USING ADAPTIVE MASTERY TESTING IN ASSESSMENT MANAGEMENT SYSTEMS

Muhittin Şahin[1], Furkan Aydın[2], Sema Sulak[3], Cennet Terzi Müftüoğlu[4], Mustafa Tepgeç[5], Gizem Karaoğlan Yılmaz[3], Ramazan Yılmaz[3] and Halil Yurdugül[6]
*[1]Ege University, Turkey; University of Mannheim, Germany **
[2]Kahramanmaras Sutcu Imam University, Turkey
[3]Bartın University, Turkey
[4]Tokat Gaziosmanpasa University
[5]Hacettepe University; Mustafa Kemal University, Turkey
[6]Hacettepe University, Turkey

ABSTRACT

The use of technology for teaching and learning has created a paradigm shifting in learning environments and learning process, and also the paradigm shifting has also affected the assessment processes. In addition to these, online environments provide more opportunities to assess of the learners. In this study, the Adaptive Mastery Testing (AMT) system in Assessment Management System was designed and developed in which students can test themselves, recognize their strengths and weaknesses, and determine their learning objective based competencies. AMT environment is structured in accordance with the rapid prototyping software developing model. In this environment, there are questions for the students about four learning objectives, which are among the basic subjects of the Statistics course. AMT environment consists of presentation, assessment, domain and learner model. Pilot implementation was carried out with 98 undergraduate students. In order to evaluate of the environment; number of tests taken, number of correct answer, number of wrong answer and number of total answer data were used. According to the findings, it is seen that most of the students are masters in the learning objectives presented to them. In addition, it was found that half of the students took an average test to become a master. In other words, half of the learners participating in the study were determined as masters by the system in the first test they took for each learning objective.

KEYWORDS

E-Assessment, Adaptive Mastery Testing, Assessment Management System, Rapid Prototyping, Statistic, Log Data

1. INTRODUCTION

The use of technology for teaching and learning has created a paradigm shifting in learning environments and learning process, and also the paradigm shifting has also affected the assessment processes. With this paradigm shifting, it’s expected that learners be aware of their learning processes and guide their learning processes. In the literature, it is stated that students who cannot guide their own learning processes cannot make use of e-learning processes adequately (Gibbons, 2003). In other words, it can be said that e-learning environments work in favor of autonomous learners (Grow, 1991). In this context, individual online learning environments and assessment environments and designs that recognize and guide the learner, enable them to learn at an individual pace, come to the fore. Assessments within the scope of increasing learners' awareness of learning and monitoring learning processes are considered as formative assessment (Yurdugül & Bayrak, 2014). Formative assessment aims to contribute to the learner’s learning process by providing information to the learner about their performance (Sadler, 1998). However, it seems that this assessment cannot be made effectively in physical environments (Mok, 2010). Online environments provide more opportunities to assess of the learners (Oosterhof, Conrad, & Ely, 2008). In this context, e-learning environments where students can test themselves and make a judgment about themselves contribute to the e-assessment process (Bayrak, 2014). In this study, the Adaptive Mastery Testing (AMT) in Assessment Management System (AMS) was designed and developed in which students can test themselves, recognize their strengths and weaknesses, and determine their learning objective based competencies. This system design is based on the Learning Management System

(LMS) architecture. It is thought that such AMS systems can be integrated with LMS in the further researches. Besides this, the AMS expression was used in the research due to the use of metrics in the assessment tasks of the students. In addition, in the research, 98 undergraduate students used this system as a pilot study and the findings according to the use of the system by students were included.

1.1 Adaptive Mastery Testing (AMT)

The key feature of adaptive tests is that each respondent is tested based on their ability (Eggen, 2004; Wainer,1990; Weiss, 1983). The most important adaptive testing application is Computerized Adaptive Testing (CAT) which is based on Item Response Theory (IRT). The critical component of adaptive tests is the selection of the items that convey the significant information regarding the ability of the examinee as well as the fact that each respondent receives items that adapted. Thus, adaptive test applications require a large item pool of computed parameters. (Kingsbury & Zara, 1991; Van Der Linden & Glas, 2000). Despite adaptive item administration is a significant feature of a learning system, models utilizing a unidimensional representation of ability, such as the standard IRT model, are not capable of exploring what aspects of the material the individual has mastered or not mastered (Deonovic et al, 2018).

On the other hand, variable-length mastery tests maximize the possibility of making correct classification decisions as well as shorten the test length (Lewis, Sheehan; 1990). Adaptive Mastery Testing (AMT) is a method in which item selection methods from variable-length mastery tests and test termination rules are applied adaptively (Kingsbury & Weiss, 1983). Mastery testing is performed to decide whether an individual is master or non-master according to the test result observed in a learning objective/subject (Vos & Glas, 2000). There are two basic approaches called sequential and adaptive mastery tests in mastery testing where the number of items is uncertain (Vos & Glas, 2000). Within the scope of the research, an AMT environment was designed and developed to decide the master or non-master levels of the students on a learning objective. The Sequential Probability Ratio Test (SPRT) algorithm has been administered to determine the master/ non-master levels of the students. Item-examine incompability is employed as the item-selection algorithm. These processes take place until it is decided the examinee is a master/nonmaster. This rule, introduced by Wald (1947), is widely used for IRT-based adaptive computer testing. When this test is utilized, it is possible to reach more consistent and optimal results with fewer questions (Spray & Reckase, 1996). Since one of the limitations of the study is the number of questions, the SPRT algorithm, which is thought to provide an optimal solution, was utilized to determine the master /non-master levels of the students.

2. METHOD

Adaptive Mastery Testing e-assessment environment which determine whether the students are master or non-master was developed. This environment was developed based on the rapid prototyping software development process (Figure 1).

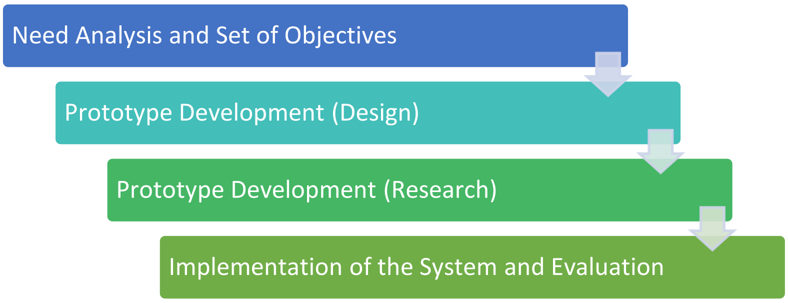

Figure 1. The steps of rapid prototyping software development process (Tripp & Bichelmeyer, 1990)

As seen in Figure 1, this process starts with the needs analysis and ends with the implementation and evaluation phase of the system. In this section, information about the participants, AMT e-assessment environment, data collection tools and data analysis were presented.

2.1 Participants

Participants of the study consist of 98 undergraduate and graduate students at different state universities. 65% (64) of these students are female and 35% (34) are male students.6

2.2 Adaptive Mastery Testing Environment

In this research, AMT in an AMS which determine whether students master or non-master about a learning objective was developed and presented to the students. In this environment, students take the tests on basic concepts, data summarization, central tendency measures and distribution measures in the basic statistics course. This AMS system is structured on the basis of LMS architecture and system architecture and processes is presented in Figure 2.

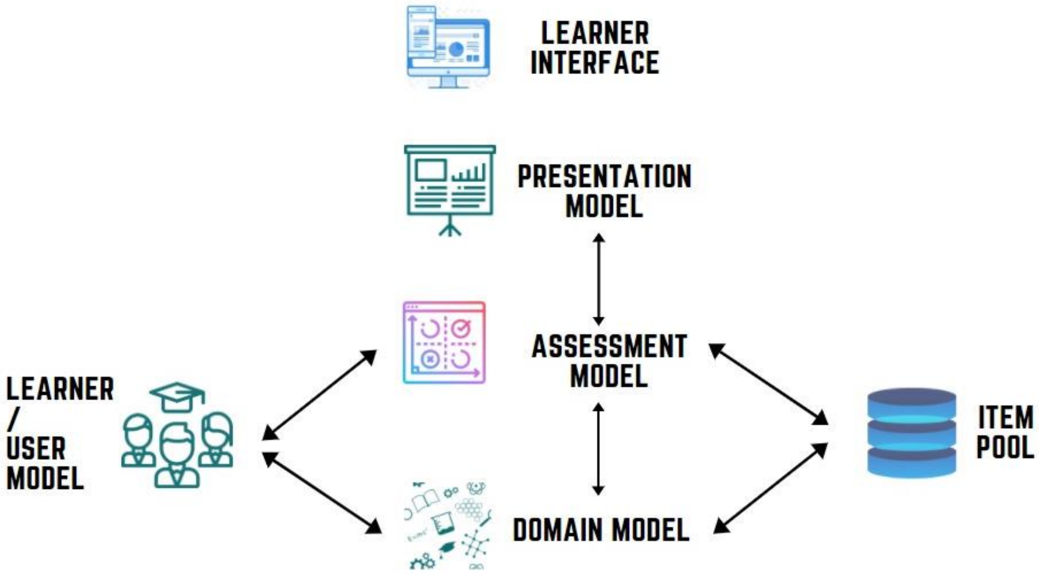

Figure 2. System architecture

In seen Figure 2, AMT consists of learner model, assessment model, domain model, item pool, presentation model, and learner interface. seen in Figure 2, AMT consists of learner model, assessment model, domain model, item pool, presentation model, and learner interface.

2.2.1 Presentation Model and Learner Interface

Students can access the system through learner interface. Students can login to the system by entering their username and password. These usernames and passwords created by the researchers and were delivered to the students. After login to the system, students can see the learning objectives whether they are master or non-master. Then they can take to test and test themselves. The screenshot of the AMT environment is presented in Figure 3.

ISBN: 978-989-8704-33-7 © 2021

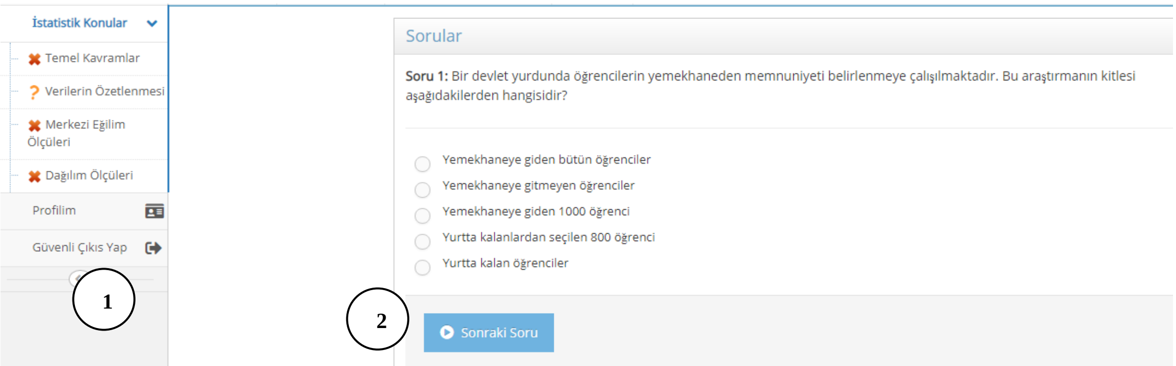

Figure 3. Screenshot of the adaptive mastery testing environment

Area 1 shows the learning objectives of the students. Here (x) sign indicates that the student has taken the test for that learning objective but is not a master, (?) Has not yet taken any test for the learning objective, and (√) indicates the learning objective that the students is master. The number 2 area shows where questions are presented to the students.

2.2.2 Learner/User Model

A learner / user model was created in order to access students to interact with the system and to store and update learner information. In this model, data is stored about the users such as a) usernames, b) user passwords, c) user codes, and d) information about the individual characteristics of the users. This model is associated with the presentation model, evaluation model and domain model.

2.2.3 Assessment Model

This model consists of the rules when configuring AMT. Within the scope of this system, SPRT algorithm has been utilized. There are more advanced versions of SPRT algorithms such as EXSPRT, EXSPRT-I, but these algorithms require item parameters. Therefore, the SPRT algorithm has been utilized within the scope of the research in order to get faster and more effective results. SPRT can produce more efficient results than CAT with fewer items (Rudner, 2001). SPRT uses the information which points located above or below the threshold score determine the classification status (Parshall et al., 2002) then, decide students is master or non-master on this subject. The test terminates when the individual is classified in one of the predetermined categories (Thompson, 2007). SPRT was used as a termination rule. There are two hypotheses for the SPRT such as H_0 (null hypothesis) and H_1 (alternative hypothesis) and presented below:

$$H_0: \theta_i \leq \theta_0 - \delta = \theta_1$$

$$H_1: \theta_i \geq \theta_0 + \delta = \theta_2$$

θ_i: ability level; θ_0: cut point; δ: indifference region

According to these hypotheses, the decision is made by the system, and the test is terminated. Learners were classified as master or non-master about the subject in the context of this research.

2.2.4 Domain Model

This model focused on the learning objective which are the smallest part of the learning process, and the competence of the learners (master or non-master) is revealed for these learning objectives. Domain model consists of a) course, b) unit, c) topic, d) learning objective (subject).

2.2.5 Item Pool

The item pool has been structured in suitability with the system. While structuring the item pool, each question was structured in accordance with the relevant learning objective and appropriate meta-data structures were used in this context.

2.3 Data Collection Tools and Data Analysis

Within the scope of this research, the system was presented to use 98 students and the log data of these students' system usage was collected as the log data. The data consists of a) the number of take test by students, b) the number of correct answer, c) the number of incorrect answer, and d) the total number of attempt to response. In addition, new metrics were calculated from these metrics. Detailed information about this metrics is presented in the findings section. Descriptive information about the log data obtained from the system was included in the analysis of the data.

3. FINDINGS

Information about the log data of the students that used the system Table 1. This information consists of number of different learning objective, number of master learning objective, number of test by taking students, number of correct answer, number of incorrect answer, and number of total response.

Table 1. Information about the log data obtained from the system usage

	N	Min	Max	X_{ave}	SD
Number of different learning objective	98	1	4	3.33	1.05
Number of master learning objective	98	0	4	2.63	1.40
Number of taking test	98	1	12	4.86	2.72
Number of correct answer	98	0	86	32.56	18.20
Number of incorrect answer	98	0	70	20.89	17.84
Number of total response	98	4	153	53.45	33.73

While in total four learning objectives were presented to the students, it can be stated that students had taken the learning objective on average approximately 3 (3.33). It's drawn the attention that the number of master learning objective is 2.63. It's seen that the number of test which taken by students is 4.86. The reason why the number of tests taken is higher than the number of learning objective is that non-master learners can take the tests over and over again. It is seen that the students responded 53.45 items on average and approximately 32.56 of these responses were answered correctly. In addition, students' total correct answer rate was 63%, and the rate of incorrect answers was 37%. On the other hand, new metrics are also calculated in order to obtain more detailed information according to these system metrics. For this purpose, learning objective ratio (Figure 4) and attempt ratio (Figure 5) were calculated.

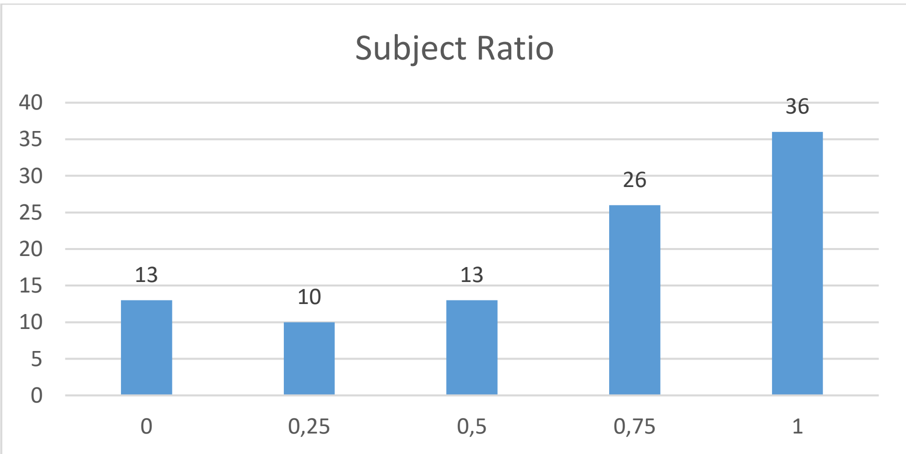

Figure 4. Learning objective/subject ratio

The learning objective/subject ratio was obtained by dividing the number of master learning objective by the total number of learning objective. If this ratio is high, it means that the student is so successful on the basis of learning objective. According to the results, it is understood that a) 37% (36) of the students in the system are masters in all learning objectives, b) 27% (26) in three learning objectives, c) 13% (13) in two learning objectives, d) 10% (10) in one learning objective. It can be stated that 13% (13) of the students are non-masters in any of the learning objectives.

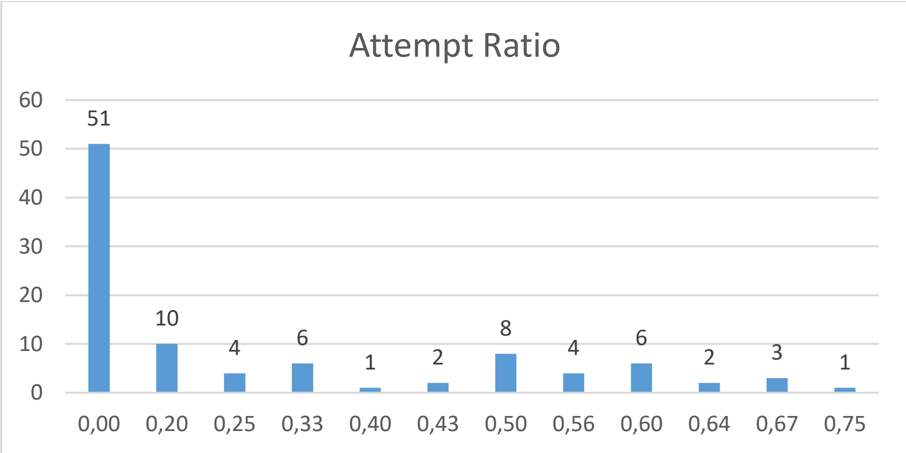

Figure 5. Attempt ratio

In order to calculate the attempt ratio in the first step, the number of learning objectives is divided by the number of tests. Then this value obtained was subtracted from 1. If the value is increasing it means that students take tests many times in the same learning objective. It shows that 52% (51) of the students were determined as a master by the system for the first time.

4. CONCLUSION

In the system developed within the scope of the research, the learning management system architecture was grounded due to the integration of these kinds of systems with learning management systems in the future. The competencies of students can be monitored and students' learning awareness can be increased through AMT during the learning processes provided that LMSs and AMSs using the same learner model are integrated. The literature reveals that the mainly used metrics in the learning analytics are learning performances, that is to say, test results (Aguilar et al., 2021; Kia et al., 2020; Tan et al., 2016). It is projected that this kind of integration will enrich learning analytics.

The AMT environment was developed in order to determine the students' master/non-master status according to a learning objective within the scope of this research. Components in this environment and results obtained from the pilot application were included. Sequential Probability Ratio Test (SPRT) algorithm was used in order to classify the students as master or non-master and item-examine incompatibility was set to work as the item-selection algorithm. The students were submitted with tests concerning four achievements towards the statistics course. According to the obtained results, it can be expressed that most of the students interacted with most of the achievements submitted to themselves. In addition, the fact that more than half of the students were determined as the master in the first test they received for each achievement is another finding we obtained.

The AMT developed as part of the research is limited to the SPRT algorithm. Alternatively, it is possible to use more advanced algorithms such as EXSPRT or EXSPRT-I using the item parameters. Another limitation of the study is the usage of log data only in order to evaluate the system. In the next stage, it is projected to perform an analysis in terms of system usability and student satisfaction.

ACKNOWLEDGEMENT

This study is financially supported by the Scientific and Technological Research Council of Turkey (TUBITAK grant ID: 119K430).

REFERENCES

Aguilar, S. J., Karabenick, S. A., Teasley, S. D., & Baek, C. (2021). Associations between learning analytics dashboard exposure and motivation and self-regulated learning. *Computers & Education*, 162, 104085.

Bayrak, F. (2014). The effectiveness of perceived self-intervention in web based self-assessment system. Hacettepe University (Unpublished doctoral dissertation), Ankara, Turkey.

Deonovic, B., Yuseldon, M., Bolsinova, M., Attali, M., and Maris, G. (2018). Learning meets assessment. Behaviormetrika, 45, 457-474. doi: 10.1007/s41237-018-0070-z

Eggen, T. J. H. M. (2004). Contributions to the theory and practice of computerized adaptive testing (Doctoral dissertation). University of Twente, Enschede.

Gibbons, M. (2003*). The self-directed learning handbook: Challenging adolescent students to excel*. John Wiley & Sons.

Grow, G. O. (1991). Teaching learners to be self-directed. *Adult Education Quarterly*, *41*(3), 125-149.

Kia, F. S., Teasley, S. D., Hatala, M., Karabenick, S. A., & Kay, M. (2020, March). How patterns of students dashboard use are related to their achievement and self-regulatory engagement. *In Proceedings of the Tenth International Conference on Learning Analytics & Knowledge* (pp. 340-349).

Kingsbury, G. G., Zara, A. R. (1991). A comparison of procedures for content-sensitive item selection in computerized adaptive tests. Applied Measurement in Education, 4, 241-261

Kingsbury, G.G. & Weiss, D.J. (1983). A comparison of IRT-based adaptive mastery testing and sequential mastery testing procedure. In D.J. Weiss (Ed.), New horizons in testing: Latent trait test theory and computerized adaptive testing (pp. 257–283), New York: Academic Press

Lewis, C., & Sheehan, K. (1990). Using Bayesian decision theory to design a computerized mastery test. *Applied Psychological Measurement*, 14, 367-386.

Mok, M. M. C. (2010). *Self-directed learning oriented assessment: Assessment that informs learning & empowers the learner*. Hong Kong: Pace Publications.

Oosterhof, A., Conrad, R. M., & Ely, D. P. (2008). Assessing learners online. New Jersey: Pearson.

Parshall, C. G., Spray, J. A., Kalohn, J., & Davey, T. (2002). Practical considerations in computer-based testing. Springer Science & Business Media.

Rudner, L. M. (2001). Measurement Decision Theory.

Sadler, D. R. (1998). Formative assessment: Revisiting the territory. *Assessment in Education: Principles, Policy & Practice*, *5*(1), 77-84.

Spray, Judith A. & Reckase, Mark D. (1996). Comparison of SPRT and Sequential Bayes Procedures for Classifying Examinees into Two Categories Using a Computerized Test. *Journal of Educational and Behavioral Statistics*, 21(4), 405-14.

Tan, J. P. L., Yang, S., Koh, E., & Jonathan, C. (2016, April). Fostering 21st century literacies through a collaborative critical reading and learning analytics environment: user-perceived benefits and problematics. *In Proceedings of the sixth international conference on learning analytics & knowledge* (pp. 430-434).

Thompson, N. A. (2007). A practitioner's guide for variable-length computerized classification testing. *Practical Assessment, Research, and Evaluation, 12*(1), 1.

Tripp, S. D. & Bichelmeyer, B. (1990). Rapid prototyping: An alternative instructional design strategy. *Educational Technology Research and Development, 38*(1), 31-44.

Van der Linden, W. & Glas, C. (2000). Computerized adaptive testing: Theory and Practice. Dordrecht, The Netherlands: Kluwer Academic Publishers

Wainer, H. (Ed.). (1990) Computerized adaptive testing: A primer. Hillsdale. NJ: Lawrence Erlbaum Associates.

Wald, A. (1947). Sequential analysis. New York: Wiley

Weiss, D. J. (Ed.). (1983) *New horizons in testing*. New York: Academic Press

Yurdugül, H., & Bayrak, F. (2014). The acceptance of web based formative assessment system for primary school students. *Journal of Educational Sciences & Practices*, *13*(26).