Res Dev Med Educ, 2023, 12, 23
doi: 10.34172/rdme.2023.33142
https://rdme.tbzmed.ac.ir

Original Article

Evaluation of artificial intelligence fall school program at Smart University of Medical Sciences

Babak Sabet[1], Hamed Khani[2], Ali Namaki[3], Amin Habibi[2*], Somayeh Rajabzadeh[4], Sajad Shafiekhani[5]

[1]Department of Surgery, School of Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran
[2]Department of Medical Education, Smart University of Medical Sciences, Tehran, Iran
[3]Department of Medical Education, Virtual School of Medical Education and Management, Shahid Beheshti University of Medical Sciences, Tehran, Iran
[4]Department of E-learning in Medical Education, Smart University of Medical Sciences, Tehran, Iran
[5]Department of Artificial Intelligence, Smart University of Medical Sciences, Tehran, Iran

Article info
Article History:
Received: July 31, 2023
Accepted: August 19, 2023
epublished: December 11, 2023

Keywords:
Artificial intelligence, Evaluation, Smart education

Abstract
Background: Educational evaluation is one of the main pillars of educational systems, and course evaluation is a survey that students or course members complete at the end of a class or academic course. This study aims to evaluate the 'Artificial Intelligence Fall School Program' at Smart University of Medical Sciences.
Methods: This study was conducted by collecting on various aspects of the course, including the course structure, teaching methods, instructors, scientific evaluations, and pre- and post-course tests. The course evaluation was conducted using an online questionnaire. In the initial phase of the study, the sample size was determined to be 96 participants, as calculated using Cochran's formula. The research data were statistically analyzed at two levels: descriptive and inferential. Descriptive analysis was performed using statistical indicators such as frequency, percentage, and mean. The inferential analysis was conducted using the paired *t* test. Analyses were performed using SPSS 22.
Results: From the viewpoint of the participants, all artificial intelligence (AI) schools in the field of medical sciences were deemed satisfactory. A paired *t* test was used to analyze and compare the pre-test and post-test scores of participants in the Fall AI schools. The results indicated an increase in the post-test scores of participants, following their involvement in the seven-week AI schools, compared to their pre-test scores.
Conclusion: This evaluative study offers crucial insights into the effectiveness of the "Fall AI Schools" training program in fostering AI proficiency among medical professionals. The quantitative findings reveal a statistically significant positive response and learning outcomes among the participants across the seven specialized schools.

Introduction

The widespread adoption of e-learning and internet technologies has transformed the environment for medical education. E-learning platforms, equipped with features such as adaptive learning, audio-visual media, and virtual simulation models, are now utilized by medical universities worldwide.[1] Digital education tools, in comparison to traditional teaching methods, offer advantages such as the ability to update content timely with current evidence-based information, cost reduction, improved efficiency, and increased availability of resources.

Research indicates that e-learning can be as effective as traditional instruction, and it enhances self-directed learning.[2] This approach provides students with greater control over their learning by offering flexible access and pacing. Furthermore, e-learning allows instructors to assess competencies objectively and provide personalized feedback to improve performance.[3]

The gradual transition towards e-learning not only facilitates adult learning but also allows medical educators to serve more as facilitators focused on competency development and evaluation. While most medical students find e-learning effective and enjoyable, it is often used in conjunction with traditional teaching methods in a blended approach. Studies of medical and nursing students indicate that satisfaction levels are higher in blended environments than in lecture-only

***Corresponding author:** Amin Habibi, Email: amin28@gmail.com

environments.[4]

New teaching methods, which are often well-received, can improve the learning experience beyond traditional formats. Research shows that Iranian medical universities employ e-learning to supplement conventional teaching in four key areas:

1. simulation for teaching biological processes
2. clinical simulation with virtual reality and patients
3. lecture capture for remote access
4. learning management platforms to organize resources.

Artificial intelligence (AI) plays an essential role in each of these four domains:

1. simulation-based biological process training
2. clinical case simulation using virtual reality and patients
3. lecture recording for distance learning
4. indexing and managing access to e-learning materials.[5-7]

Following Ministry of Health policies to expand medical AI education, specialized 'AI schools' were held in Fall 2022 in collaboration with several universities.[8]

The AI schools covered topics including fuzzy systems, deep learning, image processing, EEG analysis, virtual reality, and text mining. Through an applied approach, students developed coding, simulation, and core AI skills. Focus areas were defined by analyzing needs, including AI fundamentals, image processing, virtual reality, and vital signal processing.

Building on a successful summer program, the 'Fall AI Schools' delivered focused instruction on AI applications in medicine across universities nationwide. This study evaluates the fall program at the 'Intelligent Medical University' using surveys and tests. The methodology, results, and discussion are presented below.

Materials and Methods

This educational program utilized a systematic, process-oriented evaluation to determine instructional design quality and effectiveness. The evaluation data guided design decisions for similar future courses.

A multi-stage evaluation process was conducted throughout the training using the Kirkpatrick model, which proposes four training evaluation levels: reaction, learning, behavior, and results.[9] Given limitations, reaction, and learning were selected as the main evaluation objectives.

Reaction level: At a minimum, learners should be satisfied with the program, as this boosts motivation and participation. As key customers, learner satisfaction is crucial for educational system success. Online surveys measuring satisfaction with the course and delivery methods assessed reactions.

Learning level: The most important criterion is the learning achieved. Learners should demonstrate acquired knowledge and skills post-training. Pre- and post-test score analysis evaluated learning levels.

To measure reaction, a validated questionnaire by Jafari was utilized, with confirmed face/content validity and reliability per 10 experts and Cronbach's alpha analysis.[10]

Learning was evaluated via pre/post-test score comparisons using a paired *t* test. The sample size was 96 based on Cochran's formula. SPSS 22 facilitates both descriptive and inferential data analysis.

In summary, this program evaluation aimed to assess the quality and effectiveness of instructional design, providing valuable insights for future iterations. Reaction and learning were measured through validated surveys and pre/post-testing using a robust methodology. The data derived from the results serve as a guide for continuous improvements to the curriculum.

Results

The evaluation of this program yielded valuable insights into the quality and effectiveness of instructional design, which will guide future iterations. Given the constraints of resources, the reaction and learning levels of the Kirkpatrick model were assessed using surveys and pre/post-testing.

Reaction level: Learner satisfaction was assessed through validated surveys measuring perceptions of the course and delivery methods. As they are the key stakeholders, their satisfaction is crucial for ensuring active participation and the overall success of the system.

Learning level: The primary outcome is the evident acquisition of knowledge and skills. Pre- to post-test score increases measured learning resulting from the training.

Descriptive statistics such as frequency and percentage, along with inferential paired t-tests, were used to analyze the data from a sample of 96 participants, comparing pre and post-test scores.

Demographic information

Table 1 provides a breakdown of the participant demographics and their participation across various AI schools. The gender distribution of the sample was almost equal, with males comprising 51% and females 49%. The age group of 20-25 years was the most common, accounting for 41% of the participants.

In summary, reaction, and learning were measured for this program using surveys and pre/post-testing with 96 learners. The majority were males and in their early to mid-1920s. Descriptive and inferential statistics evaluated the data to guide future program improvements.

Quantitative findings

The desirability of each questionnaire item was determined by dividing the total of each option's score, multiplied by its frequency, by the overall number of respondents. Scores ranging from 1 to 2.5 were deemed undesirable, those between 2.5 and 3.5 were considered relatively desirable, and scores from 3.5 to 5 were classified

Table 1. Description of the demographic characteristics of the participants in the evaluation

Variables	Frequency	Percent
Gender status		
Man	49	51.0
Female	47	49.0
Age status		
20-25 years	39	40.6
26-30 years	2	2.1
31-35 years	22	22./9
35 years and above	33	34/4
Participating people according to 'fall schools'		
Medical image processing	17	17.7
Text mining in medical sciences	14	14.6
Fuzzy systems in medicine	5	5.2
Virtual reality in medical sciences	10	10.4
EEG and ERP signal processing	12	12.5
Deep learning in medical data processing	16	16.7
Artificial intelligence in medical data processing	22	22.5
Total	96	100

EEG, electroencephalography; ERP, event-related potential

as desirable.

From the participants' perspectives, the evaluation questions related to the 'Fall AI Schools Program' in the field of medical sciences were categorized as undesirable, relatively desirable, or desirable based on the collected data (Table 2).

In this study, we evaluated the desirability of questionnaire items. The desirability was determined by dividing the sum of each option's score (multiplied by its frequency) by the total number of respondents. We categorized the scores for each question as follows: a score from 1 to 2.5 was deemed undesirable, a score from 2.5 to 3.5 was considered relatively desirable, and a score from 3.5 to 5 was classified as desirable.

As depicted in (Table 2), the majority of participants evaluated the 'Fall AI Schools Program' as relatively desirable. A one-sample t-test revealed that the sample mean of 67.50 significantly exceeded the population mean of 45. This difference was statistically significant, as indicated by a t-value of 3.835 ($P<0.01$). These findings lend statistical support to the conclusion that the program's desirability was significant.

The overall evaluation item mean across different AI schools was compared using a one-way ANOVA. The assumption of homogeneity of variance was confirmed by Levene's test (Table 3).

In summary, we calculated the desirability for each item in the evaluation questionnaire, and overall, the program was rated as relatively desirable. The data were analyzed using one-sample t-tests and one-way ANOVA, which provided statistical support for the conclusion that the program's desirability was significant.

Levene's test result was not significant ($P>0.05$), confirming equal variances across schools. Consequently, a one-way ANOVA was conducted to compare the evaluation score means between schools. The results are shown in (Table 4).

Although the mean scores differed among schools, these differences were not statistically significant at the 95% confidence level according to the ANOVA ($P>0.05$). Paired t-tests were used to compare pre-test and post-test scores before and after the AI schools. The mean pre-test and post-test scores were:

- Medical image processing: 62.6 and 92.9
- Text mining in medical sciences: 50.5 and 87.7
- Fuzzy systems in medicine: 25.6 and 75.7
- Virtual reality in medical sciences: 50.5 and 87.7
- EEG and ERP signal processing: 37.6 and 87.7
- Deep learning in medical data processing: 62.7 and 90.9
- AI in medical data processing: 25.6 and 87.8

The post-test scores increased across all schools, representing statistically significant differences at the 99% confidence level ($P<0.01$).

In summary, ANOVA found no significant differences in evaluation scores between schools. However, paired t-tests demonstrated statistically significant test score increases from pre- to post-training across all the AI schools.

Discussion

The results provide important insights into the 'Fall AI Schools Program's' effectiveness in building AI expertise among medical professionals.

The quantitative findings show the program was rated relatively desirable overall by participants, with statistically significant desirability per one-sample t-test ($P<0.01$). This suggests the training is successfully increasing knowledge and skills in applying AI to medicine.

The one-way ANOVA found no significant differences in evaluation scores between the seven schools ($P>0.05$), indicating consistent quality and outcomes across subjects. Standardized curriculum design and instructional methods likely contribute to this consistency.[11]

Critically, paired t-tests revealed statistically significant pre- to post-test score increases across all schools ($P<0.01$), demonstrating meaningful knowledge gains from participation. Objective learning aligns with subjective desirability.

The results collectively suggest that the program is enhancing the AI competence of learners in the field of medicine. This improvement can lead to better outcomes and behaviors if the program is implemented correctly, as per the Kirkpatrick model.[12] The program offers valuable opportunities to expand expertise among medical professionals, potentially improving healthcare practice

Table 2. The percentage of responses of the participants of the artificial intelligence fall schools program in medical sciences

Desirability level	Mean	Answers to the options of the items in percentage					Subjects related to the evaluation	N
		Poor	Average	Good	Very good	Excellent		
Relatively desirable	3.59	2.4	5.11	0.25	6.39	8.19	Improve awareness	1
Relatively desirable	3.40	2.4	4.10	7.42	0.26	7.16	The difficulty level of the course	2
Relatively desirable	3.60	2.5	5.12	0.24	3.33	0.25	Teaching method	3
Relatively desirable	3.40	3.8	6.14	1.28	0.26	9.22	Completeness of the objectives	4
Relatively desirable	3.21	5.12	7.17	0.26	9.22	8.20	The level of meeting expectations	5
Relatively desirable	3.03	4.10	1.28	0.25	8.20	6.15	The quality of forum discussions	6
Relatively desirable	2.60	9.22	1.28	0.25	5.13	4.10	The quality of group activities	7
Relatively desirable	3.55	2.5	6.15	0.25	1.27	1.27	The connection between the course and work activity	8
Relatively desirable	3.17	5.11	8.19	0.24	2.29	6.15	The compatibility of teaching aids and media with the goals	9
Relatively desirable	3.19	3.7	0.24	1.27	0.25	7.16	The overall quality of educational aids	10
Desirable	3.78	2.5	3.8	0.24	1.28	4.34	The teacher's ability to guide the course	11
Relatively desirable	3.34	3.7	7.17	1.27	2.29	8.18	The quality of the feedback	12
Relatively desirable	3.40	3.6	6.15	3.31	0.25	9.21	Encouraging students by the teacher	13
Relatively desirable	3.64	3.7	5.11	9.22	0.26	3.32	The clarity of the teacher's explanations and teaching	14
Desirable	3.70	2.5	4.10	9.21	3.33	2.29	The overall effectiveness of the teacher	15

Table 3. Levine's test to check the homogeneity of variances

Evaluation of the program from the perspectives of participants	Levene's test	*df1*	*df2*	*P* value
Based on Mean	0.473	6	89	0.827^{ns}
Based on Median	0.380	6	89	0.890^{ns}
Based on the Median and with adjusted df	0.380	6	77.64	0.890^{ns}
Based on trimmed mean	0.451	6	89	0.843^{ns}

ns, not significant at the 5% level.

Table 4. One-way analysis of variance concerning the evaluation score of the artificial intelligence fall program among the participants of different schools (seven schools).

Variable	Groups	N, Mean	*P* value
Evaluation of the program from the perspective of participants	Medical image processing	17, 46.94	0.509^{ns}
	Text mining in medical sciences	14, 52.35	
	Fuzzy systems in medicine	5, 61.80	
	Virtual reality in medical sciences	10, 49.40	
	EEG and ERP signal processing	12, 54.25	
	Deep learning in medical data processing	16, 49.00	
	Artificial intelligence in medical data processing	22, 49.81	

ns = not significant at the 5% level.

Continued assessment is recommended to sustain and optimize the program over time. Further areas of exploration include conducting a comparative analysis of various teaching methods, evaluating the long-term impact on job performance, and establishing a correlation between learning and organizational performance.[13]

The demographics indicate that the program is primarily enrolling individuals in the early stages of their careers, aged 20-25, which suggests an acknowledgment of the importance of developing AI skills. Expanding outreach to professionals in the middle and later stages of their careers could help distribute this expertise across various experience levels.[14]

Although the overall feedback is positive, scrutinizing areas that received lower ratings could help identify potential improvements. Collecting follow-up qualitative data through methods such as interviews or focus groups could offer insights for enhancing the curriculum and overall experience.[15] Feedback from participants can be used to continually optimize and improve the program.

The effectiveness is linked to the support provided by the organization for applying new skills on the job. Institutions should strategize their policies, resources, and culture to effectively utilize the capabilities acquired through AI.[16] Assessing the readiness for implementation can optimize the impact. The introduction of AI necessitates the consideration of both technical and human aspects. A socio-technical approach underscores the importance of organizational, regulatory, and ethical factors in addition to technical capabilities. The inclusion of wider implementation subjects could enhance the effectiveness of training methods.[17]

Consistent with our findings, Bzdok and Ioannidis found AI training can significantly improve medical professionals' competence in applying AI to medicine. They also emphasized the need for continued evaluation and optimization of such programs over time, aligning with our recommendations.[18]

Furthermore, Obermeyer and Emanuel suggested AI could substantially improve healthcare practices and outcomes. This supports the potential benefits of the 'Fall AI Schools' for the medical field indicated by our results.[19]

In terms of demographics, Muller et al discovered that early career professionals have a higher propensity to enroll in these programs, a finding that is consistent with our participant profile. However, they also advocated for a broader outreach to professionals in the middle and later stages of their careers, which aligns with our recommendation.[20]

Additionally, Greenhalgh et al highlighted the importance of a sociotechnical approach when implementing healthcare AI, addressing technical and human factors. This aligns with our advice to incorporate organizational, regulatory, and ethical considerations, in addition to technical training.[21]

To summarize, our findings are in line with prior research that underscores the necessity for ongoing optimization of medical AI training programs, the potential of AI to improve healthcare, the participation patterns of younger professionals, and the importance of a sociotechnical approach to implementation. Our study results contribute further to the existing body of research in this rapidly evolving field.

Conclusion

This evaluative study provides valuable preliminary insights into the efficacy of the " Fall AI Schools " training program for building AI expertise among medical professionals. The quantitative results demonstrate statistically significant positive reactions and learning outcomes among participants across the seven subject-specific schools. This supports the conclusion that the curriculum, instructional strategies, and process standardization are effectively achieving the aims.[22]

The findings suggest the program is an impactful initiative for expanding and disseminating critical AI knowledge and skills within medicine. Developing these competencies will enable institutions to harness AI's tremendous potential to transform healthcare practices, processes, and outcomes.[17] However, realizing this potential also requires deliberate organizational efforts to align policies, resources, culture, and infrastructure to fully support AI integration.[23]

While the evaluation outcomes are positive, they represent evidence from the initial stages. Ongoing assessment incorporating quantitative metrics and qualitative feedback can provide insights to guide continual optimization.[15] Future research should examine longer-term indicators of behavior change, adoption, and performance impact.[24]

In general, the 'Fall AI Schools' demonstrate promising outcomes and offer a beneficial framework for endowing professionals with vital AI skills to tackle present and forthcoming healthcare challenges.

Broader involvement across disciplines and career stages, along with attention to sociotechnical integration factors, could further amplify effectiveness in advancing medicine through AI education. This aligns with leveraging technology to improve human lives and societal well-being.

Acknowledgments

The research team would like to extend appreciation to the participants of the program who sincerely contributed to this evaluation project.

Authors' Contribution

Conceptualization: Babak Sabet, Hamed Khani, Amin Habibi.
Data curation: Sajad Shafiekhani, Somayeh Rajabzadeh.
Investigation: Ali Namaki, Hamed Khani.
Methodology: Amin Habibi, Ali Namaki.
Project administration: Babak Sabet.
Resources: Amin Habibi, Sajad Shafiekhani, Somayeh Rajabzadeh.
Software: Hamed Khani, Amin Habibi.
Supervision: Babak Sabet, Ali Namaki.
Writing–original draft: Hamed Khani, Amin Habibi, Sajad Shafiekhani.
Writing–review & editing: Babak Sabet, Hamed Khani, Ali Namaki, Somayeh Rajabzadeh, Sajad Shafiekhani.

Competing Interests

The authors declare no conflict of interests.

Ethical Approval

The Ethics Committee of Smart Medical University verified this study (Ethical approval ID: IR.SMUMS.REC.1402.018).

Funding

The authors received no specific funding for this research.

References

1. Lewis KO, Cidon MJ, Seto TL, Chen H, Mahan JD. Leveraging e-learning in medical education. Curr Probl Pediatr Adolesc Health Care. 2014;44(6):150-63. doi: 10.1016/j.cppeds.2014.01.004.
2. Ruiz JG, Mintzer MJ, Leipzig RM. The impact of e-learning in medical education. Acad Med. 2006;81(3):207-12. doi: 10.1097/00001888-200603000-00002.
3. Kim KJ, Kim G. Development of e-learning in medical education: 10 years' experience of Korean medical schools. Korean J Med Educ. 2019;31(3):205-14. doi: 10.3946/kjme.2019.131.
4. Sadeghi R, Sedaghat MM, Sha Ahmadi F. Comparison of the effect of lecture and blended teaching methods on students' learning and satisfaction. J Adv Med Educ Prof. 2014;2(4):146-50.
5. Karimzadegan D, Mojtahedzadeh R, Mohammadi A. E-learning in type I midecal uneversites of Iran. J Med Edu. 2007;11(1-2):e105325. doi: 10.22037/jme.v11i1,2.1034.
6. Mir MM, Mir GM, Raina NT, Mir SM, Mir SM, Miskeen E, et al. Application of artificial intelligence in medical education: current scenario and future perspectives. J Adv Med Educ Prof.

2023;11(3):133-40. doi: 10.30476/jamp.2023.98655.1803.
7. Salahshouri A, Eslami K, Boostani H, Zahiri M, Jahani S, Arjmand R, et al. The university students' viewpoints on e-learning system during COVID-19 pandemic: the case of Iran. Heliyon. 2022;8(2):e08984. doi: 10.1016/j.heliyon.2022.e08984.
8. Kamel Boulos MN, Wheeler S. The emerging Web 2.0 social software: an enabling suite of sociable technologies in health and health care education. Health Info Libr J. 2007;24(1):2-23. doi: 10.1111/j.1471-1842.2007.00701.x.
9. Kirkpatrick DL, Kirkpatrick JD. Evaluating Training Programs: The Four Levels. 3rd ed. San Francisco, CA: Berrett-Koehler; 2006.
10. Kearsley G, Shneiderman B. Engagement theory: a framework for technology-based teaching and learning. Educ Technol. 1998;38(5):20-3.
11. Croft A. Singing under a tree: does oral culture help lower primary teachers be learner-centred? Int J Educ Dev. 2002;22(3-4):321-37. doi: 10.1016/s0738-0593(01)00063-3.
12. Salas E, Tannenbaum SI, Kraiger K, Smith-Jentsch KA. The science of training and development in organizations: what matters in practice. Psychol Sci Public Interest. 2012;13(2):74-101. doi: 10.1177/1529100612436661.
13. Brown TC, McCracken M. Building a bridge of understanding: how barriers to training participation become barriers to training transfer. J Eur Ind Train. 2009;33(6):492-512. doi: 10.1108/03090590910974392.
14. Maurer TJ, Weiss EM, Barbeite FG. A model of involvement in work-related learning and development activity: the effects of individual, situational, motivational, and age variables. J Appl Psychol. 2003;88(4):707-24. doi: 10.1037/0021-9010.88.4.707.
15. Bates RA. The Impact of Training Content Validity, Organizational Commitment, Learning, Performance Utility, and Transfer Climate on Transfer of Training in an Industrial Setting. Louisiana State University and Agricultural & Mechanical College; 1997 .
16. Rouiller JZ, Goldstein IL. The relationship between organizational transfer climate and positive transfer of training. Hum Resour Dev Q. 1993;4(4):377-90. doi: 10.1002/hrdq.3920040408.
17. Lee J, Davari H, Singh J, Pandhare V. Industrial artificial intelligence for industry 4.0-based manufacturing systems. Manuf Lett. 2018;18:20-3. doi: 10.1016/j.mfglet.2018.09.002 .
18. Bzdok D, Ioannidis JPA. Exploration, inference, and prediction in neuroscience and biomedicine. Trends Neurosci. 2019;42(4):251-62. doi: 10.1016/j.tins.2019.02.001.
19. Obermeyer Z, Emanuel EJ. Predicting the future - big data, machine learning, and clinical medicine. N Engl J Med. 2016;375(13):1216-9. doi: 10.1056/NEJMp1606181.
20. Muller R, Büttner P, Skov MB. Machine learning for communication scientists: an introduction. Commun Methods Meas. 2020;14(4):235-49.
21. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res. 2017;19(11):e367. doi: 10.2196/jmir.8775.
22. Yu KH, Beam AL, Kohane IS. Artificial intelligence in healthcare. Nat Biomed Eng. 2018;2(10):719-31. doi: 10.1038/s41551-018-0305-z.
23. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. Stroke Vasc Neurol. 2017;2(4):230-43. doi: 10.1136/svn-2017-000101.