mark and prompted users to be more engaged in the process of evaluating their understanding. Further, students found the application easy to use and they liked the fact that it supports various options. In the next round of evaluations we will look closer on the specific adaptive options offered by the tool with respect to their

utilization by the students and the opinions of the students. That way the default options of the system could be better customized to more effectively guide the student.

5. CONCLUSION

In this paper we briefly presented the characteristics of a mobile tool for self-assessments and we focused on understanding its overall pedagogical value in real learning settings. The experimental group has shown an increase in the utilization of the application between the trials and an increase on the average marks. A pre-experiment test was administered to ensure that the random distribution of the student population did not lead to groups with uneven distribution of knowledgeable students. Through focused interviews, students shared their opinions which were quite positive towards the tool.

In this version of the system the focus was on the acceptance by the students and the pedagogical gains. Several aspects of the tool need to be tested. For example, the importance and the extension of the criteria for customizing the assessment process is a very significant research direction. In the current experimentation we worked with senior high school students and therefore all the material was related to this educational level. Would such a system be acceptable to tertiary education students or for vocational training? This is another research path. The feedback needs to be examined. Has it been used properly or did the students find it helpful? Implementation issues could also be explored. Would the response be acceptable in case the database increases significantly?

All things considered, this tool has been proved to have a practical value and if its database is enriched with more topics it could become a very useful teaching assistant.

ACKNOWLEDGEMENT

This work has been supported by a grant from Stavros Niarchos Foundation (SNF).

REFERENCES

- Baçoğlu, E. B. and Akdemir O. (2010). A comparison of undergraduate students' english vocabulary learning: using mobile phones and flash cards. *TOJET: The Turkish Online Journal of Educational Technology* Vol. 9 No 3
- Biggs J. (1998) Assessment and classroom learning: A role for formative assessment? *Assessment in Education: Principles, Policy and Practice* 5(1): 103–110
- Bogdanović, Z., Barać, D., Jovanić, B., Popović, S., and Radenković, B., (2014) Evaluation of mobile assessment in a learning management system, *British Journal of Educational Technology*, Vol. 45, No 2. pp. 231–244
- Butler, Y. G., and Lee J. (2010). The effects of self-assessment among young learners of English. *Language Testing*, 27(1), 5–31.
- Castle S., and McGuire, C. (2010). An Analysis of Student Self-Assessment of Online, Blended, and Face-to-Face Learning Environments: Implications for Sustainable Education Delivery. *International Education Studies*, 3 (3), p. 36-40.
- Fetterman, D. M., Kaftarian, S. J., & Wandersman, A. (Eds.). (1996). *Empowerment evaluation: Knowledge and tools for self-assessment and accountability*. Thousand Oaks, CA: Sage.
- Gipps, C. V. (1999). Socio-cultural aspects of assessment. In P.D. Pearson & A. Iran-Nejad (Eds.), *Review of Research in Education*, Vol. 24, pp. 355-392.
- Harlen, W. (2007). *Assessment of learning*. Sage Publications Ltd, London.
- Hwang, G.-J., Chang, H.-F. (2011). A formative assessment-based mobile learning approach to improving the learning attitudes and achievements of students. *Computers & Education*, 56, pp. 1023–1031.
- Jacobs, C., Wilson, M. and Barker, C. (2015) Using the latest developments in mobile technology to improve students' engagement with assessment feedback, *International Conference on Interactive Mobile Communication Technologies and Learning (IMCL)*.
- Kulkarni, C., Koh, P. W., Le, H., Chia, D., Papadopoulos, K., Cheng, J., Koller, D., & Klemmer, S.R. (2013). Peer and self-assessment in massive online classes. *ACM Transactions on Computer-Human Interactions*, 9(4) Article 39.

- Lalos, P., Lazarinis, F. and Kanellopoulos, D. (2009). E-snakes and ladders: a hypermedia educational environment for portable devices. *IJMLO* 3(2), pp. 107-127
- Lazarinis, F., Verykios, V.S. and Panagiotakopoulos C. 2015. An Adaptive Web Tool for Self-assessment using Lightweight User Profiles. *7th International Conference on Computer Supported Education (CSEDU)*, Lisbon, Portugal, pp. 14-23.
- Nicol, D., Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), pp. 199–218.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18, 119–144.
- Zhu, F., Bosch, M., Woo, I., Kim, S., Boushey, C., Ebert, D., and Delp, E. (2010). The Use of Mobile Devices in Aiding Dietary Assessment and Evaluation, *IEEE J Sel Top Signal Process*, 4(4): 756–766.