Comparison of Analytic Indices of In-Person vs. Online Exams in an Iranian Medical University in the Academic Year 2020

Reza Malek Pourafshar[1], Reza Shojaeipour[2], Payam Khazaeli[3], Azam Bazrafshan[4], Amin Beigzadeh[5], Mahmoud Reza Dehghani[6*]

[1]Professor of Pathology, Medical Education Leadership and Management Research Center, Kerman University of Medical Sciences, Kerman, Iran

[2]PhD Candidate, Faculty of Psychology and Educational Sciences, Allameh Tabatabai University, Tehran, Iran

[3]Professor of Pharmaceutics, Department of Pharmaceutics, School of Pharmacy, Pharmaceutical Research Center, Kerman University of Medical Sciences, Kerman, Iran

[4]Master of Library Science, Neuroscience Research Center, Neuropharmacology Research Institute, Kerman University of Medical Sciences, Kerman, Iran

[5]PhD in Medical Education, Assistant Professor, Medical Education Leadership and Management Research Center, Kerman University of Medical Sciences, Kerman, Iran

[6]General Practitioner, Master of Medical Education, Instructor, Medical Education Leadership and Management Research Center, Kerman University of Medical Sciences, Kerman, Iran

Abstract

Background: Currently, many medical universities in Iran use e-learning programs to educate and evaluate students. This teaching and assessment method is highly significant during the coronavirus epidemic.

Objectives: The present study was conducted to compare the analytic indices of in-person vs. online exams at the Kerman University of Medical Sciences, (KMUS) Iran.

Methods: This descriptive-analytical cross-sectional study was conducted in 2020. The study samples included all exams given at the KMUS and midterm exam scores obtained from in-person and online courses in the first and second semesters in the academic year 2019-2020. The exams were selected based on courses, and the same courses were offered both in-person and online; thus, only one group was studied. Course exam indicators, including difficulty and discriminative index, were examined. Data analyzed using SPSS software version 22.

Results: The mean of the difficulty index related to in-person and online exams were (0.62 ± 0.1) and (0.68 ± 0.1), respectively. (P=0.01). The mean of the discriminative index related to in-person (0.30± 0.07) and online (0.33± 0.08) exams had no statistically significant difference (P>0.05). The frequency of easy questions in online exams was significantly higher, (55% vs. 43%) (p = 0.008). The frequency of questions with an appropriate discriminative index was significantly higher for in-person exams compared to online exams (58% vs. 54%) (p = 0.01).

Conclusion: The difficulty index was acceptable and appropriate for both in-person and online exams. Both exams had appropriate quality according to the discriminative index Therefore, it seems that e-learning programs and assessments can be a good alternative to in-person teaching in emergencies.

Keywords: Online Education, Student Evaluation, Coronavirus, Online Exam, Difficulty Index, Discriminative Index

Background

The advent of information and communication technology and the growing expansion of new communication platforms such as the World Wide Web

in universities have changed many educational activities. The emergence of "e-learning" is one of the outcomes of these developments in the field of education, which refers to various forms of learning and education based on new technologies. One of the basic features of e-learning is the use of information, knowledge, and educational technologies to establish communication among people with educational resources in the form of formal or informal education (1).

In recent years, many efforts have been made to grow and expand the efficiency of e-learning. E-learning refers to a form of learning, in which teachers and learners are separated by physical distance so that students unable to attend the class for any reason to receive in-person instruction can learn the lesson online via the Internet (2). Therefore, it is necessary to evaluate the quality of e-learning. Numerous models and frameworks have been introduced to assess the validity of e-learning programs (5-3), most of which agree on the fact that basic quality principles are the same in e-learning programs and traditional educational programs offered in-person. If educational activities are well-designed, desirable outcomes will be obtained regardless of the context, in which the activities are presented. E-learning is not only a novel method to implement traditional (class-based) teaching methods, and is a new approach to education. Thus, e-learning quality assessment indicators should be considered accordingly (6).

E-learning can have different advantages. For instance, it can create diverse educational media and provide a wider range of visual learning tools. Students have better access to educational content and individual learning. Moreover, it facilitates learner feedback and makes abstract concepts more understandable using multimedia educational tools. However, e-learning is not considered a real training and an effective type of learning due to its cold and mechanical learning environment and lack of vital and spontaneous interaction between teachers and learners. Therefore, desirable teaching goals in e-learning programs cannot easily be turned into accurate and measurable goals (7). The results of a study entitled "Assessing the impact of e-learning on the academic achievement of students of Isfahan University of Medical Sciences" showed that empowering students with an average of 3.55, education with an average of 3.60, equipment and facilities with an average of 3.65, education content with an average of 3.55, and students' awareness with an average of 3.53 was effective in the students' academic achievement (8).

Although teaching in the classroom has to date been a predominant method of teaching and learning, today, especially in the current situation, the teaching-learning process must be offered online and outside the classroom. Scientific evidence shows that classroom-based methods are no longer accepted for medical education and new methods are required (9-11). With the global spread of COVID-19 in the world and severe restrictions on physical distance in high-risk populations such as students, many universities and educational institutions are using e-learning programs and teaching courses. In this regard, 150 countries have closed educational institutions and universities, which

account for about 80% of the world's student population (12), and e-learning has replaced in-person teaching (13,14). Due to the significance of the major factors, all basic sciences, paramedical, and health courses) at the Kerman University of Medical Sciences were held through online education. Given that this is the first time that online courses are being extensively used in Iranian medical universities, there is not much evidence about the experiences of universities regarding the quality of educational programs and the challenges of implementing e-learning programs. Therefore, providing initial evidence about the quality of e-learning programs can pave the way for decision-makers in the field of education in the country to continue and promote e-learning programs more seriously. In this regard, the present study aimed to evaluate e-learning education programs at the Kerman University of Medical Sciences in the second semester of the academic year 2019-2020 by comparing students' scores in online midterm exams with their scores from exams of similar content in the previous semester in terms of difficulty index, discriminative index, mean scores, frequency of questions with different difficulty level, and discriminative index. This difference was not statistically significant Difficulty and discrimination index are major indicators examined in the analysis of questions of an exam. An appropriate difficulty index indicates the difference between subjects, while an appropriate discriminative index can distinguish between strong and weak students. Since exams as the main assessment tool indicate student's degree of learning and their achievement of educational goals, it is necessary to check the quality of questions in any exam and ensure standardization of exams.
Objectives: The present study was conducted to compare the analytic indices of in-person vs. online exams at the Kerman University of Medical Sciences, (KUMS), in the first and second semesters in the academic year 2019-2020.

Methods

In this analytical cross-sectional study, students' scores from online midterm exams in different fields of medical science were extracted through the online examination system of the university (Faradid) and compared with their scores from in-person exams of similar content in the previous semester. The study samples included all online exams held electronically at the university in the second semester of the academic year 2019-2020. Until the end of the study, 102 exams were conducted, of which 30 were selected with simple random sampling using a random number table. The inclusion criteria were exams on courses offered in both semesters of the academic year 2019-2020. Courses not offered in the first semester of the same academic year or those with different or distorted target groups were excluded from the final analysis. Finally, a total of 30 exams met the inclusion criteria. Data on exam indicators such as mean scores, difficulty index, discriminative index, frequency of easy, appropriate, and difficult questions, and frequency of questions with low, appropriate, and inappropriate discriminative index entered SPSS software version 22 after extraction from the system

and quality control of the data. Central statistical indicators and frequency tables, frequency percentages, and graphs were used to describe the data. Further, means were used to compare the distribution of scores in the online and in-person exams. Given that the data distribution was not normal in the studied samples, the Mann-Whitney U test was used to compare the means. The significance level was considered less than 0.05.

Results

In this study, from a total of 102 exams, 30 exams were selected. Table one provides the comparison of exam indicators related to in-person and online courses in the academic year 2019-2020.

According to the table one, the mean of the difficulty index had a statistically significant difference between in-person and online exams given in the first and second semesters of the academic year 2019-2020.

Comparing the mean of the discriminative index of exams given in the first and second semesters in the academic year 2019-2020 showed no statistically significant difference in this regard.

In addition, examining the frequency of easy questions in the exams in both semesters showed a significantly higher frequency of easy questions in the online exams (55%) compared to that in the in-person exams (43%). The frequency of questions with an appropriate difficulty index was measured in the exams given in both semesters and showed a statistically significant difference. A statistically significant difference was also observed in terms of the frequency of difficult questions in the exams given in the first and second semesters of the academic year 2019-2020 (10.8% vs. 9.3%, respectively).

Table 1. Comparison of Exam Indicators Related to In-Person and Online Courses at the Kerman University of Medical Sciences in the Academic Year 2019-2020

Exam indicator	In-person exams, 2019 Mean (SD)	Online exams, 2020 Mean (SD)	P
Difficulty index	0.62(0.1)	0.68(0.1)	0.01
Discriminative index	0.30(0.07)	0.33(0.08)	0.9
Cronbach's alpha coefficients	0.65(0.17)	0.68(0.15)	0.7
Easy questions (N)	43.6(22.3)	55.4(24.3)	0.01
Appropriate questions (N)	45.5(19)	35.1(21.3)	0.02
Difficult questions (N)	10.8(8.1)	9.30(8.9)	0.2
Questions with an appropriate discriminative index (N)	57.7(17.4)	53.8(15.5)	0.2
Questions with a low discriminative index (N)	36.5(14.7)	40.9(15.3)	0.2
Questions with an inappropriate discriminative index (N)	2.30(2.5)	4.20(4.9)	0.1

Moreover, the frequency of questions with a low discriminative index was not significantly different in the exams given in the first and second semesters (37% vs. 41%, respectively). Comparing the frequency of questions with an appropriate discriminative index showed no statistically significant difference between the exams given in both semesters. However, the frequency of questions with an appropriate discriminative index between 0.3 and 0.7 was slightly higher in the in-person exams (58%) compared to that in the online exams (54%). No statistically significant difference was also observed in terms of the frequency of questions with an inappropriate discriminative index between the exams given in-person (2.3%) and those given online (4.2%).

Finally, Cronbach's alpha coefficient showed almost similar mean coefficients between the in-person (0.65) and online (0.68) exams given in the first and second semesters, respectively.

Discussion

This study compared several exam indicators related to in-person and online exams on courses offered at the Kerman University of Medical Sciences, Iran, in the first and second semesters of the academic year 2019-2020, including difficulty index, discriminative index, frequency of questions with different difficulty levels, and discriminative index.

A comparison of the mean of difficulty index showed a statistically significant difference between the in-person and online exams in the first and second semesters of the academic year 2019-2020, respectively. This indicated that both types of exams had an appropriate difficulty index of 0.60 (15). These results are consistent with the findings of Baharvand et al. (16), Imam Juma, and Zahedifar (17) who reported the difficulty index of exams between 0.60 and 0.76. Considering that the best difficulty index is in the range between 30%-70% (18), it can be mentioned that the difficulty index of the exams given at the Kerman University of Medical Sciences was suitable.

The mean of discriminative index was not significantly different between the exams given in-person and online in the academic year 2019-2020. This indicates that different types of exams act similarly in terms of differentiating between strong and weak students in emergencies. Online exams can be a proper alternative to in-person exams.

Moreover, the mean of discriminative index of all exams given in both semesters was at a high level of 0.3. Therefore, online electronic exams can be used with more confidence instead of in-person exams.

Based on the mean of discriminative index of both exams, it is necessary to take more measures for both in-person and online exams to help distinguish between weak and strong students. A larger discriminative index indicates the discrimination power of the question, and a closer percentage of this index to 100 demonstrates that it is more appropriate (19).

Scientific evidence shows that the exam is the main tool of evaluation in the process of teaching and learning and if it is continuously designed based on scientific principles and standards, it will indicate the extent to which teachers and students achieve predetermined goals. The ultimate educational goal is learning achievement. Therefore, if a test does not have the desired format in terms of the taxonomy of questions, compliance with structural rules, content validity, difficulty index, discriminative index, and other test standardization measures, not only is the main role of the exam achieved but it also motivates the learners. Such tests will have negative effects on the learning process and will waste the efforts of teachers and the education system. Therefore, it is necessary to check the quality of questions and ensure the standardization of exams (20). On the other hand, one of the most significant issues in non-invigilated online exams is the issue of cheating, and medical universities must follow necessary measures to manage this issue. In this regard, it is recommended that in online electronic exams to assess the affective or psychomotor domain, which is in the form of multiple-choice questions, the following items should be considered by exam designers. Taxonomy 3, which evaluates the decision-making, application, and problem-solving power of students, should be used in testing so that students cannot easily answer questions. Therefore, Taxonomy 3 contains a higher percentage of questions (50%) than Taxonomy 1 (20%) and 2 (30%). Other factors include having a time limit, having random questions, and eliminating the possibility of switching between questions. The mean scores of all the 30 exams in the first and second semesters were compared and no significant difference was observed. This indicates that both types of exams had a similar effect on the students' academic achievement. Scientific evidence shows that 94% of learners who completed distance learning courses believe that they learned more in online classes than in traditional face-to-face classes (22). This finding suggests that e-learning can effectively increase students' learning capacity and outcomes. On the other hand, other scientific findings show that online teaching cannot be solely used as the primary approach at universities.

The mean of the frequency of easy questions was examined in the exams in the first and second semesters, showing that the frequency of easy questions was significantly higher in the online exams (55%) than in the in-person exams, (43%). This indicates that exam designers are more inclined to ask simpler questions in online exams

since it is believed that students take their online tests under special conditions which encourages professors to ask simpler questions. Comparing the mean frequency of questions with the appropriate difficulty index showed no statistically significant difference between the two types of exams in the first and second semesters, and thus, the quality of exam questions was observed to be highly similar (45% vs. 35%). Moreover, no statistically significant difference was observed between the two types of exams in the first and second semesters in terms of the mean frequency of difficult questions (10.8% vs. 9.3% for online and usual electronic exams, respectively). Despite the degree of importance, few comprehensive studies have examined theoretical exam questions in terms of various qualitative indicators. In a study conducted by Baharvand et al. at the Shahid Beheshti School of Dentistry, 30 qualitative indicators of structural validity along with the content of theoretical exams were examined. The results showed that written exams on theoretical courses performed at the Shahid Beheshti School of Dentistry were at an appropriate level in terms of communication and content coverage, compliance with structural rules, and difficulty index, but needed to be reconsidered in terms of taxonomy, discriminative index, and value for each deviant option (16). No statistically significant difference was also observed between the online and in-person exams after examining the mean frequency of questions with a low differentiation coefficient. Therefore, the theory is reinforced that in online exams, the same standard question design criteria (Millman criteria) are observed by professors, as in-person exams (37% vs. 41%, respectively). Comparing the mean frequency of questions with an appropriate discriminative index showed a statistically significant difference between the exams of both semesters in this regard. Accordingly, the mean frequency of questions with an appropriate discriminative index between 0.3 to 0.7 was slightly higher in the in-person exams (58%) compared to the online exams (54%). This shows that the number of questions with an appropriate discriminative index was the same in both types of exams, confirming that online exams have the same required standards as real face-to-face exams. However, the mean frequency of questions with an inappropriate discriminative index was almost similar in both types of exams (2.3% in the in-person exams and 4.2% in the online exams). This shows that online exam questions have the same ability to differentiate between a strong student and a weak student.

An evaluation of Cronbach's alpha coefficient showed no statistically significant difference between the exams of the first and second semesters (0.65 vs. 0.68, respectively). This indicates that the reliability coefficient of both types of exams is in an acceptable range (above 0.6). In other words, online exams have acceptable reliability and validity, and in-person exams can be used to evaluate students' academic level.

According to the results of this study, it appears that e-learning educational programs can be used as part of routine educational programs at universities, especially during the coronavirus epidemic, for evaluation of medical

sciences. Research shows that online academic education is a successful and efficient system if the educational content is properly structured and evaluated (24). One of the limitations of this study was inaccessibility to all information on the exam indicators.

Conclusion

Evaluation is one of the most significant parts of the educational process. Proper evaluation can identify the strengths and weaknesses of education (27). Exam analysis associated with testing is a major step in educational evaluation, in which the degree of accuracy and inadequacy are examined in each question to finally determine the strengths, weaknesses, and quality of an exam. The results of this study showed that at the Kerman University of Medical Sciences, the difficulty and discriminative index were acceptable and appropriate for both in-person and online exams. The experience of the Kerman University of Medical Sciences in the academic year 2010-2011 confirms the fact that if educational processes are implemented with quality and necessary technical standards are observed, e-learning programs and online exams can be of the same quality as traditional education and in-person classes. However, the continuation of this path requires preparation of all components of the educational system such as professors and students, and proper use of existing technological capacities to improve the quality of existing educational programs. It is suggested to conduct further research on the subject in other medical universities of the country. It is also suggested to examine other exam indicators such as structural validity, relevance and content coverage, and taxonomy.

References

1. Ehlers UD, Hilera JR. Special issue on quality in e-learning. Journal of Computer Assisted Learning. 2012; 28(1): 1-3. doi:10.1111/j.1365-2729.2011.00448.x
2. Vafaee Najar A, Mohammadi M, Khiabani B, Ibrahimpour H. Attitude and Performance of Faculties Towards the Implementation of the Electronic Learning System (ELS) in Mashhad University of Medical Sciences (MUMS) in 2009. Iranian Journal of Medical Education. 2011; 11(2): 120-7. [In Persian]
3. Saatz I, Kienle A. Increasing quality in large scale university courses.

Eleed. 2013; 9(1).
4. Cidral WA, Oliveria T, Di Felice M, Aparicio M. E-learning success determinants: Brazilian empirical study. Computers and Education. 2018; 122: 273-90. doi:10.1016/j.compedu.2017.12.001
5. Hadullo K, Oboko R, Omwenga E. Factors affecting asynchronous e-learning quality in developing countries university settings. International journal of Education and Development using ICT. 2018; 14(1): 152-63.
6. Masoumi D, Lindstrom B. Quality in e-learning: a framework for promoting and assuring quality in virtual institutions. Journal of Computer Assisted Learning. 2012; 28(1): 27-41. doi:10.1111/j.1365-2729.2011.00440.x
7. Manavifar L, Jamali J. Advantages and Disadvantages of Blended Teaching Method of Online and Face to Face for Practical Course of Hematology the Viewpoints of Laboratory Sciences Students in Mashhad University of Medical Sciences in 2011. Iranian Journal of Medical Education. 2012; 12(8): 619-28. [In Persian]
8. Keshavarz M, Rahimi M, Esmaeili Z. The Effect of e-Learning on the Academic Development of University Students. jms. 2013; 1(2) :13-21. [In Persian]
9. Haghani F, Rezaei H, Beigzadeh A, Eghbali B. Flipped Classroom: A Pedagogical Method. Iranian Journal of Medical Education. 2016; 16: 104-19. [In Persian]. PMCID: 27583990
10. Beigzadeh A, Rahimi M. The Use of Games in Medical Education. Res Dev Med Educ. 2015; 4(1): 1-2. doi: 10.15171/rdme.2015.001
11. Beigzadeh A, Haghani F. Active Learning Methods: A Way Of Tackling Large Classroom Setting, Perspective. Strides Dev Med Educ. 2016; 13(1): 107-13.
12. UNESCO. COVID-19 Educational disruption and response. [Cited 2020 March 25]. Available from: https://en.unesco.org/themes/education-emergencies/coronavirus-school-closures.
13. Gewin V: Five tips for moving teaching online as COVID-19 takes hold . Nature. 2020; 580: 295-6. doi:10.1038/d41586-020-00896-7 PMid:32210377
14. Lau J, Rudrani Dasgupta B Y. Will the coronavirus make online education go viral?. [Cited 2020 March 24]. Available from:https://www.timeshighereducation.com/features/will-coronavirus-make-online-educationgo-viral.
15. Sepasi H. A study of cognitive domain and analysis of psychometric characteristics of test items in final examinations of Arabic, calculus and biology courses of the third grade high school girls in different sicio-economic status in Khuzestan, Iran. Quarterly Journal of Education and Psychology 2007; 13(4): 57-78. [In Persian]
16. Baharvand M, Hoseinzadeh M, Jaberiansari Z, Abbaszadeh E, Mortazavi H. Evaluation of 30 Structural Validity and Content Quality Indices of Theoretical Dental Exams. Journal of Mashhad Dental School. 2014; 38(4): 291-302. [In Persian]
17. Emamjomeh MM, Zahedifar F. Analysis of multiple choice questions on health sciences at health faculty of Qazvin University of Medical Sciences (2009). J Qazvin Univ Med Sci. 2012; 15(4): 108-10. [In Persian]
18. Crisp G, Palmer E. Engaging Academics with a Simplified Analysis of their Multiple-Choice Question (MCQ) Assessment Results. Journal of University Teaching and Learning Practice. 2007; 4(2): 88-106.
19. Hosseini Teshnizi S, Zare S , Solati dehkordi S M. Quality analysis of multiple choice questions(MCQs) examinations of noncontinuous undergraduate medical records, Hormozgan Med J. 2010; 14(3): 177-83. [In Persian]
20. Seif A. Educational Measurement, Assessment and Evaluation. 3 rd ed. Tehran: Agah Co; 2007. 60-132. [In Persian]
21. Fealy S, Safar Pour S, RasouliAzar S. Effective factors on students' cheatingbehaviors on exams: A case study of Gorgan Islamic Azad University. IRPHE. 2014; 20(1): 57-77. [In Persian].
22. Hewitt-Taylor J. Facilitating distance learning in nurse education. Nurse Educ Pract. 2003; 3(1): 23-9. doi: 10.1016/s1471-5953(02)00052-5.

PMID: 19036314

23. Saeedinejat S, Vafaeenajar A. The Effect of E-Learning on Students' Educational Success. Iranian Journal of Medical Education. 2011; 11(1): 1-9. [In Persian]. PMID: 27012245 PMCID: PMC4807545

24. Thurmond VA. Defining interaction and strategies to enhance interactions in Web-based courses. Nurse Educ. 2003; 28(5): 237-41. doi: 10.1097/00006223-200309000-00013. PMID: 14506357

25. Okhovati M, YazdiFeyzabadi V, Beigzadeh A, Shokoohi M, Mehrolhassani MH. Evaluating the Program of Bachelor Degree in Health Services Management at Kerman University of Medical Sciences Iran Using the CIPP Model (Context Input Process Product). Strides Dev Med Educ. 2014; 11(1): 101-13. [In Persian].

26. Bazrafshan A, Haghdoost A, Rezaie H, Beigzadeh A. A practical framework for evaluating health services management educational program: The application of the mixed-method sequential explanatory design. Res Dev Med Educ. 2015; 4(1): 47-54. doi: 10.15171/rdme.2015.008

27. Beigzadeh A, Okhovati M, Mehrolhasani M H, Shokoohi M, Bazrafshan A. Challenges of the Bachelor Program of Health Services Management: A Qualitative Study. jha. 2014; 17(55): 29-42.