Multi-Criteria Model for Questions Selection in Generating e-Education Tests Involving Gamification

Daniela Borissova [1,2], Delyan Keremedchiev [1,3], Georgi Tuparov [3]

[1] *Institute of Information and Communication Technologies at Bulgarian Academy of Sciences, Bulgaria*
[2] *University of Library Studies and Information Technologies, Sofia, Bulgaria*
[3] *New Bulgarian University, Sofia, Bulgaria*

***Abstract* – In traditional and e-learning the tests play an important role to track the learners' progress. In this regard, the paper deals with the generation of test with different levels of complexity directly associated with particular evaluation. An algorithm for the generation of tests with different levels of complexity using multi-criteria optimization is proposed. It is important to note that all test items could be considered as an element of a level in an educational game or gamified environment, and their weight influences the complexity of the game. The results show that the proposed model allows the generation of tests not only with different levels of complexity, but it also provides additional flexibility in respect of the selected number of questions and their weight. The proposed algorithm could be realized as a separate module for e-testing, or it could be integrated in the learning management system.**

***Keywords* – E-learning, e-test, gamification, educational computer game, multi-criteria model.**

1. Introduction

Tests as a tool of verification and evaluation are often used not only in traditional learning but also in the various forms of e-learning.

DOI: 10.18421/TEM92-47
https://doi.org/10.18421/TEM92-47

Corresponding author: Daniela Borissova,
Institute of Information and Communication Technologies at Bulgarian Academy of Sciences, Bulgaria
Email: dborissova@iit.bas.bg

Received: 29 March 2020.
Revised: 25 April 2020.
Accepted: 04 May 2020.
Published: 27 May 2020.

Online tests should be used strategically to facilitate students' engagement. The involvement of modern ICT in academic and education has to ensure a smooth and successful transition towards technological enhancement [1], [2]. The literature review of online tests indicates that online tests can be used for different purposes when they are appropriate and optimized for assessment [3]. Contemporary development of ICT is the major circumstance that affects many different areas including education of STEM (science, technology, engineering, and mathematics) [4], [5]. For students engaged in engineering disciplines additional training is needed in order to be able to understand the system work regimes [6].

Special attention is needed to improve computer science education especially in the cases when learning functions and algorithms are concerned, since they can be hard to understand [7]. Along with functions and algorithms understanding, the critical issues in computer science are the programming languages. Computer programming skills is the core of competencies among the students involved in computer science, and many efforts have to be done in developing tools to help students to test programming skills. For this purpose, an educational environment to support students in problem-solving activities that are common for programming problems is developed [8]. The interest of students in engineering, computer science and natural sciences is decreasing because students meet difficulties with subject contents. As a result the percentage of withdrawals from such courses increases [9].

All of the above motivate us to propose a multi-criteria optimization mathematical model for the generation of tests with different degree of complexity that corresponds to different assessment or different game levels. This model can generate tests with less but heavily-weighted questions, or it can generate more but light-weighted questions. This feature can be selected by learners, but in both cases the time for test performing are with the same durations. The model is appropriate for intelligent selection of items in quest-based games or in gamification environment. The educational computer games are widely used in school or university courses and the relation to a specific learning objective is considered.

2. Literature Review

The use of different quizzes allows learners' knowledge testing, so that it is possible to monitor the learners' progress during teaching. In this regard a variety of testing tools are proposed to verify the acquired knowledge of learners in response of different learning needs. For example, the developed general web-based e-testing knowledge system allows flexible adjusting of the testing process [10]. Such systems can be used in two regimes – self-testing and official examination. Other software systems are realized as online learning tools using "Assessment as a Service" on a cloud with SOA architecture. They have the ability for adaptive testing with usage of software agents where different strategies define the agents' behavior [11]. The developed computer-based test system for schools in Indonesia includes an online exam system suitable for classroom needs, and it allows arranging the questions according to courses and competencies [12]. In the context of learning, an approach for better understanding of multi-criteria optimization is proposed by using a web-based decision support system WebOptim [13]. Depending on the topic of the learning course different supporting tools can be applied to assist the specifics of the learning materials [14]. In addition, some authors propose cooperative learning models in information systems project management courses that result in a learning model in the form of a web-based student assignment application [15]. In order to improve the effectiveness of the learning process, architecture of agent-based multi-layer interactive e-learning and e-testing platform is proposed [16].

It is shown that quizzes can enhance learning and performance, and they lead to improvements in student satisfaction [17]. Feedback provides detailed explanations of correct and incorrect options and supports learning. The usage of Classroom Response Systems to measure the students learning outcomes shows improvement in the dynamics of the class in theoretical lectures [18]. Periodically testing over the year provides students with added confidence about the gained knowledge. For this purpose, different methods can be utilized to cope with the assessment process. For example, evaluation of students can be based on different data mining classification algorithms to predict student performance [19].

Multi-objective optimization could contribute for more flexibility when generating the tests. An efficient integer linear programming approach to high-quality static test generation from large-scale question data sets is proposed [20]. It is shown that branch-and-cut for static test generation approach is useful for web-based testing and assessment for online learning environment. Other approaches based on multi-objective optimization using heuristic-based methods rely on genetic algorithm [21], or particle swarm optimization [22], ant colony optimization [23], specialized ontologies [24], etc.

Recently, gamification and educational computer games have been widely used in open source learning management systems as a part of assessment [25]. According to a number of studies the application of computer game-based learning and gamification increases students' motivation and leads to easier perception of learning content and engagement in the educational process [26]. Important elements of the quest-based educational computer games are such game elements as test items and learning situations related to particular learning content, collection of points or badges, scoreboards, and moving from easy to difficult levels of evaluation according to students' achievements [27], [28]. It is shown that games offer a low-budget alternative instructional strategy, and they could easily be applied in the classroom during a lecture. They also contribute involving additional instructional strategies [29]. Developing environments for testing the programming abilities requires specific approaches including game-based learning for the development of algorithmic thinking [30].

Although tests are widely used means for assessing the acquired knowledge, there is no information if any models are applied for test generation. The researches show that e-testing systems can take into account questions of different type, weight where the obtained result is presented in percentages and the sum of points for the correct answers.

That is why a mathematical modeling approach seems to be reasonable in generation of tests with different degrees of complexity as a part of quest based game design. For this purpose, an algorithm for test generation in which tests with different complexity levels are generated by using multi-criteria optimization is proposed. This approach could be applied for generation of test items and tasks as a part of an educational computer game.

3. Algorithm for Generating Tests with Different Complexity Levels Using Multi-Criteria Optimization

The proposed algorithm for generating tests with different complexity levels using multi-criteria optimization is illustrated in Fig. 1.

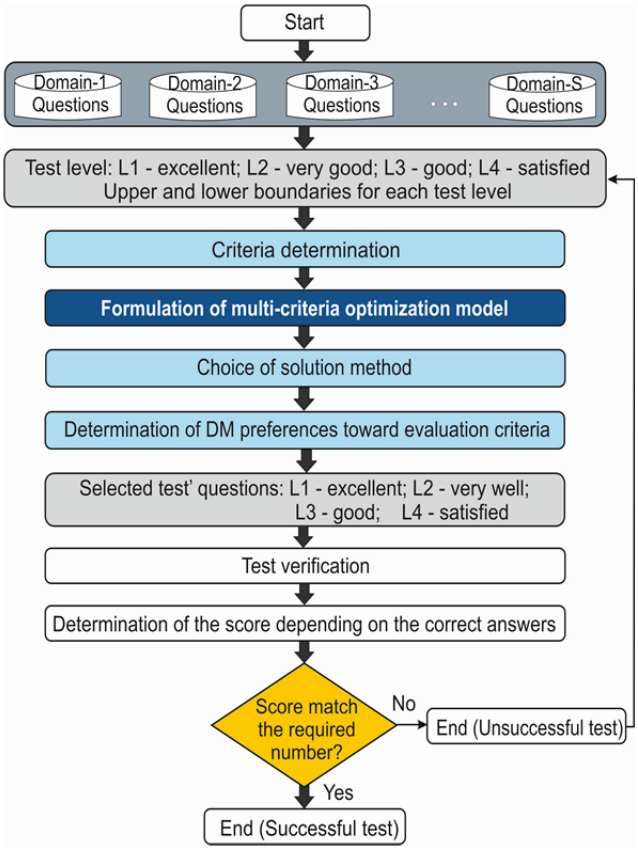

Figure 1. Algorithm for generating of tests with different levels of complexity by using multi-criteria optimization

The algorithm starts with determination of the domain for which the tests will be generated. The domain database can hold learning materials for different courses, or it can be composed of learning materials for a particular course with a set of topics organized in sub-domains. Once the domain is known, a database of questions is needed. A proper number of points correspond to each question from the database. The upper and lower level of test score is determined depending on the sum of points of all questions. Next, the evaluation criteria are to be established and the corresponding multi-criteria optimization problem is to be formulated. Then, the proper solution method has to be selected, and depending on this selection the decision maker preferences toward the evaluation criteria are to be determined.

When all of these input data are known, the corresponding multi-criteria optimization task can be solved. This will determine the number of questions for the generated test together with the particular test level. Test verification and determination are the following items of score depending on the given correct answers. Depending on the obtained score using the given upper and lower boundaries (for particular test level) the algorithm ends with a message for successful test pass or failure.

For the purpose of self-testing, in case of test failure the algorithm allows the user to select another test level and generate another test.

4. Multi - Criteria Model for Selection of Questions for Generating Tests with Different Complexity Levels

To provide the flexibility of the generated tests two main criteria are used to compose the multi-criteria model. These two criteria aim to maximize the test points (L), and minimize the number of questions (Q) that will compose the generated tests. The formulated model is as follows:

$$\begin{cases} max\left(L^k = \sum_{i=1}^{M} x_i d_i\right) \\ min\left(Q^k = \sum_{i=1}^{M} x_i\right) \end{cases} \quad (1)$$

subject to

$$L^k \leq L^k_{upper} \quad (2)$$
$$L^k_{lower} \leq L^k \quad (3)$$
$$L^k_{upper} < D \quad (4)$$
$$\sum_{i=1}^{M} d_i = D \quad (5)$$
$$x_i \in \{0,1\} \quad (6)$$

The objective function (1) simultaneously maximizes the test points (L^k) and minimizes the number of questions Q^k that will compose the test. The restrictions (2) and (3) provide satisfying the required points according to the given level of test complexity. The restriction (4) does not allow the use of a limit for the score of the generated test greater than the total number of all questions' score from which the choice is made. The choice of a particular question is realized by using binary decision variables xi determined by (6).

The weighted sum method is a scalarising approach in which the multi-criteria optimization task is transformed into a single-criterion task. It requires normalization to obtain comparable values in the same range [31]. Taking into account the formulated above model and using the weighted sum method the transformed single-criterion task is as follows:

$$max\left(w_1\left(\frac{L^{max}-L}{L^{max}-L^{min}}\right)^k + w_2\left(\frac{Q-Q^{min}}{Q^{max}-Q^{min}}\right)^k\right) \quad (7)$$

Subject to (2) – (6) and additional restriction about the coefficients concerning the criteria importance:

$$w_1 + w_2 = 1 \quad (8)$$

where w_1 and w_2 are the weighted coefficients representing the importance of the criteria from the decision maker's point of view; L^{max}, L^{min} and Q^{max}, Q^{min} are the upper and lower boundaries for the test points and the number of questions obtained as a result of solving the corresponding single criterion tasks separately.

5. Numerical Application

To test the applicability of the proposed multi-criteria model (1) – (6), the weighted sum method for determination of questions for tests with different levels of complexity is used. The domain area of the generated tests is based on the learning material for the Web programming course. Four different levels of tests complexity are used to distinguish the level of acquired knowledge. Level L^1 corresponds to the excellent performance; level L^2 means very good, level L^3 is for good performance and level L^4 express satisfactory results. These four levels are represented by the use of lower and upper boundaries for each level. These lower and upper boundaries for each level depend on the predefined numbers of questions from which the tests will be generated.

The numerical testing is accomplished using a total of 40 questions with different levels of weight given in advance. These questions belong to different types of assessment items – multiple choices, matching, short answer, essay, error finding, programming code writing, reading of programming code and description of the results of a given code.

Some examples of items in gamified assessment with different weights for introductory programming course are listed below.

Example 1 (question with 3-points weight): Someone mismatched cards with loop statements and their names/semantics. Help to match the statement with the correct names/semantics. Connect statement with its name/semantics.

for (int i=0; i<10; i++)
A) Loop with post-condition
while B do S
B) Loop with counter
do S while B
C) Loop with pre-condition

Example 2 (question with 5-points weight): John wrote the code below. Write the result from the code execution. What is the name of the method?

```
const int n=5;
int a[n]={10,20,8,3,1}, temp, i, j;
for (i=0; i<n-1; i++)
for (j=0; j<n-i-1; j++)
if (a[j]>a[j+1])
{
temp=a[j];
a[j]=a[j+1];
   a[j+1]=temp;
}
for (int k=0; k<n; k++)
cout<<"a["<<k<<"]="<<a[k]<<endl;
```

Example 3 (with 8-points weight): Ann has to write a code to calculate the positive points obtained in a test with 20 items. The points for each item could be an integer number from – 3 to 3. She wrote the code below. Find the errors and write a correct code.

```
double a[20];
int s=0.0;
.....
for (int i=0; i<20; i--)
if (a[i]<=0) s=s+a[I];
```

Example 4 (question with 10-points weight): In a computer game every player has an ID number from 1 to 20 and obtained points. The average points of all participants should be written on the game's leaderboard, as well as the points and ID numbers of the players that have obtained points higher than the average points. Write a code to solve the task.

The overall sum of the question scores is equal to 270 and the corresponding upper and lower boundaries for each level are shown in Table 1.

Table 1. Upper and lower boundaries for each level

Test boundaries	Test complexity level			
	Excellent (L^1)	**Very Good (L^2)**	**Good (L^3)**	**Satisfactory (L^4)**
Upper	200	189	175	159
Lower	190	176	160	140

The transformed optimization task with objective function (7) subject to (2) – (6), the additional restriction (8), the use of upper and lower boundaries from Table 1 in case of generation high complexity test corresponding to the excellent evaluation (L^1) and the equal importance of the criteria ($w_1 = w_2 = 0.5$) is as follows:

$$max\left(0.5\left(\frac{200-L^1}{200-190}\right)+0.5\left(\frac{Q-23}{32-23}\right)\right) \quad (9)$$

subject to

$$L^1 = \sum_{i=1} x_i d_i \quad (10)$$

$$\text{Q} = \sum_{i=1} x_i \quad (11)$$

$$177 \leq L^1 \leq 190 \quad (12)$$

$$200 = L^1_{upper} < 270 \quad (13)$$

$$x_i \in \{0,1\} \quad (14)$$

The variables in the formulated problem (9) – (14) are the scores for test level (L^1), the number of questions Q that will compose the test, and the binary integer variables x_i that make possible to determine which question is selected for a particular test. Three additional similar tasks are formulated and solved for the next three levels of complexity corresponding to the level L^2 for very good, level L^3 for good and level L^4 for satisfactory evaluation. The obtained results for the parameters of the generated tests with equal importance of both criteria (test score and number of questions within test) are shown in Table 2.

Table 2. Results for the generated tests with different level of complexity

Level of test complexity	**Test score (L^k)**	**Number of questions for test (Q^k)**
	$w_1 = 0.5$	$w_2 = 0.5$
L^1 – excellent	190	31
L^2 – very good	176	30
L^3 – good	160	28
L^4 – satisfactory	140	25

The obtained results show that the generated tests satisfy the upper and lower boundaries given in Table 1. The test scores in all four cases are at the lower boundaries – 190 for level L^1, 176 for level L^2, 160 for level L^3 and 140 for level L^4.

6. Results and Analysis

The graphical representation of the results from solving the formulated problem (9) – (14) with four different test complexity levels are shown in Fig. 2.

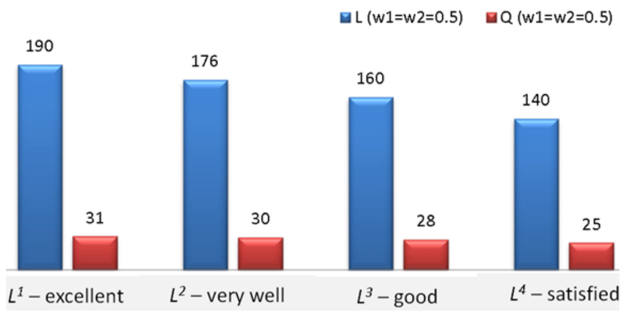

Figure 2. Parameters of generated tests in different cases

The use of the weighted sum method for solving multi-criteria optimization tasks allows the usage of different combinations to express the criteria importance satisfying the condition (8). In this relation two additional series of optimization tasks are solved for two opposite scenarios. In the first scenario the dominant criterion is the test score using: $w_1 = 1$ and $w_2 = 0$. In the second scenario the dominant role is taken by the criterion for question number using: $w_1 = 0$ and $w_2 = 1$. The parameters of the generated tests for three different cases are compared in Fig. 3.

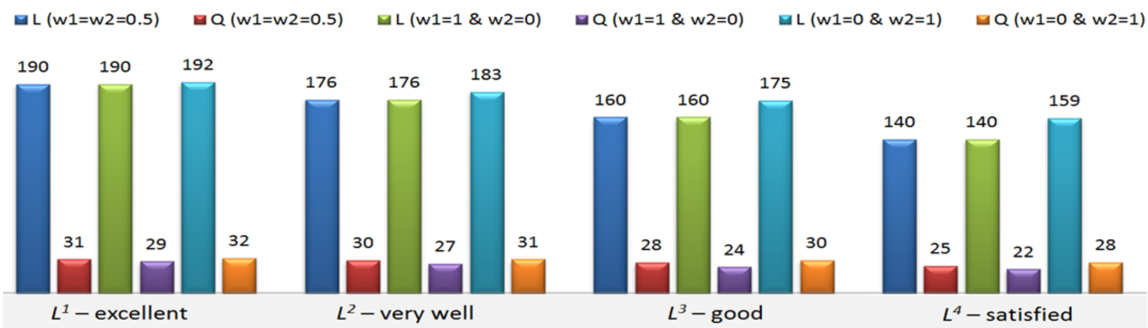

Figure 3. Parameters of generated tests under three different cases

The proposed multi-criteria model (1) – (6) using the weighted sum method is applicable to generate tests with different levels of complexity as could be seen from Fig. 3. Furthermore, by setting the user preferences using weighted coefficients (8) it is possible to provide the tests with less but heavily weighted questions and vice versa for the same level of test complexity. In case of level L^1 test generation the number of questions varies within the range of 29 to 31, and the points for this level are close to the given lower boundary of 190. When tests for level L^4 are generated, the solutions determine a comparatively large range from 22 to 28 for the number of questions and the points for this level are on the given upper and lower boundaries.

7. Conclusion

The current paper deals with the problem of test generation using multi-criteria optimization. For this purpose, an algorithm based on the proposed multi-criteria optimization model is described. The proposed model can be easily set for generation of tests with different levels of complexity. The scalarizing techniques of the weighted sum method allow explicitly setting the preferences toward the number, type of questions and testing points according to the selected level of test complexity.

The essential part of the proposed algorithm is the type of used questions that are suitable for gamification in learning management systems as a part of assessment or in quest-based educational computer games. It is shown that the proposed multi-criteria model for questions selection for generation of tests with different complexity levels does not depend on the used question type, and it can be applied in different educational subjects. It is possible to use other methods for solving the multi-criteria model (1) – (6). According to the selected method different questions can be set taking into account the particular weights of the questions.

The numerical testing of the proposed algorithm and multi-criteria optimization model for test generation is tested using a database with a limited number of questions. The obtained results show the practical applicability for both algorithm and formulated bi-criteria optimization model.

The proposed algorithm can be implemented as a separate module for e-testing or educational computer game with different weight levels, based on the question items. It can be integrated as part of the learning management system. This feature is planned as future development of the proposed multi-criteria combinatorial model for determination of questions in the process of test generation.

Acknowledgement

The research in this article is supported by the project "Pedagogical and technological aspects of educational computer games", DN-05/10, 2016, funded by Bulgarian National Scientific Fund.

References

[1]. Khan, S., & Khan, R. A. (2019). Online assessments: Exploring perspectives of university students. *Education and Information Technologies*, *24*(1), 661-677.

[2]. Shalamanov, V. (2019). Organizing for IT Effectiveness, Efficiency and Cyber Resilience in the Academic Sector: National and Regional Dimensions. *Information & Security: An International Journal*, *42*.

[3]. Boitshwarelo, B., Reedy, A. K., & Billany, T. (2017). Envisioning the use of online tests in assessing twenty-first century learning: a literature review. *Research and Practice in Technology Enhanced Learning*, *12*(1), 16. DOI: 10.1186/s41039-017-0055-7.

[4]. Terzieva, V., Paunova-Hubenova, E., Dimitrov, S., & Boneva, Y. (2018, September). ICT in STEM Education in Bulgaria. In *International Conference on Interactive Collaborative Learning* (pp. 801-812). Springer, Cham.

[5]. Atanasijevic-Kunc, M., Logar, V., Pogacnik, M., & Guliashki, V. (2018). Combined approach to modelling and simulation in engineering education. In: *LIII International Scientific Conference on Information, Communication and Energy Systems and Technologies ICEST'2018*, 51-54.

[6]. Borissova, D., & Mustakerov, I. (2009, August). A framework of multimedia e-learning design for engineering training. In *International Conference on Web-Based Learning* (pp. 88-97). Springer, Berlin, Heidelberg.

[7]. Jannah, U. R., Nusantara, T., Yulianto, F. E., & Amiruddin, M. (2019). Student's Learning Obstacles on Mathematical Understanding of a Function: A Case Study in Indonesia Higher Education. *TEM Journal*, *8*(4), 1409-1417.

[8]. Areias, C., & Mendes, A. (2007, June). A tool to help students to develop programming skills. In *Proceedings of the 2007 international conference on Computer systems and technologies* (pp. 1-7), DOI: 10.1145/1330598.1330692.

[9]. Ibanez, M. B., Di-Serio, A., & Delgado-Kloos, C. (2014). Gamification for engaging computer science students in learning activities: A case study. *IEEE Transactions on learning technologies*, *7*(3), 291-301.

[10]. Purković, D., & Kolumbić, Z. (2012, May). A concept of the web-based e-testing system. In *2012 Proceedings of the 35th International Convention MIPRO* (pp. 1328-1334). IEEE.

[11]. Gusev, M., & Armenski, G. (2014). E-assessment systems and online learning with adaptive testing. In *E-Learning Paradigms and Applications* (pp. 229-249). Springer, Berlin, Heidelberg.

[12]. Nugroho, S., Budiyanto, C., & Budianto, A. (2018). Technology Acceptance Model for Examining the Feasibility of a Computer-based Test System (an Indonesian Context). *Indonesian Journal of Informatics Education*, *2*(2), 105-116.

[13]. Guliashki, V., Genova, K., & Kirilov, L. (2013). The Decision Support System WebOptim in an E-Learning Context. In *Proc. of International Conference" Automatics and Informatics* (Vol. 2013).

[14]. Marinova, G., Guliashki, V., & Chikov, O. (2014). Concept of Online Assisted Platform for Technologies and Management in Communications–OPTIMEK. *International Journal of Business and Technology*, *3*(1), 14-21.

[15]. Ningsih, S. R., Effendi, Z. M., & Syah, N. (2019). Implementation of Cooperative Learning Model on E-Assignment Responsiveness at Higher Education. *International Journal of Emerging Technologies in Learning (iJET)*, *14*(18), 209-219.

[16]. Arif, M., Illahi, M., Karim, A., Shamshirband, S., Alam, K. A., Farid, S., ... & Balas, V. E. (2015). An architecture of agent-based multi-layer interactive e-learning and e-testing platform. *Quality & Quantity*, *49*(6), 2435-2458.

[17]. Hennig, S., Staatz, C. E., Bond, J. A., Leung, D., & Singleton, J. (2019). Quizzing for success: Evaluation of the impact of feedback quizzes on the experiences and academic performance of undergraduate students in two clinical pharmacokinetics courses. *Currents in Pharmacy Teaching and Learning*, *11*(7), 742-749.

[18]. Saleh, H., Nurdyansyah, N., Hasanah, F. N., & Rudyanto, H. E. (2019). Application of Classroom Response Systems (CRS): Study to Measure Student Learning Outcome. *International Journal of Emerging Technologies in Learning (iJET)*, *14*(14), 132-142.

[19]. Kabakchieva, D. (2013). Predicting student performance by using data mining methods for classification. *Cybernetics and information technologies*, *13*(1), 61-72.

[20]. Nguyen, M. L., Hui, S. C., & Fong, A. C. (2012). Large-scale multiobjective static test generation for web-based testing with integer programming. *IEEE Transactions on Learning Technologies*, *6*(1), 46-59.

[21]. Hwang, G. J., Lin, B. M., Tseng, H. H., & Lin, T. L. (2005). On the development of a computer-assisted testing system with genetic test sheet-generating approach. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *35*(4), 590-594.

[22]. Ho, T. F., Yin, P. Y., Hwang, G. J., Shyu, S. J., & Yean, Y. N. (2009). Multi-objective parallel test-sheet composition using enhanced particle swarm optimization. *Journal of Educational Technology & Society*, *12*(4), 193-206.

[23]. Hu, X. M., Zhang, J., Chung, H. S. H., Liu, O., & Xiao, J. (2009). An intelligent testing system embedded with an ant-colony-optimization-based test composition method. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *39*(6), 659-669.

[24]. Stancheva, N., Stoyanova-Doycheva, A., Stoyanov, S., Popchev, I., & Ivanova, V. (2017). An environment for automatic test generation. *Cybernetics and Information Technologies, 17*(2), 183-196.

[25]. Hallifax, S., Serna, A., Marty, J. C., & Lavoué, É. (2019, September). Adaptive Gamification in Education: A Literature Review of Current Trends and Developments. In *European Conference on Technology Enhanced Learning* (pp. 294-307). Springer, Cham. https://doi.org/10.1007/978-3-030-29736-7_22

[26]. Mathew, R., Malik, S. I., & Tawafak, R. M. (2019). Teaching Problem Solving Skills using an Educational Game in a Computer Programming Course. *Informatics in Education*, *18*(2), 359-373.

[27]. Stoyanova, M., Tuparova, D., & Samardzhiev, K. (2017, September). Impact of Motivation, Gamification and Learning Style on Students' Interest in Maths Classes–A Study in 11 High School Grade. In *International Conference on Interactive Collaborative Learning* (pp. 133-142). Springer, Cham.

[28]. Ivanova, G., Kozov, V., & Zlatarov, P. (2019, May). Gamification in Software Engineering Education. In *2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)* (pp. 1445-1450). IEEE. https://doi.org/10.23919/MIPRO.2019.8757200.

[29]. von Wangenheim, C. G., Savi, R., & Borgatto, A. F. (2012). DELIVER!–An educational game for teaching Earned Value Management in computing courses. *Information and software Technology*, *54*(3), 286-298.

[30]. Tuparova, D. (2019). Possibilities for development of algorithmic thinking through game based learning and unplugged activities in primary school. *Computer Science and Education in Computer Science*, *15*(1), 80-84.

[31]. Genova, K., Kirilov, L., & Guljashki, V. (2013). New Reference-Neighbourhood Scalarization Problem for Multiobjective Integer Programming. *Cybernetics and Information Technologies*, *13*(1), 104-114.