Automated e-Assessment: Students' Needs and e-Evaluation Solution Possibilities

A. Margiene and S. Ramanauskaite

***Abstract*—E-learning is increasingly gaining its usage. In the COVID-19 period, its importance even increased. One of the important e-learning processes is e-assessment. Suitable e-evaluation implementation depends on multiple factors. One of the factors for suitable testing implementation is alignment between users' needs for a specific purpose testing and its implementation possibilities. Therefore in this paper, we analyze what are students' needs for e-evaluation system in knowledge level testing and self-evaluation testing cases. Student survey and multi-criteria decision-making were used to find out how students rank different criteria of e-assessment and how linear, graph-based, and tree-based testing structures rank based on these criteria.**

***Index Terms*—E-evaluation, e-learning, needs, testing structure, Analytic Hierarchy Process (AHP).**

I. Introduction

Knowledge assessment is one of the key elements in the education process. Knowledge evaluation is used as a means to analyze, guide, and promote students' performance [1]. E-assessment can improve student learning as well, encouraging students to learn sincerely and apply a deep learning approach to get some confirmation, evaluation of his or her study results [2]. At the same time, it is very important to assign tasks of appropriate complexity and properly designed to the learner. Only tasks of the right complexity increase the learner's motivation, cause a state of flow and help learners to remain in this state for some time [3], [4]. Flow is "the state in which people are so intensely involved in an activity that nothing else seems to matter; the experience itself is so enjoyable that people will do it even at great cost, for the sheer sake of doing it" [3]. When the challenges are too small, the flow is returned by increasing them. If the challenges are too great, one can return to the state of flow by learning new skills [4]. Therefore, learners need to be given tasks and learning materials of the right complexity and presented in the right order. This is especially important in personalized and adapted e-learning systems. In adaptive or/and personalized e-learning systems the possibility for the learner to properly assess his / her level of knowledge, and for the teacher to present tasks during the assessment in such a way that the learner's level of knowledge would be properly assessed is one of the most important factors [5]-[8].

However, at the moment research on adapted e-evaluation systems design and test purpose is not analyzed enough. This does not allow estimation on which form of e-evaluation structures is more appropriate for adapted testing. Therefore, this paper aims to increase the understanding of knowledge evaluation and self-evaluation adaptive e-testing structures suitability to meet students needs.

Manuscript received June 1, 2021; revised September 26, 2021.

The authors are with the Department of Information Technologies, Vilnius Gediminas Technical University, Vilnius, LT-10223, Lithuania (e-mail: asta.margiene@vilniustech.lt, simona.ramanauskaite@vilniustech.lt).

II. Theoretical Background

The importance of knowledge testing and various aspects of test development has been explored by multiple studies and different aspects.

A. Influence of Test Question Types on Learning and Knowledge Assessment

Smith and Karpicke [9] investigated the effects of different question formats (short-answer, multiple-choice, hybrid questions) on learning. These authors investigated the appropriateness of multiple-choice tests as a means of assessing student achievement in human anatomy. Bulgakov and Dedikova [10] found that using single test tasks with the ordering of answers, unlike other types of tests (single choice, multiple-choice, establishing a sequence, and establishing compliance), provides the developer with significantly more opportunities to shape the assessment scale and provides a more interesting analysis of student performance monitoring results. Therefore AlMahmoud *et al*. [11] describe their approach to the standardized format used for multiple-choice questions (MCQs) assessment and provide recommendations on how to improve the assessment of high-stakes. Nnodim [12], as well as Thompson and Husmann [13] investigated the appropriateness of multiple-choice tests as a means of assessing student achievement in human anatomy and provided recommendations for the creation of better multiple-choice questions too.

Despite the question type appropriateness for a specific purpose, the content of the question should be aligned with the test purpose as well. Therefore Hartell and Strimel [14] examine content validity in teacher-made tests for elementary technology education. Lopez de Arana *et al*. [15] describe the procedure used to develop a self-assessment questionnaire for university service and learning experiences and validate it using a modified online Delphi method.

B. Advantages and Disadvantages of e-Testing

E-assessment is based on mostly automated knowledge testing, which does not require interactive teachers' interactivity. The automation influences faster evaluation process, increase test availability however at the same time might lose the interactivity and adaptability, which can be provided by face-to-face evaluation with the teacher.

doi: 10.18178/ijiet.2022.12.3.1612

Peytcheva-Forsyth and Aleksieva [16] presented the results of a student survey on students' experience in e-assessment during the COVID-19 pandemic and students' views on e-assessment. The results proved the e-assessment solutions require some additional work, not just computerization of the same face-to-face used testing solutions. Otherwise, it will decrease the trust of students in e-assessment and reinforce their preference for a face-to-face assessment.

Cigdem and Öncü [17] conducted a survey study that examined how students' perceived usefulness of e-testing influences course grades and students' attitudes towards the e-assessment system. The results revealed perceptions about question contents significantly affected perceived usefulness. At the same time, a suitable introduction to the testing system increases students' satisfaction and usefulness of the e-testing system.

Nuha *et al*. [18] point to the benefits and obstacles of using e-assessment in learning from different domains: student, teacher, institution, and educational aims. They identified the main advantages (providing direct and immediate feedback for students, improving student performance, reducing the time and effort of the teacher, decreasing the cost for the institution, and encouraging high-order thinking). At the same time, they present the disadvantages of e-assessment (poor technical infrastructure, unfamiliar students with computers). Meanwhile, Appiah and Tonder [2] focuses on concepts such as tasks that can be accessed through e-assessment, benefits, and challenges of e-assessment, and principles of e-assessment.

A study by Peytcheva-Forsyth and Aleksieva [16] found that when students were asked to indicate the disadvantages of e-assessment, as many as 28.44% of respondents indicated that they could not demonstrate their knowledge well enough. They consequently were asked if they would rather be assessed online than face-to-face. 18.57% of respondents indicated that they would prefer to be assessed in face-to-face mode rather than online distance mode, while 31.84% of respondents indicated that e-assessment is appropriate only in some courses. The results reveal that e-assessment still has shortcomings that need to be addressed.

C. Pre-tests in Personalized and Adaptive e-Learning Systems

Jagadeesan and Subbiah [5] have developed a Skill-based e-learning environment in which all learners are divided into three levels (basic, regular, and advanced) after the skills test. The learning content is provided only based on their skills test assessment reports. Athanasiadis *et al*. [6]developed the "Learning' platform" to introduce a personalization mechanism that automatically changes the levels of complexity of the system according to the student flow. Nabizadeh *et al*. [7] describe two approaches to maximizing user grades for a course while respecting their time constraints. These approaches recommend successful paths based on available time and user initial knowledge level. Trouss *et al*. [19] provide a hybrid model to detect misconceptions using machine learning, and a technique to automatically model student learning and forgetting process using a fuzzy inference system. The fuzzy inference system uses each student's level of knowledge in one language and can diagnose his or her level of knowledge in another language, allowing the system to create an adaptive learning environment for each student. Hariyanto and Kohler [8] proposed an adapted e-learning system based on the different contributions of learning style and initial learner knowledge measured by a pre-test in five sections. If the test result meets or exceeds the standard grade set by the teacher, the student passes this section. These conditions affect the appearance of links in the menu area.

D. The Adaptability of Knowledge Assessment Tests

Mustakerovo and Borissova [20] proposed an educational Web-based e-testing system that students can use to test their knowledge and teachers to take a formal exam. The test questions are divided into easy or advanced questions, and this allows students to choose between easy, advanced, or all questions passing modes. This e-evaluation system is based on task difficulty only. Arif *et al*. [21] expanded a conceptual approach for developing an educational web-based e-testing system proposed by Mustakerov and Borissova. They proposed a multi-layered architecture of intelligent agents for e-testing and e-learning systems. The test difficulty levels were increased to three (easy, medium, and difficult) and further divided into sub-levels. Learner at first has to pass three easy level sub-tests and two sub-tests from medium level and only then he will be able to take the advance level test. This approach is similar to the competence tree solution [22] for e-evaluation, where the competence tree is designed and tests are associated with one or multiple competencies in the tree. Based on the testing needs, the student might choose which competencies he or she wants to demonstrate or in another case, top-to-bottom or bottom-to-top approaches can be used to go from hardest tasks to easier ones or via versa, based on whether the student solved the previous task or no.

Another solution for adapted e-testing, based on students' success on previous tasks is contextual graphs [23]-[25]. The teacher or designer of the contextual graph is responsible for designing possible testing paths, which students will follow, based on the results of each test question.

III. Methodology for e-Assessment Criteria Evaluation

An analysis of the literature shows that adaptive testing is important in both the development of personalized e-learning systems and adaptive knowledge assessment systems. However, research that reveals what test criteria for selection and presentation of tasks are important in self-control tests and final knowledge tests is lacking. When developing adaptive e-testing systems, it is important to know whether the same criteria apply to the case of final knowledge (FK) evaluation and the case of self-testing (ST). Therefore, a survey on students' needs for FK and ST in e-assessment was executed.

To get students' need the Analytic Hierarchy Process (AHP) method of Multiple-criteria decision-making was used. Application of multi-criteria decision making allows estimation of users' needs (by gathering criteria importance) as well as a ranking of analyzed alternatives, based on their

quantitatively estimated property values [26]. To clarify the relationships between the criteria and its sub-criteria, we created the criteria tree, where three main criteria (test structure, adaptability, and feedback) were selected and each of them was divided into a sub-criteria (see Fig. 1).

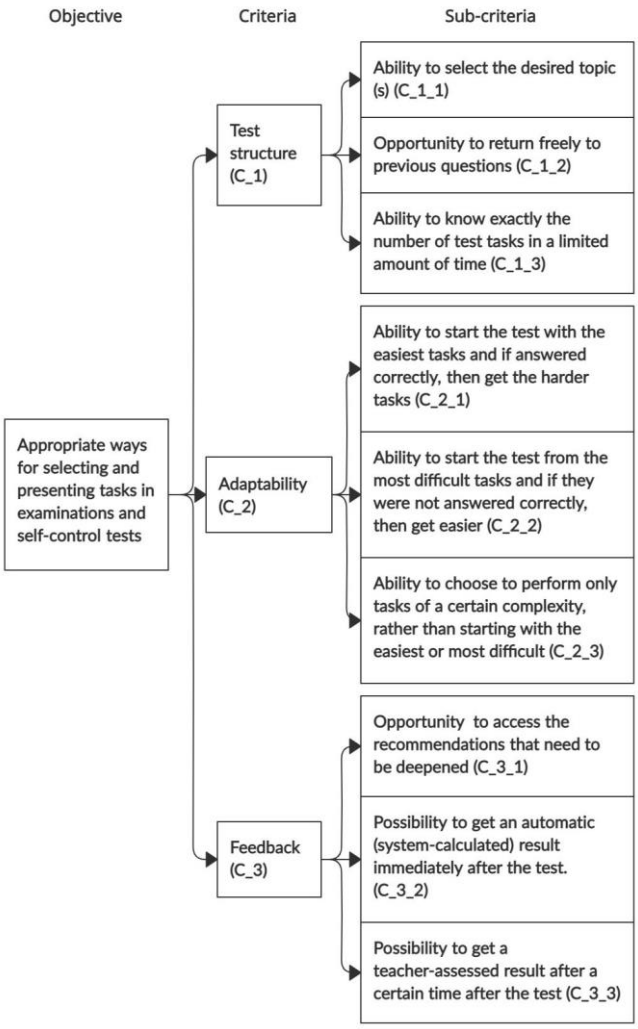


Fig. 1. Test criteria for selection and presentation of tasks tree.

The structure of the test is composed of the ability to know or even select a specific topic on which the test questions should be presented. As well, it is important to know whether students want to know in advance how many questions will be presented in the test and have a possibility to return to the previous tasks, questions.

Usually, when conducting knowledge assessment directly, the teacher asks questions of varying complexity. If a student does not answer a difficult question, he or she is asked an easier question from the same area. If a student answers an easy question, then he is presented with a more difficult one. This determines the student's level of knowledge in specific areas by adaptively selecting the directly selected questions. Therefore, the criteria of test adaptability are raised to determine how important this factor is for the student.

Test adaptability can have several options: start with the easiest task, start with the most difficult task, or allow the student to choose a task of a specific difficulty.

In addition to test adaptability and structure, feedback might be important in e-assessment as well. It is important to determine how important are different assessment methods to the students: get recommendations for further improvement, based on the test results; get the test results immediately after the test, evaluated in an automated way; get manual evaluation results after some time after the test is finished.

To determine which criteria are more important to students in ST and FK cases, a survey form was designed and distributed between students. Students were first asked to compare the importance of the criteria branches for FK and ST cases and then the sub-criteria by criteria (separate test structure sub-criteria, test adaptability sub-criteria, and feedback sub-criteria). By comparing the two criteria or sub-criteria, students were able to choose from the following statements: only A is important (value 9), A is more important than B (value 5), equally important (value 1), B is more important than A (value 1/5), only B is important (value 1/9). The system was selected based on AHP methodology and simplifying it to have 5 possible values only, rather than 17. This was done to assure the survey will not be too difficult for students to understand and answer.

At the same time, three alternatives were investigated: tree-based testing structure; graph-based testing structure; list-based testing structure. Currently, the majority of testing systems are using list-based testing structures, while the graph-based testing structure is gaining popularity but it is a slow process because of the need for additional test design and implementation requirements. The tree-based testing structure has similar application limitations as graph-based, therefore is not as popular as list-based testing structures too.

We assigned values for each of the three criteria based on whether the alternative supports appropriate sub-criteria (value 1) or not (value 0). The sum of appropriate sub-criteria value and importance coefficient (obtained from students' survey) products define the score for each alternative and allow its ranking.

The analysis has to be executed separately for FK and ST cases to see the differences between those two cases.

IV. Research Results

A. Students Survey Results

In the survey participated 25 students of higher education in the field of computer sciences and 49 students from last year's secondary school. Each of the students had to express his or her opinion on e-testing needs both for FK as well as ST cases. Based on students' answers the importance of each criterion was calculated. AHP usage allowed estimation of consistency ratio (CR). We take it into account and analyze consistent opinions only (CR<0.1). Summarized results of e-testing needs criteria coefficients are presented in Table I.

As seen from the result summary in table 1, the Higher education students have more different preferences for FK and ST cases. Meanwhile, secondary school students' opinions a very similar both for FK as well as ST case. This might be an indicator higher education students understand the purpose of each of those cases, while secondary school students do not see a bigger difference between final testing and self-evaluation testing.

To define the similarity between students' opinions, hierarchical clustering was used with Pearson's distance for

record similarity estimation. The data included all criteria coefficients, testing case code (value 0 for FK and value 1 for ST), and student type (value 0 for higher education and value 1 for secondary school students). The clustering results are presented in Fig. 2.

TABLE I: SUMMARY OF E-TESTING CRITERIA IMPORTANCE COEFFICIENTS

Criteria code	Higher education students		Secondary school students	
	FK case	ST case	FK case	ST case
C_1	0.40	0.30	0.40	0.40
C_1_1	0.24	0.46	0.38	0.43
C_1_2	0.40	0.29	0.35	0.32
C_1_3	0.36	0.25	0.27	0.24
C_2	0.33	0.32	0.32	0.31
C_2_1	0.41	0.29	0.38	0.38
C_2_2	0.20	0.22	0.25	0.25
C_2_3	0.39	0.49	0.38	0.37
C_3	0.27	0.38	0.28	0.29
C_3_1	0.34	0.50	0.40	0.42
C_3_2	0.38	0.27	0.27	0.27
C_3_3	0.29	0.23	0.33	0.30

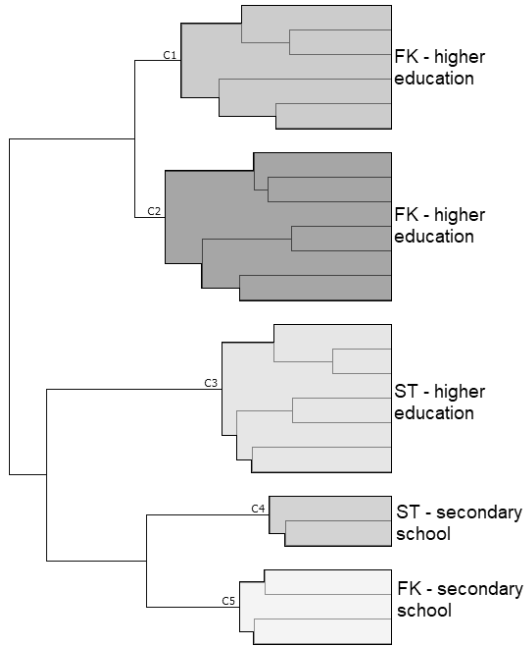


Fig. 2. Clustering results, based on students type and testing case.

As seen in Fig. 2, five clusters define a clear separation between cases and student types. It shows two different clusters are indicated for higher education students in FK case. This is mostly influenced by some variations in the field of C_2_3, C_2_1, and C_1_3 sub-criteria importance.

For further analysis, we will not go into details of all five clusters and will select the two main clusters: set of C1 and C2 clusters (Cl1) and set of C3-C5 clusters (Cl2). This is the main threshold for the clustering and usage of two clusters will simplify the further analysis.

B. Alternative Ranking Results

As alternatives three testing system structures were analyzed: linear, graph-based, tree-based. Each alternative was evaluated by assigning one of three possible values for each criterion. Value 1 is assigned if the criteria might be implemented in the alternative without editing the existing test or creating an additional test to fit the needs. Value 0.5 is assigned if the criteria can be met theoretically, by modifying the existing solution or creating multiple test variants for different situations. Value 0 is assigned if the criteria can not be met in the existing testing solution. The results of the alternative evaluation and calculated score for Cl1 and Cl2 cases are provided in Table II.

TABLE II: SUMMARY OF ALTERNATIVE EVALUATION AND SCORING

Criteria code	Coefficient		Testing structure types		
	Cl1	Cl2	Linear	Graph-based	Tree-based
C_1	0.40	0.38	-	-	-
C_1_1	0.24	0.42	0.5	0.5	1.0
C_1_2	0.40	0.33	1.0	0.0	0.0
C_1_3	0.36	0.26	1.0	0.5	0.0
C_2	0.33	0.32	-	-	-
C_2_1	0.41	0.36	0.5	0.5	1.0
C_2_2	0.20	0.24	0.5	0.5	1.0
C_2_3	0.39	0.40	0.5	0.5	1.0
C_3	0.27	0.30	-	-	-
C_3_1	0.34	0.43	0.0	1.0	1.0
C_3_2	0.38	0.28	1.0	1.0	1.0
C_3_3	0.29	0.29	1.0	0.0	0.0
		Cl1 score	0.70	0.48	0.62
		Cl2 score	0.63	0.50	0.69

Alternative evaluation results show the linear test structure meets the structure requirements the best, the highest adaptation scores are achieved by tree-based testing structure, while in the field of feedback each solution has different values, but the average score is the same – 2 out of 3.

By multiplying the criteria value and coefficient for each case, the final score for each case and testing structure type was calculated. The scores prove the linear testing structure has the highest score for Cl1 cases and the average score for both Cl1 and Cl2 cases. Meanwhile, a tree-based testing structure is very close to it and is the best solution for Cl2 case.

It is worth mentioning linear testing structure has a higher score for Cl1 case, while graph- and tree-based testing structures a better scored for Cl2 case. This is mostly influenced by the linear testing structure's ability to freely navigate between all test questions and constant question numbers in the test. At the same time, it limits or complicates the adaptability and feedback possibilities, which are better implemented by other testing structures.

V. CONCLUSION

The e-evaluation area is dependent on multiple factors. Various research papers exist on e-learning and e-evaluation, e-testing, however, it is lacking a deeper view on different e-testing structure suitability to meet user need for a different type of cases – final knowledge testing and self-evaluation testing. Therefore this paper gathers students' opinions on e-evaluation system requirements and scores three possible e-testing structures on their possibility to meet the needs.

The difference between final knowledge testing and self-evaluation testing is noticeable for higher education students. While secondary school students have a very similar opinion on final knowledge testing and self-evaluation testing. Therefore higher education students' opinion of final knowledge testing was analyzed as one area (Cl1) and all the rest cases as another area (Cl2). For higher

education students knowledge testing the test structure is the most important (importance coefficient is 0.40) while test adaptability and feedback have lower importance (respectively 0.33 and 0.27). Meanwhile, for higher education students' self-evaluation testing and all cases of secondary school students, the importance coefficients values are more similar (structure – 0.38, adaptability – 0.32, feedback – 0.30). This shows these two cases have similarities, however, the internal structure of these criteria varies.

Linear, graph-based, and tree-based testing system structures were scored as alternatives in this paper. The results revealed linear structure testing systems have the highest average ranting for both knowledge as well as self-evaluation testing (66%). Tree- and graph-based testing structures achieved on average 65% and 49% respectively. However average scores should not be the main criteria to choose from, as for higher education students' final knowledge testing the linear structure testing has the highest score (70%), while the tree-based testing structure has the highest score for rest analyzed case testing (69%).

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

AM conducted the research; SR analyzed the data; AM and SR wrote and approved the final version of the paper.

References

[1] I. Gil-Jaurena, "Learning assessment in open, distance and flexible education," *Open Praxis*, vol. 5, no. 3, pp. 195-196, 2013.

[2] M. Appiah and F. Tonder, "E-assessment in higher education: A review," *International Journal of Business Management and Economic Research(IJBMER)*, vol. 9, no. 6, pp. 1454-1460, December 2018

[3] M. Csikszentmihalyi, *Flow: The Psychology Of Optimal Experience*, 1st ed. USA: Harper Collins Publ., 2008.

[4] M. Csikszentmihalyi, *Finding Flow: The Psychology of Engagement with Everyday Life*, 1st ed. Basic Books, 1998.

[5] S. Jagadeesan and J. Subbiah, "Real-time personalization and recommendation in adaptive learning management system," *J Ambient Intell Human Comput*, vol. 11, pp. 4731–4741, February 2020.

[6] C. Athanasiadis, E. Hortal, D. Koutsoukos, C. Z. Lens, and S. Asteriadis, "Personalized, affect and performance-driven computer-based learning," in *Proc. the 9th International Conference on Computer Supported Education*, pp. 132-139, vol. 1, 2017.

[7] A. H. Nabizadeh, D. Gonçalve, S. Gama, J. Jorge, and H. N. Rafsanjani, "Adaptive learning path recommender approach using auxiliary learning objects" *Computers & Education*, vol. 147, April 2020.

[8] D. Hariyanto and T. Koehler, "An adaptive user interface for an e-learning system by accommodating learning style and initial knowledge," in *Proc. the International Conference on Technology and Vocational Teachers (ICTVT 2017)*, pp. 16-23, vol. 102, September 2017.

[9] M. A. Smith and J. D. Karpicke, "Retrieval practice with short-answer, multiple-choice, and hybrid tests," *Memory*, vol. 22, no. 7, pp. 784-802, September 2013.

[10] O. Bulgakov and A. Dedikova, "Testing educational achievements: From knowledge test to test of understanding," *Vestnik of the St. Petersburg University of the Ministry of Internal Affairs of Russia*, no. 2, pp. 183-190, July 2020.

[11] T. AlMahmoud, M. A. Elzubeir, S. Shaban, and F. Branicki, "An enhancement-focused framework for developing high quality single best answer multiple choice questions," *Education for Health*, vol. 28, no. 3, pp.194-200, September 2015.

[12] J. O. Nnodim, "Multiple-choice testing in anatomy," *Medical Education*, vol. 26, no. 4, pp. 301-309, July 1992.

[13] A. R. Thompson and P. R. Husmann, "Developing multiple-choice questions for anatomy examinations," in *Teaching Anatomy*, L. K. Chan, W. Pawlina, Eds. Cham: Springer, 2020, pp. 405-416.

[14] E. Hartell and G. J. Strimel, "What is it called and how does it work: examining content validity and item design of teacher-made tests," *Int J Technol Des Educ*, vol. 29, pp. 781–802, July 2019.

[15] E. Prado, P. A. Higuera, and H. O. Carvajal, "Design and validation of a questionnaire for self-assessment of university service-learning experiences, " *Educación XX1*, vol. 23, no. 1, pp. 319-347, 2020.

[16] R. Peytcheva-Forsyth and L. Aleksieva, "Forced introduction of e-assessment during COVID-19 pandemic: How did the students feel about that? (Sofia University case)," *AIP Conference Proceedings*, vol. 2333, March 2021.

[17] H. Cigdem, and S. Öncü, "E-assessment adaptation at a Military Vocational College: Student perceptions," *Eurasia Journal of Mathematics, Science and Technology Education*, vol. 11. pp. 971-988, August 2015.

[18] A. Nuha, G. Wills, and M. Waldm. "Advantages and challenges of using e-assessment," *International Journal of Information and Education Technology*, vol. 8, pp. 34-37, August 2018.

[19] C. Troussas, K. Chrysafiadi, and M. Virvou, "An intelligent adaptive fuzzy-based inference system for computer-assisted language learning," *Expert Systems with Applications*, vol. 127, pp. 85-96, March 2019.

[20] I. Mustakerov and D. Borissova, "A conceptual approach for development of educational Web-based e-testing system," *Expert Systems with Applications*, vol. 38, pp. 14060-14064, April 2011.

[21] M. Arif, N. Abdullah, M. Illahi, A. Karim, S. Shamshirband, K. Alam, S. Farid, Shahid S. Iqbal, Z. Buang, and V. Balas. "An architecture of agent-based multi-layer interactive e-learning and e-testing platform," *Qual Quant*, vol. 49, pp. 2435–2458, October 2015.

[22] S. Ramanauskaitė and A. Slotkiene, "Hierarchy-based competency structure and its application in e-evaluation," *Applied Sciences*, vol. 9, no. 17, August 2019.

[23] A. Slotkiene, "Context aware information model for active learning object," *Baltic Journal of Modern Computing*, vol. 6. no. 2, pp. 155-161, June 2018.

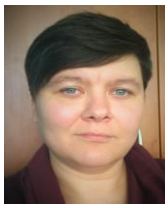
[24] P. Brézillon, "Task-realization models in contextual graphs," *Modeling and Using Context*, vol. 3554, Berlin: Springer, pp. 55-68.

[25] N. Aukstakalnis, K. Baniulis, J. Pauliute, and A. Slotkiene, "Graphical model: The means for simulation-based learning," in *Proc. ITI 2008 - 30th International Conference on Information Technology Interfaces*, vol. 16, 2008, pp. 471-476.

[26] E. H. Forman and S. I. Gass, "The analytic hierarchy process-an exposition," *Operations Research*, vol. 49, no. 4, pp. 469-486, August 2001.

Copyright © 2022 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited (CC BY 4.0).

Asta Margiene was born in Lithuania on June 19, 1977. She received the M.Sc. in computer science in 2006 from Šiauliai University. Now she is a PhD student of informatics engineering at Vilnius Gediminas Technical University. PhD thesis is about the architecture of adaptive knowledge assessment systems. Currently, she is a lecturer at Vilnius Gediminas Technical University and at Vilnius University Šiauliai Academy.



Simona Ramanauskaite was born on December 18, 1983. She received her M.Sc. in informatics in 2008 and PhD in informatics engineering in 2012 from Vilnius Gediminas Technical University. Currently she is working as a professor at Vilnius Gediminas Technical University, Department of Information Technologies. Her current research interests include different aspects of information security, human-computer interaction and e-learning.