Using Multi-criteria Evaluation of E-learning System: A Methodology Based on Learning Outcomes

Randa Aljably[1(✉)] and Salah Hammami[2]

[1] Computer Department, College of Science and Human Studies, Shaqra University, Shaqraa, Kingdom of Saudi Arabia
raljebly@su.edu.sa

[2] Computer Science Department, King Saud University, Riyadh, Kingdom of Saudi Arabia
shammami@ksu.edu.sa

Abstract. Academic research on E-learning system has been widely influenced by COVID-19 pandemic. As evaluating online learning can directly promote the design of more effective learning environments. As evaluating E-learning systems while considering factors such as system content, pedagogical issues, adaptation, learning agents and learning outcomes dimensions can be considered as a complex multi-criteria decision making problem. This paper presents a methodology to support researchers and practitioners in a effectively adopting analytical techniques such as Analytic hierarchy process, Entropy Method, and the Criteria Importance through Inter-criteria Correlation in E-learning evaluation. The paper experiments with multiple learning scenarios and the results provide various alternatives for decision making.

1 Introduction

When the education in the world was affected by the COVID-19 pandemic. The shift to digital environments was the solution to continue the teaching and learning processes. The transition was carried out by complete dependency through the learning management system (LMS), blended E-Learning or Traditional blackboard Teaching-Learning with E-Learning [1].

However, these systems have yet to prove their success in engaging users a quality learning process by providing necessary information in a timely and effective way [2, 3] with respect to pedagogical, learning and adaptation dimensions [4]. To address the evaluation necessity of E-learning systems, this current study attempts to present a systematic and analytical approach to evaluation criteria using analytic hierarchy process (AHP) [5, 6], entropy method (EM) [7], the criteria importance through inter-criteria correlation (CRITIC) [8] and simple derivation weighting (SDW) [9, 10] on the proposed evaluation criteria. The present research is undertaken to accomplish two major objectives, namely:

1. To develop and implement an evaluation model that enables making decisions based on indicators with respect to pedagogical, adaptive, student and course learning outcomes dimensions in a multiple criteria perspective.

Y. Tian et al. (Eds.): ICBDS 2021, CCIS 1563, pp. 563–574, 2022.
https://doi.org/10.1007/978-981-19-0852-1_44

2. To generalize the proposed model to be applied under three different scenarios reflecting worse and better conditions of system functionality, student behavior and learning outcomes (LO) to evaluate the impacts of proposed alternatives under different conditions.

This paper is organized as follows. Section 2 presents the previous research in literature on evaluating E-Learning. Section 3 discuss the presented methodology, Next, Sect. 4 experimental setup and results, while a discussion of the findings are presented in Sect. 5. Finally, Sect. 6 concludes the present study.

2 State of the Art of Evaluating E-learning

E-learning systems are multidisciplinary by nature. Many researchers from fields such as computer science, information systems, psychology, education, and educational technology, have been trying to evaluate E-learning systems. There has been a good recognition for the evaluation process even if it's still in its early stages and has dealt only with partial components, each evaluation project or research had its own justification and reasoning. The main categories of the literature are:

Learning during COVID-19 has influence many experts to evaluate the performance of E-learning tools and websites. In [11] the MCDM integration with linguistic hesitant fuzzy sets to evaluate and select the best E-learning website for network teaching. The results indicate that MCDM can be practical and effective when used for website selection under vague and uncertain linguistic environment.

The researchers in [5] employed the analytic hierarchy process (AHP) with group decision-making (GDM) and Fuzzy AHP (FAHP) to study the diversified factors from different dimensions of the web-based E-learning system. The MCDM approach produced realistic results in categorizing each dimension and critical success factors of the E-learning system. This categorization was intended as a decision tool for stakeholders' resources.

On the other hand, when MCDM was used in [12] for problem structuring, it was brought down by weak inter- and strong intra- connection between criteria. And highly affected by criteria definition and the number of criteria in each cluster (cluster size). Another factor that plays a great role in using MCDM is scale selection [13]. In statistical methods and AHP ranking of indicators using a scale of rating [0 to 100], where the higher the rating indicates the higher the performance of the approach under the evaluation criteria.

3 Proposed Methodology

The methodology starts with defining system evaluation criteria and generating alternatives that relate system capabilities to goal. It then follows with evaluating all alternatives in terms of criteria. After applying multi-criteria analysis method. A selection is made on one alternative as optimal. This process iterates until reaching better multi-criteria optimization. So to start with the first step in the research experts and decision makers definition of evaluation criteria we selected the list described in Table 1.

Table 1. The hierarchal structure for evaluation

Dimensions	Level	Criteria
Student Learning Outcomes	Knowledge	C01.Ease of knowledge appliance.
	Thinking skills	C02.Capability of problem definition and analysis.
	Psychomotor	C03.Capability of conducting computer based system design and implementation
		C04.Ease of functioning within teams.
	Community Relationships / Ethics	C05.Perception of ethical and legal responsibilities.
	Generic Learning skills	C06. Ease of communication with teachers.
	Subject content	C07.Undersandment of computing impact.
		C08.Recognition for continuous professional development importance.
		C09. Ability to apply technologies in new context.
Course Learning outcome		C10.Usage of testing strategies.
		C11.Capability of producing soft ware requirements.
System content		C12. Providence of appropriate learning material.
		C13.Suffecient content.
		C14.Latest content.
Pedagogical issues		C15.Fleaxibility of learning.
		C16.Providance of experience to learner.
		C17.Connection between knowledge and realty.
Adaptation	Learning Styles / Motivation	C18.Capability of recording learner's performance.
		C19.Capability of providing gaudiness to learner.
		C20.Capability of controlling learner's progress.
System agents		C21.Privacy of usage and performance.
		C22.Intelctual response to search / Recommending Agents.

After defining the criteria, we represent the criteria as scenario indicators which are obtained either from computer simulation models or model results. The model outputs the priority and performance of alternatives under proposed criteria scenario. This study delineates the best alternative on the basis of three different multi-criteria decision making (MCDM) methods, including simple additive weighting (SAW) [9] and compromise programming (CP) [14]. Each method is also applied with a set of criteria weights that represent objective judgments represented by expert's reviews as well as subjective preferences of decision makers.

As for the methodology for calculating criteria values, we suggest that each criterion is specified with a range of values starting from a lower threshold to an upper threshold and measured in a finite amount of time. In this paper the range of the criteria is selected between [0.0–2.0] representing the lower and upper thresholds respectively, in certain conditions these values may be fixed to a specific number to serve better judgment. As for criteria weights, we use AHP [5, 6] as an elicitation technique as one way, the other involves assigning performance scores to each alternative in performance matrix.

In order to derive the preference structure and performance matrix we used analytic hierarchy process (AHP) [5, 6], entropy method (EM) [7], the criteria importance through inter-criteria correlation (CRITIC) [8] and simple additive weighting (SAW)[9]. We choose AHP because of its superiority in performing judgment and scales over others [14]. Using the performance matrix (PM) where the columns correspond to criteria (C1, C2.., Cm) and rows correspond to alternatives (A1, A2, … An), with the entries (a_{ij}) being the indicators for all alternatives across all criteria. Once the matrix is set up, the next step for the decision process is to define the weights (w_1, w_2, …,w_m) of the criteria. The value of each alternative or solution will be calculated according to the value n Eq. (1), where A_i represents the suggested alternative or in our case system or solution and where A_{ij}, d_{ik} denotes the performance value of the i^{th} system under criteria j and dimension k respectively.

$$A_i = \sum\nolimits_{k=1}^{p} d_{ik} = \sum\nolimits_{j=1}^{n} A_{ij} W_j \tag{1}$$

When defining weights of Criteria's, these weights in MCDM do not have a clear economic significance, but their use provides the opportunity to model the preference structure. That is why they are assigned by decision makers (DM) using importance weights which represent the relative importance of criteria [15].

In evaluating the current system against the criteria; we will use rating-based and ranking-based methods [16]. Rating-based involves rating an alternative under each dimension then calculating the sum of its weighted performance under each criterion to give it an overall weight with which it will compete with other alternatives [19]. While in ranking-based the alternatives will be pair wise compared with respect to each criterion or dimension to derive normalized relative priorities of each alternative. The overall priority of each alternative and their rankings can then be used as the basis for selection. To apply this on our proposed system, it should be assessed against criteria or dimensions. For each criterion, if its performance value is less than the pre-defined threshold then it indicates that this is an area which needs improvement. However, if there so many areas, then the priority will be given to the area with the greatest weighted distance from perfection, defined by:

$$\text{WDP} = \text{CW (DW)} * \text{(Perfect score} - \text{Performance score)} \tag{2}$$

Where WDP: weighted distance from perfection; CW: criteria weight; DW; Dimension weight.

We depict our implementation for AHP in the following Table 2.

Table 2. Algorithm 1. Procedure of analytic hierarchy process

Algorithm 1. ***Procedure of Analytic Hierarchy Process***

This procedure computes the performance evaluation under reference scenario.

Input:

C: Criteria Matrix of size n.

A: Alternative Matrix, of size m.

Output:

W: Weights Matrix.

For each criterion $c \in C$, learn comparison matrix P according to expert reviews or DM.

- -Compute its eigenvector V, repeat steps until difference between successive solutions is less than a predefined threshold.
- -Raise P to powers that are successively square,
- -Obtain N matrix after calculating rows sum of P.
- -Normalize N.

For on each $c_i \in C$ $(i = 1,2,\ldots.n)$ $learn$ $a_j \in A(j1,2,\ldots..m)$, Obtain R_i $(m \times m)$by calculating pair-wise comparisons of each a_j according to c_i.

For each R_i compute its eigenvector $E_{j,(j=1,\ldots,n)}$.

Obtain W by Multiplying E ×V.

The next step in the experimental setup is defining the management alternatives as seen in Table 3. The first alternative A01, aims to enhance the educational process and the E-learning performance by adding positive modifications to courses, these modifications are determined after an assessment process for each course to determine its weakness points. Alternative A02 assumes that the deletion of a course is more appropriate as there is no use of modifying it, or that its participation in system performance is significantly poor, this is usually followed by a suggestion for another replacement of the deleted course.

Another alternative similar to A01 is A03, assuming modifications or cancelation A04 are determined and applied to system learning agents to enhance their productivity and there by boost the systems performance, modifications may include change of platform, addition in functionalities, these actions are automated unless necessary.

Alternative A05 is designed and applied to provide an indicator to the decision makers of wither the student or course LO have been written correctly, or need additional revision, sometimes the regression in performance is caused by wrongly written LO which could not by achieved.

If the value for Alternatives A05 or A06 is high, this will be an indicator that improvements are not only necessary for LO but also to course and agents respectively.

Table 3. The alternative matrix

Alternative	Description
A00. Do nothing	No changes are made to the system
A01. Modify Course	Modify course in order to enhance performance
A02. Cancel Course	Replace course with one higher in performance
A03. Modify Agent	Modify Agent in order to enhance performance
A04. Cancel Agent	Replace Agent with one higher in performance
A05. Modify learning outcome (LO)	Enhancing LO in favor of achievement
A06. (A01 + A05)	Improving the course joined by modifying LO
A07. (A03 + A05)	Modify Agent joined by modifying LO

As for the reference scenarios, we based the scenarios on changing factors affecting the E-learning system such as changes in learning outcomes, since it had direct effect on our system and was initially the goal of evaluation. The simulation of evaluation methods is run to identify the impact of changing learning outcomes on our system.

The values that are entered into the performance matrix were obtained by:

A. Pair wise comparison, where it's the responsibility of the decision makers (DM) or experts to make such a comparison between each criterion and the others usually based on their formal or informal judgment.
B. Random distribution of values until a satisfactory state or result is reached.

The first scenario is based on increased demand on the system to fulfill the LOs. The second scenario represents the best conditions containing random values in the range [0.0–2.0] assigned to criteria and alternatives assuming the focus on preserving the current LO achievements with respect to an increase in available agents, content, and courses. Meanwhile the decrease in these mentioned factors were projected in the third scenario represented.

4 Experimenral Results

We start with the performance matrix (PM) for each scenario, containing 6 dimensions versus 8 alternatives, thereby we have 3 matrices shown in Table 4 for the three scenarios.

In scenario (1) the feasible alternatives could by arranged where the top alternatives are A03, A04, A07 these alternatives focus on agents' modification or removal to enhance performance, followed by A01, A05, A06 which focus on modifying course and LO. As for A00 and A02, they were not feasible enough to overcome the domination of other alternatives as their values were slightly less that all others, This indicates that the top alternatives deserve deeper analysis. The same holds for scenarios (2) and (3).

Table 4. Performance matrix for different reference scenarios

Performance matrix for scenario1						
Alternatives	SLO	CLO	Adaptation	Pedagogical issues	System content	System agents
A0	0.56	0.19	0.68	0.19	0.30	0.69
A1	0.68	0.24	0.63	0.25	0.35	0.73
A2	0.55	0.19	0.69	0.16	0.29	0.72
A3	1.88	0.43	0.56	0.17	0.42	0.73
A4	0.80	0.34	0.36	0.15	0.36	0.56
A5	1.72	0.26	0.64	0.12	0.34	0.75
A6	0.89	0.49	0.57	0.18	0.42	0.74
A7	0.81	0.40	0.62	0.19	0.36	0.73
Performance matrix for senario2						
Dimensions	SLO	CLO	Adaptation	Pedagogical issues	System content	System agents
A0	0.66	0.29	0.78	0.29	0.4	0.79
A1	0.78	0.34	0.73	0.35	0.45	0.83
A2	0.65	0.29	0.79	0.26	0.39	0.82
A3	1.98	0.59	0.66	0.27	0.52	0.85
A4	0.9	0.54	0.46	0.25	0.46	0.66
A5	0.82	0.36	0.74	1.22	0.44	0.85
A6	0.99	0.53	0.67	0.28	0.52	0.84

(*continued*)

Table 4. (*continued*)

Performance matrix for senario2						
Dimensions	SLO	CLO	Adaptation	Pedagogical issues	System content	System agents
A7	0.91	0.5	0.72	0.29	0.46	0.83
Performance matrix for senario3						
Dimensions	SLO	CLO	Adaptation	Pedagogical issues	System content	System agents
A0	0.46	0.09	0.58	0.09	0.20	0.59
A1	0.72	0.16	0.54	0.02	0.28	0.63
A2	0.45	0.09	0.46	0.06	0.29	0.62
A3	1.78	0.39	0.59	0.07	0.32	0.65
A4	0.58	0.14	0.53	0.15	0.25	0.63
A5	0.70	0.24	0.26	1.05	0.26	0.46
A6	0.79	0.33	0.47	0.08	0.32	0.64
A7	0.71	0.30	0.52	0.09	0.26	0.63

The next step is calculating the criteria weights according to the methods explained before which are: Analytical Hierarchy Process (AHP), Entropy Method (EM), Inter-criteria Correlation (CRITIC) and simple derivation weighting (SDW).These objective weights are displayed in Table 5. The objective weights ranked the most important criteria to assess the alternatives performance as the SLO, Pedagogical issues, CLO in Scenario (1) but this was changed in Scenarios (2) and (3) as the pedagogical issues gained more importance than SLO, followed by CLO, after these come the adaptation criteria in all scenarios.

Table 5. Criteria weights obtained by weighting methods.

Dimensions	Student learning outcomes	Course learning outcome	Adaptation	Pedagogical issues	System content	System agents
Wight matrix for scenario 1						
Entropy method	**0.860**	**0.425**	−0.228	**0.054**	−0.021	−0.091
CIRTIC method	**0.515**	**0.214**	0.085	**0.116**	0.040	0.028
SDW method	**0.350**	**0.240**	0.118	**0.143**	0.089	0.058

(*continued*)

Table 5. (*continued*)

Dimensions	Student learning outcomes	Course learning outcome	Adaptation	Pedagogical issues	System content	System agents
Wight matrix for scenario 2						
Entropy method	**0.460**	**0.287**	−0.159	**0.485**	−0.013	−0.060
CIRTIC method	**0.170**	**0.083**	0.040	**0.666**	0.022	0.016
SDW method	**0.235**	**0.150**	0.079	**0.437**	0.055	0.041
Wight matrix for scenario 3						
Entropy method	**0.224**	**0.132**	0.030	**0.455**	0.037	0.119
CIRTIC method	**0.082**	**0.073**	0.031	**0.783**	0.016	0.013
SDW method	**0.169**	**0.161**	0.065	**0.527**	0.044	0.031

These criteria were considered important, while the others like system agents or system content were not considered to have a significant role in the decision making process as their scores are less divertive. The sum of these 3 criteria weights derived by the CIRTIC method and SDW ranges from [0.73–0.92] for all scenarios, with one interesting remark was that the pedagogical issues criteria gained importance as the scenarios moves towards better and worst values like in scenario (2) and (3).

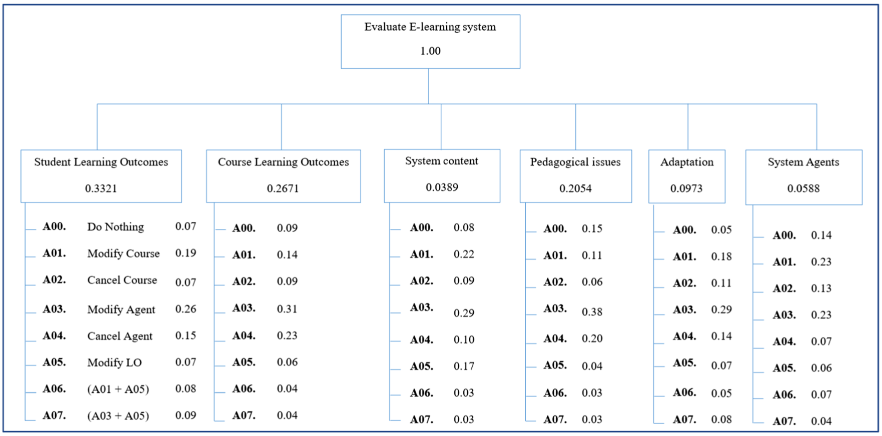

Fig. 1. Results obtained from analytic hierarchy process.

The results collected from the AHP method is illustrated in Fig. 1. The number of performed pair-wise comparisons between the criteria (6*(6 − 1)/2 = 15) and the alternatives are evaluated against each criterion to determine those who are not considered important in the objective weighting methods. With respect to the preference judgments of DMs who gave priority to environmental effects, SLO were considered to have strong importance over CLO, and demonstrated importance over pedagogical issues.

As for ranking the alternatives according to the distance methods, which calculate the Euclidean distance demonstrated by compromise programming (CP), and simple additive method (SAW). The alternative rankings summarized in Table 6 were as follows:

In Scenario (1), using EM, CIRTIC, SDW; A03, A05 and A06 were given the highest ranking amongst all other alternatives for all methods used. Following are alternatives A07, A04 and A01 while the worst were A00, and A02.

Table 6. Alternatives ranks from different MCDMs.

	SAW-EM	SAW_CRITIC	SAW_SDW	CP_EM	CP_CRITIC	CP _SDW
Ranking for Scenario1						
A00	7	7	7	7	7	7
A01	6	6	6	6	6	6
A02	8	8	8	8	8	8
A03	1	1	1	1	1	1
A04	4	5	5	4	5	5
A05	2	2	2	2	2	2
A06	3	3	3	3	3	3
A07	5	4	4	5	4	4
Ranking for Scenario2						
A00	7	7	7	7	7	7
A01	6	4	5	6	4	5
A02	8	8	8	8	8	8
A03	1	2	2	1	2	2
A04	4	6	6	4	6	6
A05	2	1	1	2	1	1
A06	3	3	3	3	3	3
A07	5	5	4	5	5	4
Ranking for Scenario3						
A00	7	6	6	7	6	6
A01	6	7	7	6	7	7

(*continued*)

Table 6. (*continued*)

	SAW-EM	SAW_CRITIC	SAW_SDW	CP_EM	CP_CRITIC	CP _SDW
A02	8	8	8	8	8	8
A03	2	2	2	2	2	2
A04	5	3	5	5	3	5
A05	1	1	1	1	1	1
A06	3	4	3	3	4	3
A07	4	5	4	4	5	4

As for AHP the alternatives were arranged according to Table 7 as follow; Alternatives A04, A03 and A01 were giving greater importance according to the DM who were concerned with outcomes criteria in a scenario aimed at enhancing the educational process, the alternatives modifying the agent was most preferable since it will cost less and provide better results than the others. The interesting thing is that when modifying the LO was not preferred by the DM, all alternatives including this choice were given low values and were placed on bottom of the list.

Table 7. Ranking of alternatives using AHP

Alternative	Weight
A00. Do nothing	0.0943
A01. **Modify Course**	**0.1628**
A02. Cancel Course	0.0815
A03. **Modify Agent**	**0.3003**
A04. **Cancel Agent**	**0.1740**
A05. Modify learning outcome (LO)	0.0645
A06. (A01 + A05)	0.0536
A07. (A03 + A05)	0.0581

5 Discussions and Findings

It has become quite important to develop evaluation methods for E-learning systems and be applicable with changing policies and environmental constrains. It is also necessary to focus this evaluation process on systems including their learning agents of all types and layers, in a form of simulator capable of handling all kind of scenarios and producing accurate results to DM, this is an important step in improving such systems.

Considering all presented scenarios, the alternative of modifying system agents was elected as most efficient management alternative followed by combining it with modifications to LOs. By that it would be possible to solve problems vastly if implemented immediately and recover from failure to improve current performance.

Traditional measures such as canceling a course or agents aren't considered efficient for enhancing the learning process and insuring learners' achievement of LOs. All the MCDM methods priorities the alternatives in a similar arrangement, there by the decision that is deduced from any method will be relatively similar to the other. With minor difference based on the method chosen and major difference on the weights assigned to the criteria.

The EM, CIRTIC, AHP, SDW methods can prove reliable in making use of all the information contained in row data, thereby considered reasonable ways in criteria weighting, and guaranteed to produce robust and fair decisions.

6 Conclusion

The success of E-Learning is highly depending upon adaptation and fulfillment of teaching objectives. The factors affecting the E-Learning systems success are many therefore it is essential to evaluate them so that the stakeholders, such as educational authority, students, and instructors, will be able to control the negative effects of each of these E-Learning factors and their dimensions in an effective manner. The Multi criteria decision making (MSDM) approach presented in this paper could prove successful in prioritizing each of the pedagogical, learning and adaptation s dimension according to a presented evaluation criterion. This categorization of factors will help stakeholders in ensuring the continuous improvement in E-learning systems and the design of more effective learning environments.

References

1. Mohammed, H.J., Mat Kasim, M., Mohd Shaharanee, I.N.: Evaluation of e-learning approaches using AHPTOPSIS technique. J. Telecommun. Electron. Comput. Eng. **10**(1–10), 7–10 (2018)
2. Mulla, Z.D., Osland-Paton, V., Rodriguez, M.A., Vazquez, E., Plavsic, S.K.: Novel coronavirus, novel faculty development programs: rapid transition to eLearning during the pandemic. J. Perinat. Med. **48**(5), 446–449 (2020)
3. Barteit, S., Guzek, D., Jahn, A., Bärnighausen, T., Jorge, M.M., Neuhann, F.: Evaluation of e-learning for medical education in low-and middle-income countries: a systematic review. Comput. Educ. **145**, 103726 (2020)
4. Supriyatno, T., Susilawati, S., Ahdi, H.: E-learning development in improving students' critical thinking ability. Cypriot J. Educ. Sci. **15**(5), 1099–1106 (2020)
5. Naveed, Q.N., et al.: Evaluating critical success factors in implementing E-learning system using multi-criteria decision-making. PLoS ONE **15**(5), e0231465 (2020)
6. Taherdoost, H.: Decision making using the analytic hierarchy process (AHP); A step by step approach. Int. J. Econ. Manag. Syst. **2** (2017)
7. Luo, L., et al.: Blended learning with Moodle in medical statistics: an assessment of knowledge, attitudes and practices relating to E-learning. BMC Med. Educ. **17**(1), 1–8 (2017)

8. Abdelkader, E.M.: A hybrid decision support system for holistic evaluation of material alternatives. Int. J. Strateg. Decis. Sci. (IJSDS) **12**(1), 18–36 (2021)
9. Wang, Y.J.: A fuzzy multi-criteria decision-making model based on simple additive weighting method and relative preference relation. Appl. Soft Comput. **30**, 412–420 (2015)
10. Sreelatha, K., Anand Raj, P.: Ranking of CMIP5-based global climate models using standard performance metrics for Telangana region in the southern part of India. ISH J. Hydraul. Eng. **27**, 1–10 (2019)
11. Gong, J.W., Liu, H.C., You, X.Y., Yin, L.: An integrated multi-criteria decision making approach with linguistic hesitant fuzzy sets for E-learning website evaluation and selection. Appl. Soft Comput. **102**, 107118 (2021)
12. Kadin, N., Ređep, N.B., Divjak, B.: Structuring E-learning multi-criteria decision making problems. In: 2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), pp. 705–710. IEEE (2017)
13. Rahmani, F., Ahmadi, H., Ghanbari, E., Khorasani Kiasari, S.M.: Evaluate and ranking the affecting factors in developing E-learning in higher education with fuzzy multi-criteria decision-making approach. Technol. Educ. J. (TEJ) **13**(2), 284–298 (2019)
14. Srinivasa Raju, K., Sonali, P., Nagesh Kumar, D.: Ranking of CMIP5-based global climate models for India using compromise programming. Theoret. Appl. Climatol. **128**(3–4), 563–574 (2016). https://doi.org/10.1007/s00704-015-1721-6
15. Koksalmis, E., Kabak, Ö.: Deriving decision makers' weights in group decision making: an overview of objective methods. Inf. Fusion **49**, 146–160 (2019)
16. Song, Y., Gui, Y., Gehringer, E.F.: An exploratory study of reliability of ranking vs. rating in peer assessment. Int. J. Educ. Pedag. Sci. **11**(10), 2405–240 (2017)