Eliciting preferences of professors and medical group students for evaluation methods of theoretical courses: An application of discrete choice experiment analysis

Ali Kazemi Karyan, Satar Rezaei, Shokooh Etesami[1], Leyla Pezhman[1], Behzad Karami Matin, Sajad Delavari[2]

Abstract:

BACKGROUND: Evaluation has become an inseparable part of education process which gives feedback to students and professors to improve education quality. This study aimed to elicit preferences of professors and students about attributes of evaluation methods in theoretical courses in Kermanshah University of Medical Sciences, Iran, in 2018.

MATERIALS AND METHODS: Discrete choice experiment (DCE) method used for eliciting preferences of participants of the study. A narrative literature review and interview with eight professors and ten students conducted to determine attributes and levels of evaluation methods in the university. Furthermore, experimental design used for making final choice sets of the evaluation methods. We included 213 students and 30 professors in the study. Conditional logistic regression model performed to data analysis.

RESULTS: Most of the professors (36.67%) preferred to allocate up to 30% of evolution scores to midterm examination. However, the most percentage of students (30.45%) were agree to include midterm examination up to 15% of total scores. The majority of students prefer to examination questions compromise just presented materials, while 70% of professors prefer to include additional texts for evaluation examinations. In case of quiz examination, professors in comparison with students prefer that quiz should have higher proportion of total scores. DCE analysis indicated that professors and students preferred a mix of questions in examinations. In addition, additional resources beyond what is taught in class made utility for professors and disutility for students. Quiz, also, increased the utility of an evaluation package in professors.

CONCLUSION: The findings showed that there is a gap between preferences of professors and students regarding some attributes of evaluation methods such as student's discipline, examination materials, and quiz. Further studies are needed to examining other attributes of evaluation methods in theatrical and practical courses in Iran and other contexts.

Keywords:

Discrete choice experiment, evaluation, preferences, professors, students, theoretical courses

Research Center for Environmental Determinants of Health, Health Institute, Kermanshah University of Medical Sciences, Kermanshah, Iran, [1]Students Researches Committee, Kermanshah University of Medical Sciences, Kermanshah, Iran, [2]Heath Human Resources Research Center, School of Management and Information Sciences, Shiraz University of Medical Sciences, Shiraz, Iran

Address for correspondence:
Dr. Behzad Karami Matin, Research Center for Environmental Determinants of Health, Health Institute, Kermanshah University of Medical Sciences, Kermanshah, Iran.
E-mail: bkm_1344@yahoo.com

Introduction

Education is one of the important factors in sustainable development of countries.[1-3]. It could improve economic growth, equity, and welfare.[3-5] Thus, educational institutions are trying to transfer knowledge and skills to the population to improve economic development of nations.[6]

The evaluation has become an inseparable part of education which gives feedback to

How to cite this article: Kazemi Karyan A, Rezaei S, Etesami S, Pezhman L, Matin BK, Delavari S. Eliciting preferences of professors and medical group students for evaluation methods of theoretical courses: An application of discrete choice experiment analysis. J Edu Health Promot 2021;10:82.

students and teacher to improve education process.[7] This could be a continual process during education or a point at the end. The results of evaluation could be very useful in designing other education or redesigning current courses to improve quality.[8] Thus, it could improve education process and reduce deficiencies. On the other hand, it could rank students which encourage them for more effort and let other agents to know students' learning and expertise.

In medical education, evaluation is more important than other fields. Medical students should learn their required knowledge, skills, and expertise with the minimum error to practice them for patients.[9] Professors need to evaluate their students and rank them based on their learning and skills. Thus, they try to do several types of evaluation methods to reach a reliable conclusion. However, different professors use different methods and as a result they reach to different outcomes which may not be reliable.[10] In fact, education systems need valid and reliable evaluation methods to evaluate university outcomes. If student evaluation be similar in universities, health systems could rely more on students' scores in universities and could easily compare their knowledge.[9]

However, there is a plenty of research on students' evaluation method, there is not enough literature on preparing a valid model for student and learning evaluation at universities. Most of the researches are done in descriptive form and tried to present the methods which are used by universities' professors and lecturers for student evaluation. Some others such as Pearson *et al*. (2012), tried to find evaluation criteria in scoring students.[11] Other researches use techniques such as analytical hierarchy process,[12,13] neural network,[14] fuzzy logic[15] or data mining[16] to evaluate education. Most of the mentioned models and methods are not applicable in practice and mainly are not based on preferences of instructors and students. Thus, it seems that these researches could not fulfill requirements of education system.

Considering preferences of students in different aspects of medical education can lead to improving the quality of health-care services that they will provide in the future.[17,18] Discrete choice experiment (DCE) could present models for analyzing preferences which have previously used in education valuation. Kennelly *et al*. (2015) found that business students prefer to have online assignments instead of paper assignments.[19] In addition, another study at McMaster University used DCE to investigate preferences of medical students about problem-based learning. This study indicated that students preferred small tutorial groups, and programs that have web-based support.[20] An increasing number of studies in Iran investigates learning styles and evaluation

methods among students and faculty members.[21-24] However, the evidence is rare about preferences of medical students and faculty members about attributes of current evaluation methods in Iran. This evidence could help education planners to analyze students and professors' preferences in scoring and evaluation. In this way, we could find evaluation criteria and their importance in scoring from different viewpoints. Thus, in the present research, we use DCE approach to find professors and students preferences about attributes of evaluation methods in theoretical courses in Kermanshah University of Medical Sciences (KUMS) in the west of Iran. The results could help educational planner to select a logical and valid attributes for evaluations.

Materials and Methods

This study was a descriptive-analytical study that used DCE method of analysis to elicit preferences of professors and students of KUMS about attributes of evaluation methods in theoretical courses during 2018. For this purpose, following steps of DCE method were done:

Determining attributes and levels

The basic hypothesis of DCE method is that we can describe the goods, services, methods, etc., using their attributes. It is suggested to use literature review and qualitative studies such as interview with experts for determining attributes and attributes-levels.[25,26] Therefore, we applied a narrative literature review and interview with eight professors and 10 students to determine attributes and levels of evaluation methods in the university. The experience of at least 5 years of teaching in the university was inclusion criteria for professors to include in qualitative study. In addition, students passed at least 20 theoretical courses in the time of study. All professors had academic degrees related to medical sciences such as health-care management, health education, health economics, nursing, and medicine. We continued interviews until achieving saturation in information. After reviewing literature and the interviews, the research team determines six main attributes to include in final design. The attributes and levels that included in final design are shown in Appendix 1.

Experimental design

By defining attributes and levels, scenarios were created using a combination of attributes and levels. In other words, this step of the study involves deciding which choices to make to present to the study participants.[27] The output of experimental design is set of choice sets that use to eliciting the preferences of participants. A choice set consists of the choices that the individual chooses. There are several ways to design the choice sets of study. However, in many cases, respondents are

offered two or more options to choose one among them. In this study, 7 choice sets, each with two scenarios, made by D-efficiency criteria to optimize the final design. Appendix 2 shows a choice set included in the study.

Finally, based on first phases of the study, the questionnaire of this study was developed. The questionnaire had two sections. First section included demographic, descriptive variables, and opinion of participants about attributes of evaluation methods. Second one included choice sets and definition of the attributes.

Data analysis

The base for analysis the DCE data is the random utility model (RUM). In according to RUM, in a situation that we request a person to choose between two evaluation plan of A or B, when he/she chose option A, it shows that this alternative have higher utility than alternative B. It can be write as follows:[28]

$$U(A, C) > U(B, C) \tag{1}$$

Here, U devotes the derived utility from the evaluation plans A and B. C is personal characteristics that are a common factor in two sides of equation 1, then the formula can be rewrite as:

$$V(A-B) = U(A, C) - U(B, C) \tag{2}$$

Here, V is indirect utility obtained from evaluation plan A in compare with evaluation plan B. the following linear regression model fitted to estimate the utility of attributes of evaluation plans:

$$V = \beta1 dis + \beta2\ mid + \beta3 que + \beta4 que + \beta5 que + \beta6\ sour + \beta7\ final + \beta8\ quiz + \mathcal{E} \tag{3}$$

Where, β1 to β8 are the coefficients of the attributes of evaluation plans: Dis (student's discipline), mid (midterm examination), que (question types), final (final examination), and quiz (quiz). Ɛ devotes error term. With considering the logistic distribution for error term, the conditional logit regression model was performed for data analysis.

The restrictive assumption of this model is independent of irrelevant alternatives. However, the interpretations of coefficients of this model is easier than other complicated models, such as nested or mixed logit models and it is shown this method is enough strong to use for analysis DCE data.[29]

We performed a subgroup analysis to elicit preferences of students by schools. Goodness of fit of models was investigated using McFadden R^2 and χ^2 tests.

The statistical software packages SAS 9.1 and Stata 14 (StataCorp, College Station, TX, USA) were used to perform the analysis.

Sampling

The sample size of the DCE studies is complex to determine. At least 30 participants should include in the studies. However, others suggested some formula to determining the sample size as follow:[30]

$$S = \frac{nta}{c} \tag{4}$$

Where, S is sample size, n is the total number of participants, t is the total number of choice sets, a is the number of alternatives in the choice sets, and c is the highest number levels in the included attributes. Accordingly, we included 213 students and 30 professors in the study. The convenience sampling method used for data collection by interview with participants.

Ethical statements

This study has been approved by the Ethics Committee of KUMS (ethics code: IR.KUMS.REC.1396.633).

Results

The descriptive results of the study showed that mean (standard deviation: [SD]) of students was 22.55 (2.49) years. Of total students, 213 (68.27%) were female and others were male. Samples were included from all schools of health (28.53%), nursing and midwifery (17.95%), medicine (4.49%), paramedical sciences (21.79%), pharmacy (6.41%), nutrition sciences (4.17%), and dentistry (16.67%). The mean (SD) age of professors was 37.93 (7.18) years and 76.67% were male. In addition, mean (SD) of number of courses that professors teach during their employment in the university was 108 (143.70). The highest percentage of professors (70%) had philosophy of doctor (PhD) degree; follow by degrees of master of sciences (26.67%), and medical doctor (3.33%).

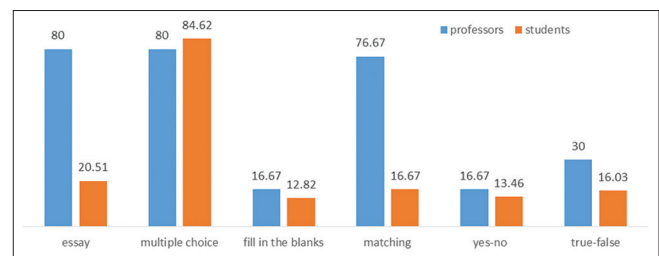Figure 1 shows the percentage of agreement with including different types of questions in evaluation

Figure 1: Percentage of agreement with inclusion of different types of questions in examinations

of students by professors and students. Professors believed that essay (80%), multiple choice (80%), and matching (76.67) questions were the most popular types of questions. On the other hand, multiple choice (84.62%) was the most preferred question by students and other types of questions were not very popular.

Preferences of participants about related items to the evaluation of students showed that 13.33% of professors and 33.65% of students were not agree to allocate some scores to student's discipline in class. In case of midterm examination, the most of professors (36.67%) prefer to allocate up to 30% of evolution scores to this item. However, the most percentage of students (30.45%) were agree to include midterm examination up to 15% of total scores. The majority of students prefer to examination questions compromise just presented materials, while 70% of professors prefer to include additional text for evaluation examinations in theoretical courses. Table 1 shows preferences of participants about attributes of evaluation plans in theoretical courses.

The results of DCE model to eliciting preferences of professors showed that in univariate logit regression model midterm examination, type of questions (at the level of a mix of essay, multiple choice, fill in the blanks, yes-no, true-false), source of examination, and quiz significantly increased the probability of selecting evaluation packages. However, multivariate model for professors' preferences indicated that all attributes increased the probability of choosing an evaluation method. With increasing a percent of midterm scores, the probability of selecting

an evaluation method increased by 2.2% ($P < 0.01$). The attributes of question types increased the utility of professors so that mix types of questions increased the probability of choosing the evaluation method by 17.7% than essay ($P > 0.1$). In case of source of examination, presented sources and additional references increased the utility of professors than presented sources in class (OR: 2.424, $P < 0.01$). Quiz, also, increased the utility of an evaluation package in professors. One percent increase in quiz share in total score increased the probability of selection an evaluation method by 7.3% ($P < 0.01$). The fitted model was statistically significant ($P < 0.001$). Table 2 presents the results of DCE analysis for eliciting the preferences of professors for attributes of evaluation plans in theoretical courses.

Univariate analysis of attributes of evaluation method showed that student's discipline and adding additional references to presented material in class have disutility for students ($P < 0.001$). Multivariate regression analysis indicated that student's discipline ($P < 0.001$), type of questions ($P < 0.05$), and quiz ($P < 0.001$) decrease the probability of selecting the evaluation method by the students. A mix of essay, multiple choice, fill in the blanks, yes-no, true-false questions in an evaluation method decreased the probability of choosing that method by 76%. One percent increase in final examination's score significantly increased the probability of the selecting evaluation method by 2.5% ($P < 0.01$). The fitted model was statistically significant ($P < 0.001$). Table 3 presents the results of DCE analysis for eliciting the preferences of students for attributes of evaluation plans in theoretical courses.

Table 1: Preferences of participants about attributes related to evaluation plans of students

Evaluation items	Professors, *n* (%)	Students, *n* (%)
Student's discipline		
No	4 (13.33)	105 (33.65)
Up to 5% of total score	18 (60)	124 (39.74)
Up to 10% of total score	8 (26.67)	83 (26.6)
Midterm exam		
No	2 (6.67)	68 (21.79)
Up to 15% of total score	9 (30)	95 (30.45)
Up to 30% of total score	11 (36.67)	79 (25.32)
Up to 40% of total score	8 (26.67)	70 (22.44)
Materials of exam		
Presented sources in class	9 (30)	269 (86.5)
Presented sources in class + additional references	21 (70)	42 (13.5)
Final exam		
Up to 50% of total score	13 (43.33)	108 (34.62)
Up to 70% of total score	12 (40)	144 (46.15)
Up to 90% of total score	5 (16.67)	60 (19.23)
Quiz		
No	4 (13.33)	154 (49.36)
Up to 5% of total score	21 (70)	120 (38.46)
Up to 10% of total score	5 (16.67)	38 (12.18)
Total	30 (100)	312 (100)

Eliciting of the preferences of students by schools of the university showed that students not prefer to discipline have some scores for all schools. A mix of essay, multiple choice, fill in the blanks, yes-no, true-false questions had disutility for students of all schools. Regarding the source of examination, the level of presented material and additional references made disutility for students of nursing and midwifery, and dentistry schools. Finally, quiz had disutility for students from all schools. The results of DEC analysis for eliciting the preferences of students for attributes of evaluation plans in theoretical courses by schools are presented in Table 4.

Discussion

This study primarily revealed that professors think essay, multiple choice, and matching questions are more appropriate than other questions for the evaluation of students' knowledge. They less prefer using other questions such as the fill in the blanks, yes-no, and true-false questions. On the other hand, students prefer multiple questions much more than other question types. Thus, it could be concluded that MCQ could be an evaluation tool that prefer by both students and professors. MCQs are using widely in universities and entrance examinations.[31,32] They have many pros and cons; but, if they are not designed well, barley considered as a suitable tools for deep knowledge assessment.[33] On the other hand, some studies revealed are a suitable assessment for in-depth learning, if they are designed properly.[32,34] Another issue about MCQs is the chance for cheating. Since this study revealed that MCQs are much more popular for students in compare with other types of questions, this maybe because of their easiness for cheating.[35] Now, there are several ways that could reduce the probability of cheating in MCQs such as test randomization,[36] computer-aided assessment,[37] and automatic test generator.[38] The

Table 2: Preferences of professors about attributes of evaluation plans using discrete choice experiment analysis

Attributes	Multivariate analysis			Univariate analysis		
	aOR	95% CI	*P*	OR	95% CI	*P*
Student's discipline	1.023	0.945-1.108	0.568	1.027	0.979-1.087	0.265
Midterm examination	1.022	1.007-1.037	<0.010	1.031	1.018-1.048	<0.001
Questions (ref: essay)						
Multiple choice	0.701	0294-1.667	0.421	0.514	0.298-0.888	0.017
Essay and multiple choice	1.177	0.654-2.116	0.586	0.739	0.447-1.223	0.240
A mix of essay, multiple choice, fill in the blanks, yes-no, true-false	1.177	0.493-2.809	0.714	1.126	0.658-1.931	0.666
Materials of exam (ref: presented material)						
Presented material and additional references	2.424	1.396-4.211	<0.010	2.307	1.592-3.346	<0.001
Final examination	1.004	0.984-1.024	0.689	0.989	0.979-0.999	0.043
Quiz	1.073	1.021-1.128	<0.010	1.064	1.019-1.111	<0.010
Observation		540			-	
LR χ^2 (8)		51.27			-	
P		<0.001			-	
McFadden R^2		0.104			-	

aOR=Adjusted odds ratio, OR=Odds ratio, CI=Confidence interval, LR=Likelihood ratio

Table 3: Preferences of students about attributes of evaluation plans using discrete choice experiment analysis

Attributes	Multivariate analysis			Univariate analysis		
	aOR	95% CI	*P*	OR	95% CI	*P*
Student's discipline	0.913	0.875-0.953	<0.001	0.973	0.960-0.986	<0.001
Midterm examination	1.009	1.003-1.016	0.007	1.000	0.997-1.004	0.559
Questions (ref: essay)						
Multiple choice	0.587	0.348-0.988	0.045	1.671	1.436-1.945	<0.001
Essay and multiple choice	0.568	0.329-0.980	0.042	1.517	1.311-1.776	<0.001
A mix of essay, multiple choice, fill in the blanks, yes-no, true-false	0.240	0.106-0.545	<0.001	1.014	0.867-1.187	0.857
Materials of examination (ref: presented material)						
Presented material and additional references	0.984	0.853-1.136	0.828	0.803	0.738-0.874	<0.001
Final examination	1.025	1.010-1.040	0.001	1.002	0.999-1.005	0.114
Quiz	0.947	0.919-0.975	<0.001	1.006	0.993-1.020	0.357
Observation		4368			-	
LR χ^2 (8)		93.930			-	
P		<0.001			-	
McFadden R^2		0.031			-	

aOR=Adjusted odds ratio, OR=Odds ratio, CI=Confidence interval, LR=Likelihood ratio

Table 4: Preferences of students about attributes of evaluation plans using Discrete choice experiment analysis by schools

Attributes	Health		Nursing and midwifery		Medicine		Paramedical sciences		Pharmacy		Nutrition sciences		Dentistry	
	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI
Student's discipline	0.875	0.791-0.967	0.930	0.848-1.019	0.928	0.765-1.125	0.973	0.880-1.076	0.846	0.682-1.050	0.995	0.797-1.240	0.881	0.800-0.969
Midterm examination	1.006	0.990-1.021	1.010	0.996-1.024	1.030	0.998-1.062	1.003	0.987-1.018	1.036	1.002-1.070	0.993	0.959-1.027	1.011	0.996-1.025
Questions (ref: essay)														
Multiple choice	0.360	0.102-1.270	0.652	0.216-1.962	1.000	0.098-10.166	1.000	0.287-3.479	1.200	0.087-16.440	1.333	0.088-20.108	0.355	0.112-1.123
Essay and multiple choice	0.307	0.079-1.176	0.883	0.287-2.721	1.001	0.090-11.082	1.038	0.277-3.886	0.584	0.036-9.507	1.736	0.101-29.944	0.340	0.107-1.077
A mix of essay, multiple choice, fill in the blanks, yes-no, true-false	0.105	0.015-0.731	0.477	0.083-2.755	0.655	0.016-26.134	0.388	0.056-2.691	0.444	0.007-25.748	0.495	0.007-34.301	0.132	0.021-0.818
Materials of examination (ref: presented material)														
Presented material and additional references	1.118	0.852-1.467	0.690	0.486-0.977	1.213	0.560-2.452	1.025	0.753-1.394	1.090	0.545-2.180	1.067	0.522-2.177	0.960	0.666-1.383
Final examination	1.044	1.007-1.080	1.014	0.983-1.046	1.006	0.941-1.074	1.016	0.981-1.053	1.005	0.933-1.081	1.009	0.935-1.089	1.031	0.998-1.065
Quiz	0.932	0.868-1.000	0.950	0.893-1.011	0.947	0.827-1.083	0.967	0.901-1.038	0.865	0.747-1.003	0.995	0.852-1.162	0.954	0.896-1.015
Observation	1246		784		196		952		280		182		728	
Log likelihood	-414.60		-258		-62.12		-316.17		-78.16		-58.64		-242.25	
LR χ^2 (8)	34.470		27.33		0.169		27.54		37.75		8.87		20.1	
P	<0.001		<0.001		0.169		<0.001		<0.001		0.354		<0.01	
McFadden R^2	0.040		0.05		0.085		0.042		0.194		0.070		0.039	

aOR=Adjusted odds ratio, OR=Odds ratio, CI=Confidence interval, LR=Likelihood ratio

results of a study for investigating evaluation methods in theoretical and practical courses of laboratory sciences students in Tehran University of Medical Sciences showed that the midterm exam is used in 35% of the total theoretical and practical courses, and among the evaluation methods, the MCQs method with 70% was among the common methods of assessments.[22] The study of Kojury *et al.* (2017), about types of evaluation methods of medical students at the Shiraz University of Medical Sciences indicated that MCQs is the most frequent method of evaluation. Based on this fact that MCQs cannot evaluate higher level of learning in students, this study suggested a mix of different types of evaluation methods such as descriptive examinations, and Mini-CEX in order to evaluate higher level of cognitive knowledge.[39]

In brief, in medical sciences assessment should be done on knowledge and skills of the student. However, MCQs fail to assess students' skill.[40] Thus, the limitations of MCQs should be considered by universities and professors who use it. It could be concluded that all types of question especially essay should be mixed with MCQs for better evaluation. Furthermore, evidence suggests that viva voce or oral examination method is a valid and reliable technique of student evaluation. This method can deeply evaluate the knowledge and skills of students and provide more opportunity for dialectic communication between examiner and student. However, examiner should manage stress of students for increasing effectiveness of this method.[41]

Many variables can effect on students preferences about attributes of evaluation methods. According to the current knowledge, personality and gender are among the main factors that influence individuals' preferences for student evaluation methods, including the type of questions.[42] In addition, gender, age, and emotional variables are known as variables that effect on academic achievement. Females more probability to have higher commitment and higher final degree grade point average compared to male students. In comparison to extraverts' personality, openness students favor essays and oral exams and do not prefer multiple choice and group work. In addition, females prefer coursework evaluation and oral examinations are more favorable for males. Besides personality, learning style, and intelligence supposed to be associated with preferences of student for examination techniques.[43]

According to our findings, more students preferred to discipline in class not include in the evaluation process, while more than 75% of professors agreed to consider the discipline of students as an item in this process. In addition, students had a higher tendency to have a

midterm exam. About half of the students did not prefer to have a quiz in their evaluation.

According to DCE analysis, professors preferred to student's discipline, final examination, and Quiz have higher scores in evaluation of students. Furthermore, they rather a mix of essay, multiple choice, fill in the blanks, yes-no, true-false questions to include in the evaluation exams. Moreover, they tended to consider presented material in the courses and additional related references as sources of examinations. On the other hand, considering a proportion of total score to discipline of students, additional sources for examinations, and quiz by professors for evaluation made disutility for students. Furthermore, students prefer essay and MCQs than other type of questions. MCQs made higher utility for pharmacy, and nutrition sciences schools' students than other ones. Similarly, another study in Iran showed there are some match and mismatch points between learning preferences of professors and students.[44]

Aragon (2010) indicated that if face-to-face and online training courses are designed appropriately and based on learning theories, they will equally affect people's learning. Therefore, carefully designing training courses and evaluating tests and adapting these courses to students' preferences can play an important role in learning and evaluating them.[45]

One of the strength of this study is employing an econometric technique to find students and professors which deal with the research novelty. It is the first attempt in Iran to applying this method in eliciting preferences in medical education. Another strength is considering students' and professors' preferences simultaneously. The research achieved sound knowledge about educational assessment in universities. However, it faced with several limitations. The first limitation is related to sample selection method which was convenient sampling. Thus, the results may not be completely generalizable to study population or other universities. Study participants were included from KUMS. Therefore, the results could not representative of preferences of professors and students in other medical university of Iran. As there are limitations for number of included attributes in DCE method, we could not consider all important attributes. Therefore, future works should consider other aspects of evaluation methods.

Conclusion

The findings showed MCQs as a favorite test form for both professors and students. However, professors prefer other types of question including essays and matching questions too. Both professors and students prefer a mix of questions in the exams. There is a gap between preferences of professors and students regarding attributes of evaluation methods such as student's discipline, examination materials, and quiz. Further studies are needed to examining other attributes of evaluation methods in theatrical and practical courses in Iran.

Conflicts of interest
There are no conflicts of interest.

References

1. Hu J, Huang Q. Higher Education Sustainable Development in Sichuan Province: Problems and Solutions; 2013. p. 3663-6.
2. Chen X. Harmonizing ecological sustainability and higher education development: Wisdom from Chinese ancient education philosophy. Educ Philos Theor 2019;51:1080-90.
3. Guo L, Huang J, Zhang Y. Education development in China: Education return, quality, and equity. Sustainability 2019;11: 13, 10.3390/su11133750.
4. Shin JC. Higher education development in Korea: Western university ideas, Confucian tradition, and economic development. High Educ 2012;64:59-72.
5. McGrath S. Education and development: Thirty years of continuity and change. Int J Educ Dev 2010;30:537-43.
6. Darling-Hammond L, Flook L, Cook-Harvey C, Barron B, Osher DJ. Implications for ducational practice of the science of learning and development. Applied Developmental Sciences, 2020;24:97-140. https://doi.org/10.1080/10888691.2018.1537791.
7. Zhao C, Zhao Y, Luo L, Li Y. Fractional modeling method research on education evaluation. J Softw 2011;6:901-7.
8. Zhao P. Study on Higher Education Quality Evaluation System of Local Universities; 2013. p. 2113.
9. Lewallen LP. Practical strategies for nursing education program evaluation. J Prof Nurs 2015;31:133-40.
10. Botek M. Comparison of education evaluation models. Qual Access Success 2018;19:63-7.
11. Pearson T, Garrett L, Hossler S, McConnell P, Walls J. A progressive nurse practitioner student evaluation tool. J Am Acad Nurse Pract 2012;24:352-7.
12. Sun X, Fang X. Construction and operation of analytic hierarchy process about moral education evaluation in colleges and universities. Adv Inf Sci Serv Sci 2011;3: DOI: 10.4156/aiss.vol3. issue11.15.
13. Wang K, editor. Ideological and Political Education Evaluation of Universities by a Novel AHP. 2014 Sixth International Conference on Measuring Technology and Mechatronics Automation, Zhangjiajie, 2014, pp. 309-312, doi: 10.1109/ICMTMA.2014.77.
14. Bao Q, Gu X, Shen Y, editors. An Improved Neural Network Model for Graduate Education Evaluation. 2007 International

Conference on Computational Intelligence and Security (CIS 2007), Harbin, 2007, pp. 550-554, doi: 10.1109/CIS.2007.190.

15. Xu J, Zhang G, editors. The Construction of Enterpreneurship Education Evaluation Index System Based on Dynamic Fuzzy Logic Theory. 2012 DOI: 10.1109/ISRA.2012.6219149.

16. Cheng Y, Wang Y. Design of Network Education Evaluation System Based on Data Mining Technology. Applied Mechanics and Materials 241-244:2849-2852 و DOI: 10.4028/www.scientific.net/AMM.241-244.2849.

17. Shams A, Yarmohammadian MH, Abbarik HH. Determination of rate of customer focus in educational programs at Isfahan University of Medical Sciences(1) based on students' viewpoints. J Educ Health Promot 2012;1:24.

18. Danaei SM, Mazareie E, Hosseininezhad S, Nili M. Evaluating the clinical quality of departments as viewed by juniors and seniors of Shiraz dental school. J Educ Health Promot 2015;4:75.

19. Kennelly B, Flannery D, Considine J, Doherty E, Hynes S. Modelling the preferences of students for alternative assignment designs using the discrete choice experiment methodology. Pract Assess Res Eval 2014;19:1-13.

20. Cunningham CE, Deal K, Neville A, Rimas H, Lohfeld L. Modeling the problem-based learning preferences of McMaster University undergraduate medical students using a discrete choice conjoint experiment. Adv Health Sci Educ Theory Pract 2006;11:245-66.

21. Sadeghi T, Loripoor M. Usefulness of 360 degree evaluation in evaluating nursing students in Iran. Korean J Med Educ 2016;28:195-200.

22. Abbasi S, Einollahi N, Gharib M, Nabatchian F, Dashti N, Zarebavani MJ. Evaluation methods of theoretical and practical courses of paramedical faculty laboratory sciences undergraduate students at tehran university of medical sciences in the academic Year 2009-2010. Payavard 2013;6:342-53.

23. Bastani P, Amini M, Taher Nejad K. Shaarbafchi Zadeh NJJoAiME, professionalism. Faculty members' viewpoints about the present and the ideal teacher evaluation system in Tehran. Univ Med Sci 2013;1:140-7.

24. Sarabi-Asiabar A, Jafari M, Sadeghifar J, Tofighi S, Zaboli R, Peyman H, *et al*. The relationship between learning style preferences and gender, educational major and status in first year medical students: A survey study from iran. Iran Red Crescent Med J 2015;17:e18250.

25. Ryan M, Kolstad JR, Rockers PC, Dolea C. How to Conduct a Discrete Choice Experiment for Health Workforce Recruitment and Retention in Remote and Rural Areas: A User Guide with Case Studies. The World Bank; 2012.

26. Karyani AK, Rashidian A, Sari AA, Sefiddashti SEJMjotIRoI. Developing attributes and levels for a discrete choice experiment on basic health insurance in Iran. Med J Islam Repub Iran 2018;32:26. doi: 10.14196/mjiri.32.26.

27. Johnson FR, Lancsar E, Marshall D, Kilambi V, Mühlbacher A, Regier DA, *et al*. Constructing experimental designs for discrete-choice experiments: Report of the ISPOR conjoint analysis experimental design good research practices task force. Value Health 2013;16:3-13.

28. Hauber AB, González JM, Groothuis-Oudshoorn CG, Prior T, Marshall DA, Cunningham C, *et al*. Statistical methods for the analysis of discrete choice experiments: A report of the ISPOR conjoint analysis good research practices task force. Value Health 2016;19:300-15.

29. Cushing B. Conditional logit, IIA, and Alternatives for Estimating Models of Interstate Migration; 2007. Available from: https://researchrepository.wvu.edu/rri_pubs/65.[Last accessed on 2020 Jun 24].

30. Orme BK. Getting started with conjoint analysis: Strategies for product design and pricing research: Research Publishers, LLC; 2006.

31. Shigli K, Nayak SS, Gali S, Sankeshwari B, Fulari D, Shyam Kishore K, *et al*. Are multiple choice questions for post graduate dental entrance examinations spot on?-Item analysis of MCQs in prosthodontics in India. J Natl Med Assoc 2018;110:455-8.

32. Justice P, Marshman E, Singh C. Improving student understanding of quantum mechanics underlying the Stern-Gerlach experiment using a research-validated multiple-choice question sequence. Eur J Phys 2019;40:1-15. DOI: 10.1088/1361-6404/ab2135.

33. Mkrtchyan A, Abrahamyan A, editors. Combined Multiple Choice Questions: An objective and efficient assessment of knowledge. Proceedings of the IADIS International Conference e-Learning 2011, Part of the IADIS Multi Conference on Computer Science and Information Systems 2011, MCCSIS 2011; 2011.

34. Villafranca JJ, Ash DE, Wedler FC. Manganese(II) and substrate interaction with unadenylylated glutamine synthetase (*Escherichia coli* w). II. Electron paramagnetic resonance and nuclear magnetic resonance studies of enzyme-bound manganese(II) with substrates and a potential transition-state analogue, methionine sulfoximine. Biochemistry 1976;15:544-53.

35. Chen M, editor. Detect Multiple Choice Exam Cheating Pattern by Applying Multivariate Statistics. Proceedings of the International Conference on Industrial Engineering and Operations Management, Bogota, Colombia, October 25-26, 2017.

36. McLeod AI, Zhang Y, Yu H. Multiple-choice randomization. J Stat Educ 2003;11:https://doi.org/10.1080/10691898.2003.11910695.

37. Tang JK, Yeung HC, Man YF, Li KK, Wong TL, Pang WM, *et al*. A Web-Based Computer-Aided Assessment Creation and Invigilation System; 2015. p. 74-85.

38. Wang JY, Chen FG, Tai DW, Chen JS, editors. An Automatic Test Generator for Enhancing the Technolgy Profciency of Senior High School Students. 2016, 5th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI).

39. Kojury J, Rivaz S, Amini M, Rivaz MJBJoMEEDCBUoMS. Assessment of educational group's status based on types of evaluation methods of medical students at the Shiraz University of Medical Sciences 2014. Journal of Medical Development Center (EDC), 2017;5:7-13.

40. Tangianu F, Mazzone A, Berti F, Pinna G, Bortolotti I, Colombo F, *et al*. Are multiple-choice questions a good tool for the assessment of clinical competence in Internal Medicine? Ital J Med 2018;12:88-96.

41. Pearce G, Lee G. Viva voce (oral examination) as an assessment method: Insights from marketing students. J Marketing Educ 2009;31:120-30.

42. Lakhal S, Sévigny S, Frenette É. Personality and preference for evaluation methods: A study among business administration students. Stud Educ Evaluat 2013;39:103-15.

43. Sheard M. Hardiness commitment, gender, and age differentiate university academic performance. Br J Educ Psychol 2009;79:189-204.

44. Abolfaz Khademi. Preference for students to learn English. Daneshvar Raftar 2003;11:29-42.

45. Aragon SR. Facilitating Learning in Online Environments: New Directions for Adult and Continuing Education, Number 100: John Wiley & Sons; 2010.

Appendix

Appendix 1: Attributes and attributes-levels of evaluation plans of students in theoretical courses in Iran

Row	Attributes	Levels
1	Student's discipline	1. Not included (0%), 2. up to 5% of total score, 3. up to 10% of total score
2	Midterm exam	1. Not included, 2. up to 15% of total score, 3.up to 30% of total score, 4.up to 40% of total score
3	Questions	1. Essay, 2. Multiple choice, 3. A mix of essay and multiple choice questions, 4. a mix of essay, multiple choice, fill in the blanks, yes-no, true-false questions
4	Materials of examination	1. Presented material, 2. Presented materials and additional references.
5	Final examination	1. Up to 50% of total score, 2.up to 70% of total score, 3. up to 90% of total score
6	Quiz	1. Not included (0%), 2. up to 5% of total score, 3. up to 10% of total score

Appendix 2: One of the choice set included in the study

Attributes	Plan A	Plan B
Student's discipline	Up to 5% of total score	Not included (0%)
Midterm examination	Up to 40% of total score	Not included
Questions	A mix of essay and multiple choice questions	A mix of essay, multiple choice, fill in the blanks, yes-no, true-false questions
Materials of examination	Presented materials and additional references	Presented material
Final examination	Up to 50% of total score	Up to 70% of total score
Quiz	Not included (0%)	Not included (0%)

Which of the evaluation plans would you like to choose (please tick just one box) plan A ☐ plan B ☐