scientific reports

OPEN

Evaluation of medical education mobile apps with MARS and MARuL in a TPACK informed model

Mohammadsadegh Rezaei[1✉], Sara Owrangi[2✉] & Sahar Jafari[3]

The evaluation of medical education apps is paramount for ensuring their effective implementation, particularly in the post-COVID-19 landscape. The study aims to evaluate medical education mobile apps using two current and validated evaluation tools. It also explores how effectively these tools measure the mobile app's impact on the medical education outcomes. Following the screening of 409 records using the PRISMA method, 30 mobile apps in the field of medical education were evaluated employing the MARS (Mobile Application Rating Scale) and MARuL (Mobile App Rubric for Learning) assessment tools. A majority of mobile apps received acceptable scores: 20 mobile apps (67%) met the MARS threshold (> 0.3), and 22 mobile apps (73%) met the MARuL threshold (> 51). Findings indicate that while MARS and MARuL provide useful evaluations, both inadequately address pedagogical aspects within the TPACK model; MARS overlooks educational objectives, and MARuL omits key criteria such as cognitive development, practice targets, motivation, goal orientation, effective scaffolding, self-directed learning, curriculum oriented, critical thinking and authenticity.

Disease treatment is constantly advancing. Medical technology is increasingly integrated. Diagnostic and treatment approaches are growing more complex. Particularly in the post-COVID-19 landscape, public health maintenance adds another layer of complexity[1,2]. This requires more sophisticated medical education. Ongoing improvements in medical education are needed[3]. Information and communication technologies offer the possibility of educational experiences beyond the confines of the therapeutic environment, without the need for expensive therapeutic equipment. One crucial aspect of ICT technologies is the accessibility of learning opportunity through mobile apps for enhancing medical education[4,5]. Mobile phones are an ideal platform for hosting medical education mobile apps, as they offer improved educational opportunities both within and beyond traditional educational environments. These devices are readily available, portable, and can be easily connected to other electronic tools. Furthermore, their use does not require extensive training and is user-friendly[4,6].

Choosing the appropriate mobile app for medical education, which aligns with the necessary technical and functional requirements for both theoretical and clinical instruction, poses a significantly greater challenge. Hence, it is crucial to evaluate the quality of mobile apps in the field of medical education using suitable and reliable tools and criteria[6,7]. The purpose of this study is to evaluate a medical education mobile app using two valid and current assessment tools, and to critically appraise the ability of these tools to measure the effectiveness of the mobile apps in achieving medical education outcomes.

The research questions are as follows: What scores do medical education mobile apps receive when evaluated with the Mobile App Rating Scale (MARS) tool and the recently developed Mobile App Rubric for Learning (MARuL) tool, designed specifically for clinical education mobile apps? To what extent do these scores reflect the quality of the mobile apps in delivering effective education? Which criteria should be integrated into the MARS and MARuL tools to improve their effectiveness in evaluating medical education mobile apps based on the TPACK model?

The central hypothesis is that the qualitative features of medical education mobile apps are not fully covered by existing tools. Previous evaluations using the MARS tool were conducted by experts with both medical knowledge and experience in educational technologies. These studies found that many mobile apps received acceptable MARS scores. However, experts rated them poorly in their subjective assessments of usefulness for

[1]Department of Computer Engineering and Information Technology, Shiraz University of Technology, Shiraz, Iran. [2]Medical School, Shiraz University of Medical Science, Shiraz, Iran. [3] Department of Computer Engineering, Sharif University of Technology, Tehran, Iran. ✉email: rezaei@sutech.ac.ir; sara.owrangi1998@gmail.com

medical education. By contrast, all mobile apps that scored acceptably on the subjective quality criterion also achieved acceptable scores on MARS. This suggests that the MARS tool alone is insufficient for comprehensive qualitative evaluation of medical education mobile apps[8]. In this study, we extended the evaluation by applying a correlation test between mobile app scores on the newly developed MARuL tool and the subjective quality criterion.

The structure of this article is as follows: In the next section, a review of previous research in the field of evaluating mobile apps and the tools used is discussed. The third part of the study introduces the TPACK model, providing the conceptual foundation for reviewing the performance of MARS and MARuL in evaluating medical education mobile apps. The fourth part describes the research methodology The fifth section presents the results of the evaluation and ranking of mobile apps. The sixth section discusses the results and findings of the research. Finally, the last section concludes the article.

Literature review

Articles[9–13] evaluate mobile apps addressing chronic, postoperative, and menstrual pain. All of these articles used the PRISMA method for mobile app selection. Articles[11,12] employ the MARS tool for evaluation, while article[13] evaluates mobile apps based on content and functionality criteria. Articles[14,15] evaluate mobile apps related to Depression. These articles utilized the PRISMA method for mobile app selection and employed the MARS tool for evaluation. Articles[16,17] evaluate mobile apps related to Weight Management. Both articles used the PRISMA method for mobile app selection. Article[17] further evaluated mobile apps based on evidence-based strategies, healthcare expert involvement, and scientific evaluation. Articles[7,18] evaluate mobile apps related to Clinical Skills Training. These articles used the PRISMA method for mobile app selection and employed both the MARS and MARuL tools for evaluation. Article[19] evaluate mobile apps for Family Caregivers. Evaluation criteria included cost (purchase and in-app), device requirements (Fitbit, iHealth, tablets, etc.), interoperability (platform, web-app, EMR/EHR integration), disease-specific focus (e.g., dementia care), target caregiver age group, evaluation claims, and functions. Article[20] evaluated mobile apps for inflammatory bowel disease management using the MARS tool. Article[21] examines mobile apps designed for patients with chronic obstructive pulmonary disease (COPD) who are receiving home oxygen therapy. The study evaluates the mobile apps' usability from the patient perspective and their efficiency from the perspective of healthcare experts. Article[22] examined Brazilian mobile apps focused on the care of patients with anxiety disorders. The MARS tool was employed for evaluation, and the PRISMA method was used for mobile app selection. Articles[23] evaluated health and fitness mobile apps[24], support patients undergoing bariatric surgery mobile apps[25], patients managing bipolar disorders mobile apps[26], cancer mobile apps[27], cardiovascular disorders mobile apps[28], diabetes mobile apps[29], prevention of healthcare-associated infections mobile apps[30,31], viral immunodeficiency mobile apps[32], conservative dentistry and endodontics[33], psychological skills training[34], Schizophrenia mobile apps[15], smoking cessation mobile apps using the MARS tool.

Previous research has emphasized the importance of evaluating mobile apps in the fields of health and education. One of the key areas of focus in evaluating health-related mobile apps is the content. Consequently, all evaluation tools developed for these mobile apps include criteria that evaluate their content. Other important evaluation factors commonly considered in research include user-friendliness and applicability, as well as criteria related to the mobile app's aesthetics. To ensure unbiased results and comprehensive evaluation outcomes, the researchers primarily utilized the PRISMA framework for mobile app selection and evaluation. This framework helps reduce bias and ensures a comprehensive evaluation process. Additionally, in many studies, the MARS evaluation tool has been widely used for evaluating mobile apps. This tool is renowned for its simplicity, reliability, and objective goals. It has been utilized in numerous research projects within the health field. The tool's reliability and validity have been reported as satisfactory in various studies.

TPACK model

The TPACK model, developed by Koehler and Mishra (2008)[35], integrates three dimensions pedagogical knowledge (PK), technological knowledge (TK), and content knowledge (CK) to guide effective use of ICT in teaching (Fig. 1). PK denotes teachers' understanding of evidence-based instructional methods, TK represents the skills to deploy technology, and CK denotes expertise in a specific subject area. The intersections of these dimensions, form three constructs: technological pedagogical knowledge (TPK), technological content knowledge (TCK), and pedagogical content knowledge (PCK). TPK involves applying technology to support pedagogical approaches, TCK focuses on using specific technologies within a subject, and PCK represents the integration of pedagogy and content, such as teaching mathematics effectively to students.

Research method

In this study, medical education mobile apps were evaluated and ranked according to their importance. To ensure a comprehensive search for relevant mobile apps, the researchers utilized the PRISMA framework, which has been widely used in reputable medical research[36–40]. PRISMA, a reporting guideline for systematic reviews and meta-analyses, provides evidence-based standards to improve transparency. The framework consists of four phases: identification, screening, eligibility, and inclusion[41]. The research method employed in this study is a systematic review, designed and implemented based on the PRISMA framework. The PRISMA flow diagram is shown in Fig. 2.

Selection of mobile apps

The mobile app review was conducted from September 2023 to January 2025 at Shiraz University of Technology, Shiraz, Iran. Mobile apps were identified using Google Search, restricting results to the Google Play Store with the "site:" operator, complemented by direct searches within the Play Store. To reduce bias, all authors performed
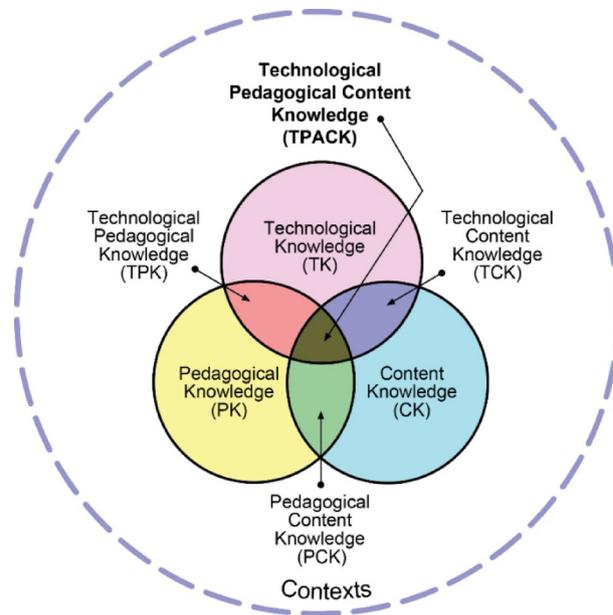
Fig. 1. The TPACK Model.Source: http://tpack.org.

parallel searches at different times and from distinct IP addresses. Duplicate mobile apps were removed. Multiple search terms, including "Medical Education," "Medicine Education," "Medical Learning," "Medicine Learning," and "Medical Student Guideline," were employed to locate relevant mobile apps.

Initially, titles and descriptions were screened for relevance to medical education. Mobile apps classified as computer games, unrelated to medical education, or presented in languages other than English were excluded. In the second stage, duplicates from different search terms and similar versions (e.g., "pro" or "lite") were removed. Mobile apps not designed for medical students, those aimed at patient or paramedic training, and mobile apps with limited availability were also excluded (see Fig. 2).

We included mobile apps that (a) deliver content in English, (b) are freely accessible without requiring payment, and (c) are applicable to at least one medical education course. Mobile apps were excluded if they (a) were games intended for children, (b) targeted nursing education, (c) only provided a demo version, or (d) could not be installed.

Mobile app quality assessment tool

In this study, the mobile app was qualitatively evaluated using the MARS and MARuL tools.

The MARS offers a comprehensive evaluation of mobile app quality across five sub-criteria: Engagement (e.g., Entertainment, Interest, Customization, and Interactivity), Functionality (e.g., Ease of use, Efficiency, Navigation, and Gestural design), Aesthetics (e.g., Layout, Graphics, and Visual appearance), Information quality (e.g., Accuracy, Objectives, Quality, Quantity, Visual information, Credibility, and Evidence-based content), and Subjective app quality. It consists of 19 items, The mean score across all sub-criteria reflects overall mobile app quality, while subscale averages scores highlight specific strengths and weaknesses[39]. A score ≥ 3 is considered the threshold for acceptable quality[32,33,42]. The tool demonstrates good interrater reliability (intraclass correlation coefficient (ICC) = 0.79) and strong internal consistency (Cronbach's α = 0.9)[39].

The MARuL tool evaluates mobile apps across four sub-criteria: Teaching and learning (e.g., Purpose, Pedagogy, Learning capacity, Quantity of information, Relevance to study/course, Instructional features, User interactivity, Feedback, Efficiency), User-Centered (e.g., Subjective quality, Satisfaction, Perceived usefulness, Perceived importance, User experience, Intention to reuse, Engagement) and Professional (e.g., In line with professional standards, Credibility of mobile app, Information quality) and Usability (e.g., Aesthetics, Functionality, Differentiation, Ease of use, Absence of advertisements, Technical specifications, Advantages over web-based or conventional equivalent)[7]. The tool includes 26 items scored on a 0–4 scale, with total scores categorizing mobile apps as not at all valuable (< 50), potentially valuable (51–69), or probably valuable (> 69). MARuL demonstrates good interrater reliability (ICC = 0.66), and excellent internal consistency (Cronbach's α = 0.96)[7]. and its content validity has been verified by experienced educational experts.

In Table 1, the evaluation criteria for educational mobile apps were categorized according to the TPACK model, based on the initial reference table proposed in[43]. The correspondence of items from the MARS and MARuL tools with each criterion was also determined. The results show that, within the Technology dimension, the MARS tool covers 11 of 18 criteria (61%), while the MARuL tool covers 12 of 18 criteria (67%). In the Pedagogy dimension, MARS addresses 2 of 11 criteria (18%), compared with 5 of 11 (45%) in MARuL. For the Content dimension, MARS includes 5 of 11 criteria (45%), whereas MARuL covers 7 of 11 (67%). Table 2 presents a comparative analysis of the similarities between MARS and MARuL.
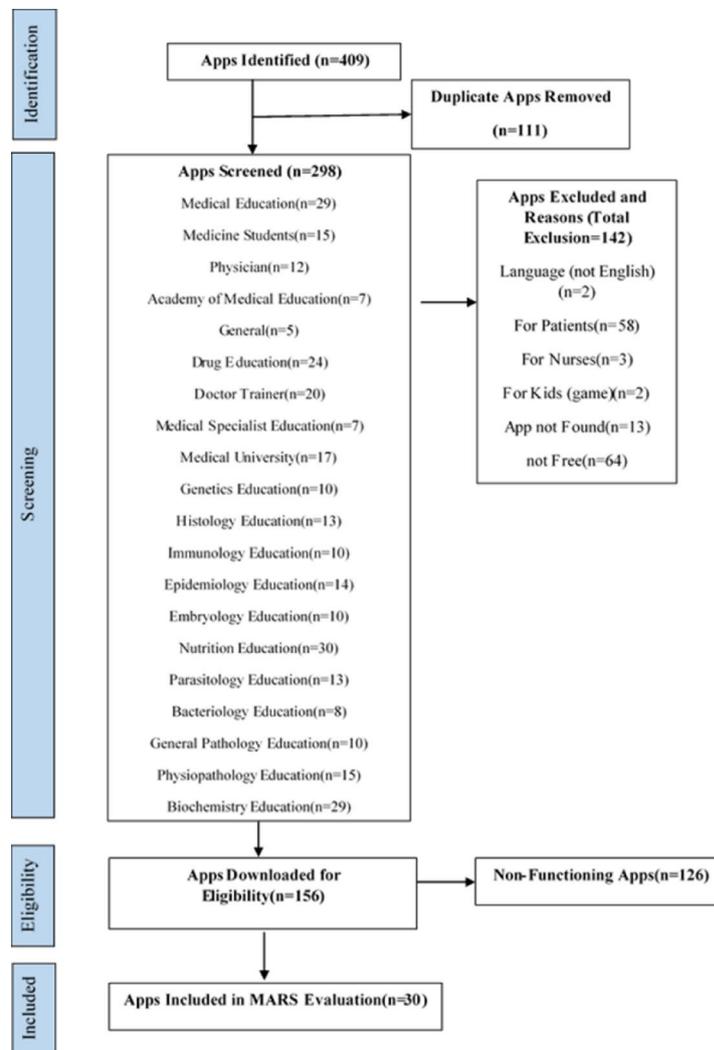
Fig. 2. PRISMA 2020 Flow Diagram for the Selection of Medical Education Mobile Apps.

Mobile app review process

For this research, four expert evaluators were invited to collaborate, all of whom held at least a professional doctorate in the field of medicine. The reviewers disclosed no conflicts of interest pertaining to the industry of the developers whose mobile apps were being evaluated.

Four evaluators received necessary guidance on the content and objectives of the Mobile Application Rating Scale (MARS) during an online session conducted by IT experts. Table 3 details the evaluators and the tools used for the evaluation. These experts possessed at least a master's degree and had 5 years of experience in developing mobile apps. An educational video was prepared and made available to the evaluators for reference. these experts have undergone 10 h of training on how to use the MARS and MARuL tools and actively participated in the testing phase of utilizing this tool. The MARS training video, provided by its developers, was used, while the MARuL training video was created by the researchers to train evaluators.

Each evaluator independently installed and reviewed the app for at least 30 min. The evaluators unanimously agreed on the relevance of all evaluation tools measures for evaluating educational mobile apps.

Before rating the assigned mobile apps, all evaluators evaluated two mobile apps specifically selected for educational purposes (previously excluded from the analysis). They discussed their results to ensure a shared understanding of the MARS and MARuL items and the evaluation process.

All mobile apps were accessible to all evaluators for review. The mobile apps were downloaded and rated between November 1 and December 30. Furthermore, information was collected regarding the target group, educational domain, type of content and learning activities, developer, and current mobile app store rating of each mobile app. The availability of scientific studies was verified through Google Scholar, PubMed, and the developer's website.

Statistical analysis of data

To evaluate the interrater reliability of MARS and MARuL sub-criteria and the total score, the third form of ICC (known as "Two-Way Mixed" in SPSS) was employed in this research[57]. It should be noted that the evaluators

TPACK Dimensions	Metric		MARS	MARuL	Definition
Technology	Options for user support			MuD6	Provide technical support to enhance user experience with the app, including tutorials, search functionality, help resources, and automatic updates[7,46–48]
	Accessibility				Ensure the mobile app is accessible across smartphones with all essential features, requiring minimal cognitive effort, while complying with web and multimedia accessibility standards[48,49].
	Design	Layout	MaC1	MuD1	Ensure an aesthetically pleasing and user-friendly design with a clear, consistent layout, high-quality media, and intuitive navigation. Use bright colors for learners, appropriate fonts and sizes, well-structured touch-screen menus, colorful animated buttons, and icons that accurately represent their functions[45–47,49,50]
		Graphics	MaC2	MuD1	
		Visual appeal	MaC3	MuD1	
		Ease of use	MaB2	MuD4, MuD1	
	Purpose		MaA5	MuA1	Ensure well-defined objectives with coherence between goals, skills, and target audience, supported by appropriate content and guidance for teachers and students[46,49].
	User control				Users should be able to select and sequence tasks within the mobile app, with the freedom to make choices throughout the mobile app[47,51].
	Stability				Assess whether the mobile app experiences freezing or crashes[46].
	Portability			MuD7	The mobile app should function seamlessly across platforms[52].
	Multimodal options				Evaluate the role of embedded visual and verbal features in motivating and sustaining student engagement, while ensuring they do not distract from learning or serve merely as entertainment[53]
	Functionality		MaB4	MuB5	We can fully utilize all of its features
	Communication				Enable learners to share mobile app-generated content, as well as their learning progress, issues, or concerns, with others via email or print at any time[47,50,51,54–56].
	Performance		MaB1	MuD2	Measure the mobile app's loading speed, responsiveness, and task completion efficiency[47]
	Gamification		MaA1		The process of integrating game-like elements into mobile apps[46].
	Interoperability				Developed in standard formats for universal compatibility, including complete metadata and accessibility features, and packaged for easy transfer and reuse[49].
	Navigations		MaB3	MuD4	The ease with which students can navigate the app independently, supported by clearly labeled menus and icons, functional status links, and clear instructions[46,47,49,53–56].
	Working mode				Determine whether the mobile app requires an internet connection to function[46].
	Error prevention		MaB1	MuD2	Facilitate error prevention and recovery by providing concise error messages[47,51].
	Reusability				High modularity, enabling reuse across disciplines and learning environments, including face-to-face, online, and blended settings[49].
	UI customization		MaA3	MuA7	Customization of voice, color, layout, and other user interface elements[7,57].
Pedagogy	Teaching			MuA6	Determine whether the mobile app presents, explains, or models subject content, rather than solely testing it[46].
	Feedback			MuA8	The mobile app's ability to track progress and provide meaningful, specific, and timely feedback that reinforces correct responses, enabling learners to conduct self-evaluation[46,47,50,54–56].
	Engagement		MaA2	MuB7	Evaluates whether the mobile app maintains student engagement and attention, or if it is perceived negatively by the learner[46,47,54].
	Learning values			MuA9	Typically evaluated through teacher or student perceptions, focusing on learner control, knowledge construction, information-seeking skills, and overall achievement. Effective mobile apps structure learning into manageable, connected steps and offer strategies to support task completion[58].
	Cognitive development				Emphasize recognition over recall, using appropriate language and content grounded in familiar concepts[47]
	Practice targets				A mobile app that enables users to practice learning targets either in a realistic, problem-based environment or in isolation, such as through flashcards[50,54]
	Motivation				The mobile app delivers pedagogical feedback from instructors through activities or content designed to engage learners' interests[45,49,51,55].
	Goal orientation adaption				Evaluate mobile app flexibility based on applicability to diverse curricular areas[54]
	Effective scaffolding				Evaluate whether the app's activities progress in difficulty in a manner that supports the learner[46]
	Self-Directedness				Educational mobile apps are recommended to provide learners with personalization options[52]
	Frequency		MaE2	MuB6	The frequency of mobile app use by teachers and learners varies by purpose; mobile apps with potential for daily use are considered more valuable than those used occasionally[48]
Continued					

TPACK Dimensions	Metric	MARS	MARuL	Definition
Content	Customization		MuD3	Mobile apps should allow personalization and individual profiles to match learners' abilities and needs[45–47,54,55].
	Authenticity			Include classroom activities that provide real-life practice through simulations, such as exploring spacecraft or planets, or constructing atoms using electrons, protons, and neutrons[49,51]
	Interactive elements	MaA2, MaA4	MuA7	Engages users and supports learning by allowing content manipulation for creativity and expression, with appropriate gestures and easy interaction[47,49,53].
	Content quality	MaD3, MaD5		Content should be balanced, up-to-date, accurate, objective, respectful of copyright, and presented with clear instructions[46,49].
	Critical thinking			The mobile app should promote higher-order thinking by encouraging questioning, analysis, evaluation, and openness. It should help students connect prior experiences with in-app activities and provide opportunities for scientific inquiry through observation, experience, reflection, reasoning, and communication[45,49,50,54,56,59]
	Relevance	MaD2	MuA5	The mobile app's content should be aligned with and reinforce the targeted goals and skills of the lesson[45,48,49,56].
	Accuracy	MaD1, MaD4, MaD6	MuB3, MuC3	Content should be accurate and complete, with graphics that enhance understanding of scientific concepts and correctly represent experimental procedures and measurements[56].
	Curriculum oriented			Mobile apps should align with curriculum skills or concepts and effectively incorporate them into meaningful learning experiences[45,50,52]
	Up-to-date		MuC1	Delivering information and representations in a timely manner[51]
	Credibility of app	MaD6	MuC2	Evaluate whether the mobile app was developed by experienced experts in the relevant learning domain[7,57]
	Evidence base	MaD7	MuC1	Check for evaluation and validation in academic sources[57]
General	Satisfaction	MaE4	MuB2	The evaluator's level of satisfaction with the mobile app[7,57]
	Perceived importance	MaE3	MuB4	Mobile app's impact on student learning[7,57]
	Advertisements		MuD5	Evaluate whether ads affect usability and if a paid version eliminates them[7]
	Recommend to other	MaE1,	MuB1	Evaluator's willingness to recommend the mobile app[7,57]

Table 1. Categorization of evaluation criteria within TPACK dimensions and mapping to MARS and MARuL sub-criteria.

MARul items	MARS related item	Overlap with MARS items
Teaching and learning	MaD2, MaD3, MaD4, MaD7, MaA3, MaA4	6/9
User-centered	MaA2, MaA4, MaA5, MaB4, MaD4, MaE1, MaE2, MaE3, MaE4,	7/7
Professional	MaD4, MaD6, MaD7	3/3
Usability	MaC1, MaC2, MaC3, MaB1, MaB2	4/7

Table 2. MARS and MARuL criteria relation.

Evaluator	Gender	OS version	Hardware
Evaluator 1	Female	Android 10	Xiaomi Mi 10
Evaluator 2	Male	Android 10	Samsung Galaxy A51
Evaluator 3	Male	Android 11	Redmi Note
Evaluator 4	Female	Android 10	Samsung Galaxy A30s

Table 3. Evaluator description.

involved in rating the medical education mobile apps using the MARS and MARuL tools were all trained and qualified individuals with expertise in medical knowledge and proficiency in using the evaluation tools. The evaluators' average ratings were used to evaluate the reliability of the mobile app evaluation. The statistical analysis was performed using SPSS version 26 software.

Results
Selection of mobile apps
A total of 409 mobile apps were initially identified on Google Play. After removing duplicates, 298 mobile apps remained. Two mobile apps were excluded because they were not available in English. An additional 58 mobile apps were considered unsuitable for medical education. Three apps were specific to training nurses and two apps were categorized as children's games. 13 mobile apps were no longer available. Another 64 mobile apps were excluded because they were not free during the second phase of monitoring. In the third stage, the remaining downloaded mobile apps were evaluated for their functionality on mobile phones. As a result, 126 mobile apps were found to be either unable to run or encountered errors during execution and were therefore removed.

Ultimately, 30 mobile apps were identified as eligible for inclusion in the study. The process of mobile app identification, following the PRISMA framework, is illustrated in Fig. 3.

Mobile app features

Considering the educational services provided by the mobile apps, they were further classified into four groups[1]: content for clinical education[2], content for basic science instruction[3], questions for evaluation[4], content for skill training. Based on this classification, 12 mobile apps offered content specifically for medical clinical training. Among these, two mobile apps also provided instruction in disease diagnosis, patient treatment, and prescription writing skills. In addition, one mobile app supplied assessment questions or supported the creation of exam question banks alongside educational content. For basic medical courses that support clinical training, 12 mobile apps provided relevant instructional material. One of these mobile apps focused exclusively on laboratory skill training. Furthermore, six mobile apps provided only question-based resources for examinations in the basic and clinical sciences.

The analysis of medical education mobile apps showed that most mobile apps (17 out of 30) primarily provided knowledge-based content. only two mobile apps focused on delivering skill-based instructional material. Additionally, eight mobile apps offered multiple-choice examination services for students and graduates (see Fig. 3).

Thematically, the mobile apps covered a limited but frequently used set of topics in medical education (15 out of 51). The topics represented included pharmacology (3 out of 30), physiology (1 out of 30), transcription skills (1 out of 30), embryology (3 out of 30), genetics (4 out of 30), histology (2 out of 30), nutrition science (5 out of 30), parasitology (2 out of 30), epidemiology (1 out of 30), bacteriology (1 out of 30), pathology (7 out of 30), anatomy (2 out of 30), immunology (2 out of 30), infectious disease (3 out of 30), microbiology (1 out of 30), and health (1 out of 30) (see Fig. 4).

However, no mobile apps were identified that provided instruction in a wide range of additional medical subjects. These included anatomy, dentistry, dermatology, forensic medicine, obstetrics and gynecology, neurology, oncology, ophthalmology, psychiatry, radiology, sports medicine, health, immunology, infectious diseases, informatics and telemedicine, internal medicine, otolaryngology, pediatrics, social medicine, surgery, and medical ethics. Although these fields fall within the scope of medical education, no corresponding mobile apps were available.

Ranking of mobile apps

Mobile apps were evaluated using the MARS and MARuL tools, and the resulting scores are shown in Figs. 5 and 6. Detailed scores for each sub criterion of both instruments are provided in Tables 1 and 2 in the appendix. The overall MARS scores ranged from 1.03 to 4.41 with a standard deviation of 0.85. The MARuL scores ranged from 20.75 to 82.75, with a standard deviation of 14.01.

In the MARS evaluation, performance received the highest mean score (3.79), while Subjective quality received the lowest mean score (2.02). The mean overall MARS score was 3.05 out of 5. A total of 56% of apps (17 out of 30) achieved an acceptable MARS score (> 0.3).

In the MARuL evaluation, the Professional sub-criterion achieved the highest mean score (7.82 out of 12; 65%), whereas the User-centered sub-criterion had the lowest mean score (10.27 out of 28; 37%). The mean
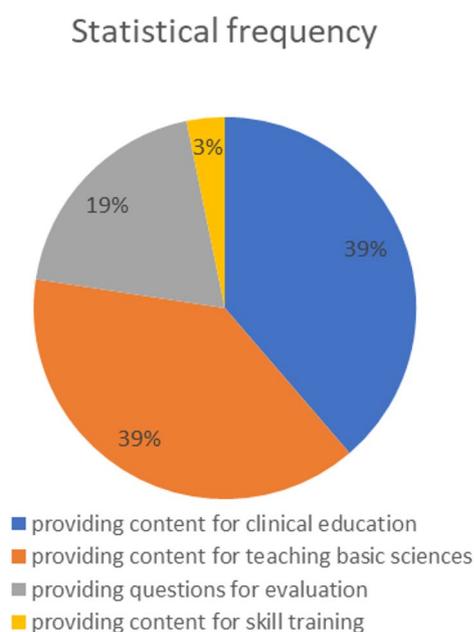
Statistical frequency

- providing content for clinical education
- providing content for teaching basic sciences
- providing questions for evaluation
- providing content for skill training

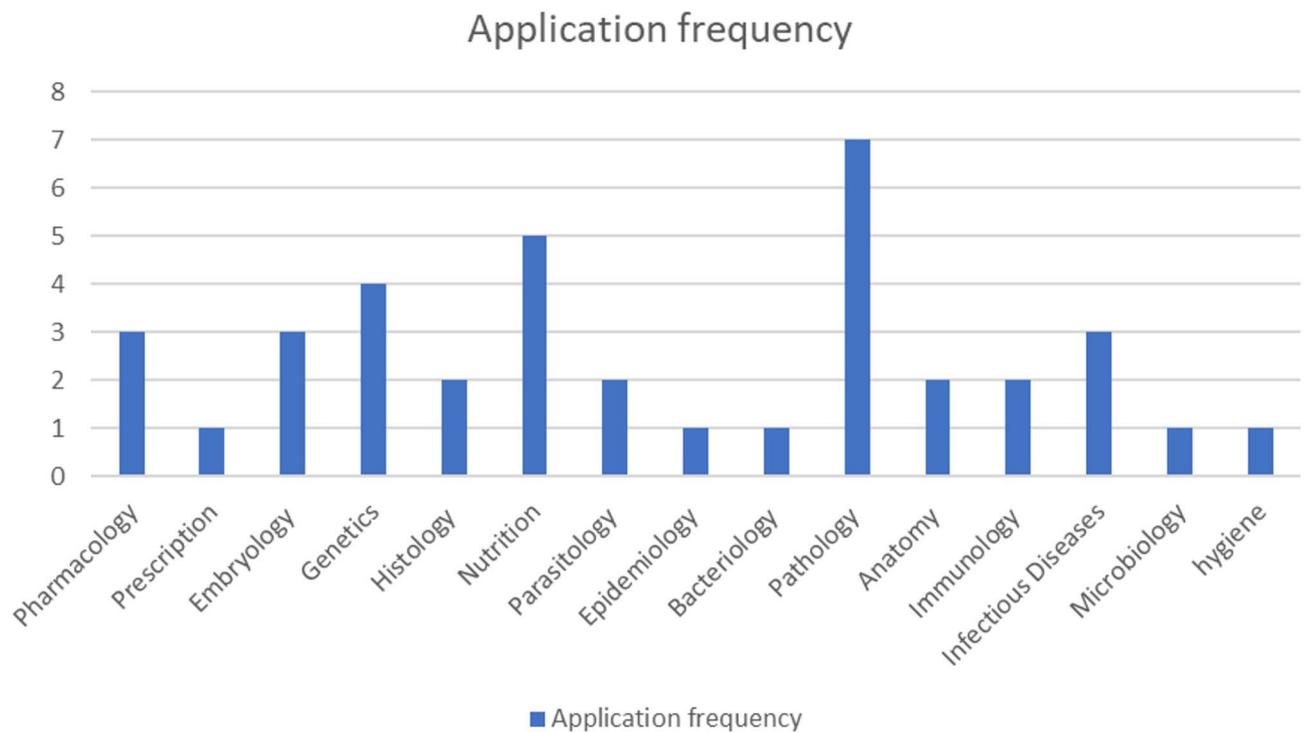**Fig. 3**. Statistical frequency of selected mobile app in leaning subject.

Fig. 4. Statistical frequency of selected mobile app in medical science subject.
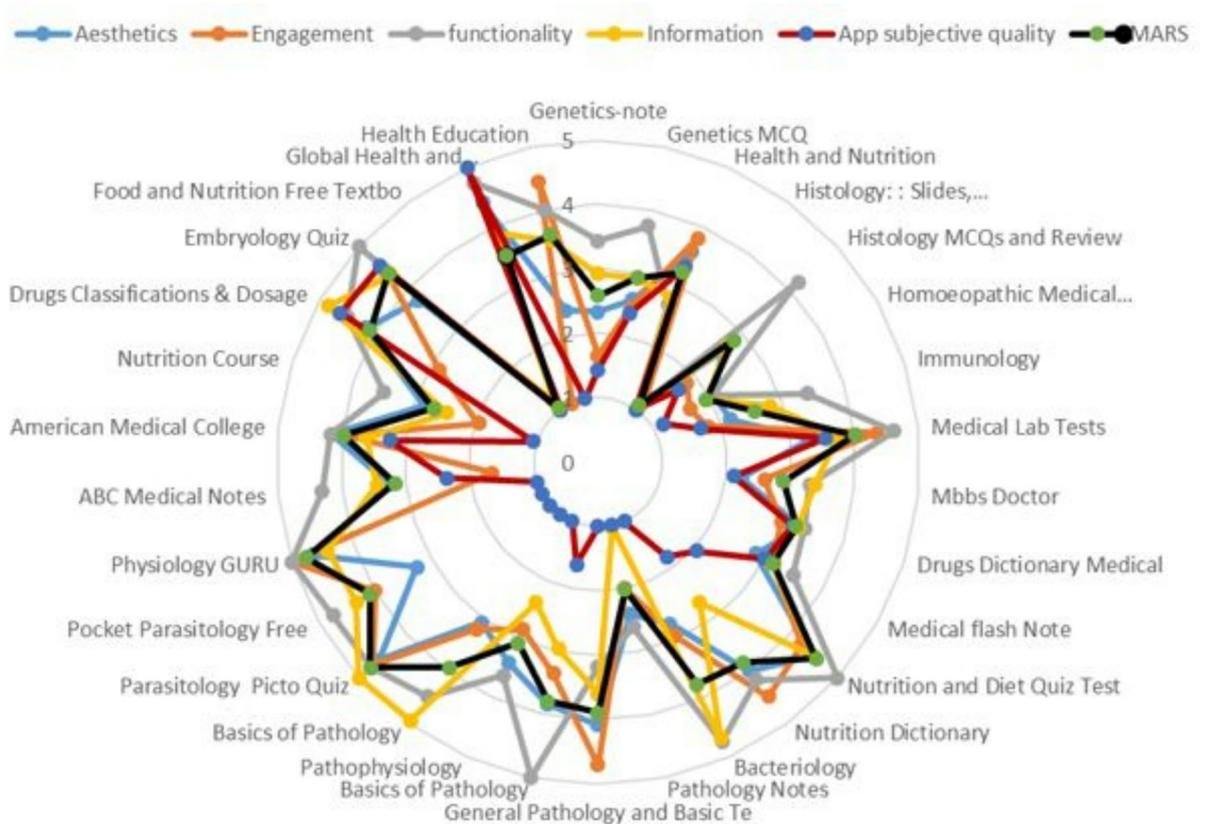

Fig. 5. Total MARS and sub criteria score for medical education mobile app.
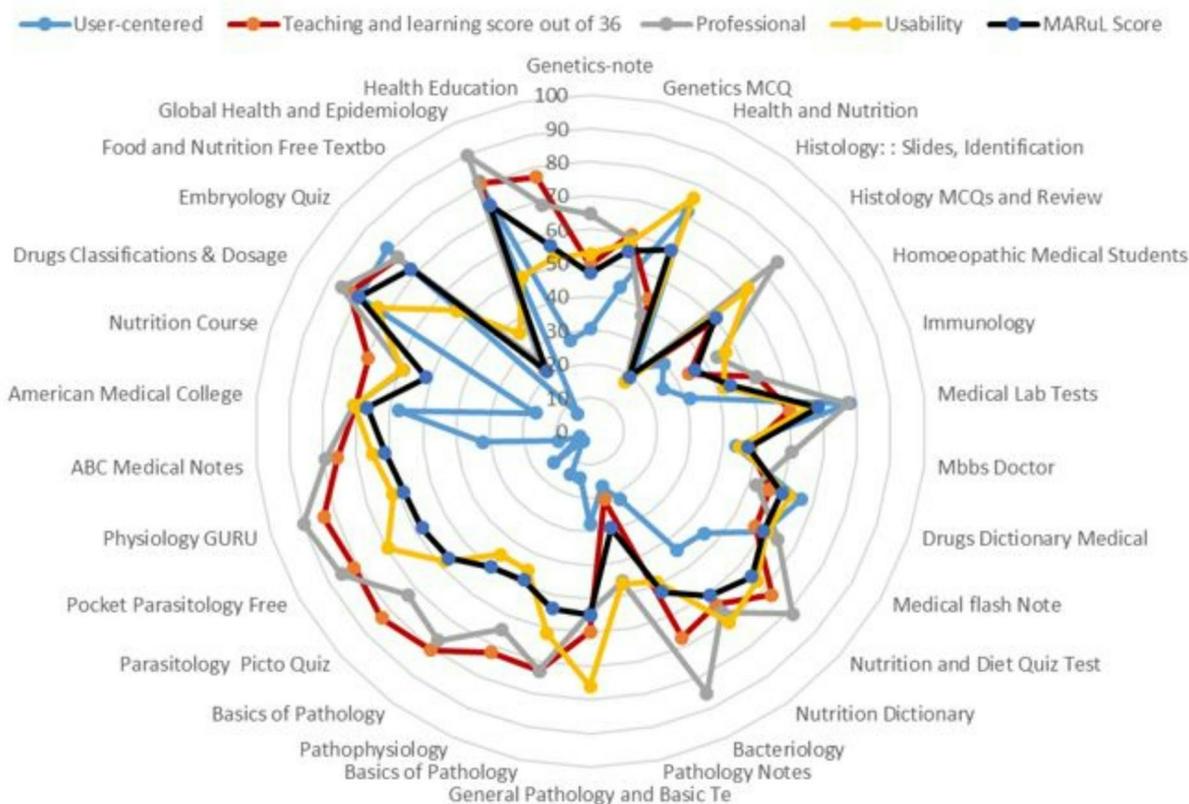
Fig. 6. Total MARuL and sub criteria score for medical education mobile app.

MARS sub-scale	Internal consistency	Cronbach's Alpha
Engagement	0.96 (95% CI: 0.93–0.98)	0.97
Aesthetics	0.97 (95% CI: 0.95–0.98)	0.95
App subjective quality	0.97 (95% CI 0.95–0.98)	0.97
Functionality	0.97 (95% CI: 0.95–0.98)	0.97
Information	0.97 (95% CI: 0.95–0.98)	0.97
Total MARS score	0.99 (95% CI: 0.98–099.98)	0.99

Table 4. Result of intraclass correlation coefficient (ICC) analysis.

overall MARuL score was 56.04 out of 104. A total of 73% of mobile apps (22 out of 30) obtained an acceptable MARuL score (> 51).

The inter-rater agreement for the overall MARS score was excellent and strong at the mobile app level, with an ICC(3, k) of 0.97 (95% CI: 0.95–0.98). The evaluation results also showed excellent inter-rater agreement for the Aesthetic sub-criterion (ICC(3,k) = 0.97 (95% CI: 0.95–0.98), Engagement (ICC(3,k) = 0.96 (95% CI: 0.93–0.98)), Performance (ICC(3,k) = 0.97 (95% CI: 0.95–0.98)), Information (ICC(3,k) = 0.97 (95% CI: 0.95–0.98)), and Subjective quality of the mobile app (ICC(3,k) = 0.97 (95% CI: 0.95–0.98)). The results obtained for the sub-criterion of Subjective quality and the overall MARS score in this study differ from other research findings[39,58,59]. The agreement analysis results for the evaluators of the mixed-effects model are displayed in the Table 4.

The inter-rater agreement for the overall MARuL score was excellent, with an ICC(3,k) of 0.94 (95% CI: 0.89–0.97) at the mobile app level. The evaluation results also showed excellent inter-rater agreement for the User-centered sub-criterion (ICC(3,k) = 0.93 (95% CI: 0.88–0.96)), Teaching and learning (ICC(3,k) = 0.83(95% CI: 0.70–0.91)), Usability (ICC(3,k) = 0.85 (95% CI: 0.75–0.92)) and Professional (ICC(3,k) = 0.93 (95% CI: 0.87–0.96)). The agreement analysis results for the evaluators of the mixed-effects model are displayed in the Table 5.

Based on the MARS scores, the top three mobile apps were Embryology Quiz with a score of 4.41 (MARuL score: 74.75), Drugs Classifications & Dosage with a score of 4.19 (MARuL score: 82.75), and Nutrition Quiz with a score of 4.06 (MARuL score: 67.50). Based on the MARuL scores, the top three mobile apps were Drugs Classifications & Dosage with a score of 82.75, Global Health and Epidemiology with a score of 76.50 (MARS score: 3.81) and Embryology Quiz with a score of 75.75.

MARuL sub-scale	Internal consistency	Cronbach's Alpha
User-centered	0.93 (95% CI: 0.88–0.96)	0.93
Teaching and learning	0.83 (95% CI: 0.70–0.91)	0.83
Usability	0.85 (95% CI: 0.75–0.92)	0.85
Professional	0.93 (95% CI: 0.87–0.96)	0.93
Total MARuL score	0.94 (95% CI: 0.89–0.97)	0.93

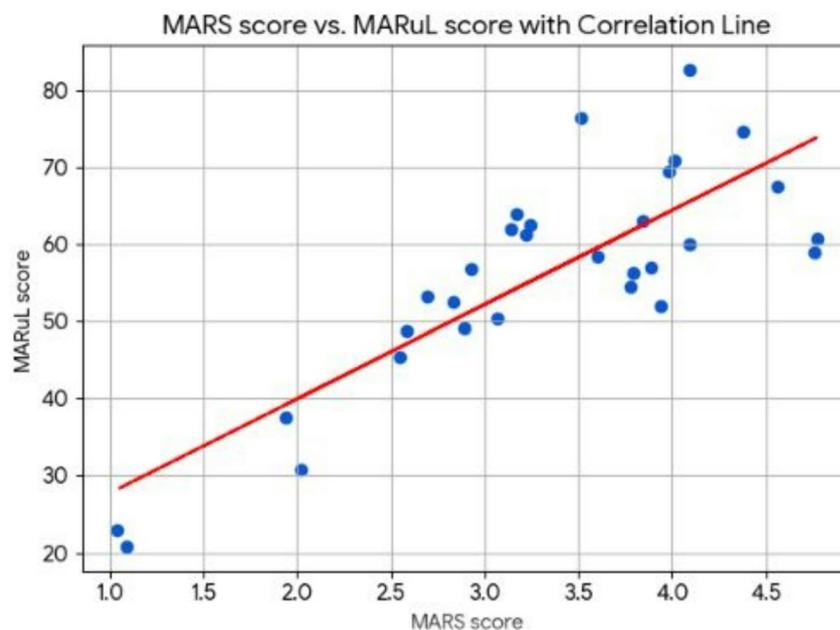**Table 5**. Result of intraclass correlation coefficient (ICC) analysis for MARuL.

Fig. 7. Correlation between MARS and MARuL scores.

The correlation between overall MARS scores and the overall MARuL score was 0.82, indicating a strong positive relationship between the two evaluation tools (Fig. 7). This result suggests a significant degree of alignment in how both tools evaluate mobile apps. Among the sub-criteria, the strongest correlation was observed between the User-centered sub-criterion of the MARuL tool and the subjective quality sub-criterion of the MARS tool ($r = 0.92$). Other notable correlations include: Teaching and learning (MARuL) and Information (MARS) ($r = 0.85$), Professional (MARuL) and Information (MARS) ($r = 0.82$), Usability (MARuL) and Aesthetics (MARS) ($r = 0.71$). The correlation between the overall MARuL score and Subjective quality was 0.70 (Fig. 8), whereas the correlation between the overall MARS score and Subjective quality was 0.27 (Fig. 9).

These findings indicate that the MARuL tool shows stronger alignment with evaluators' qualitative judgments regarding the usefulness of mobile apps for medical education. This pattern may reflect the greater emphasis of the MARuL tool on pedagogical quality criteria with the TPACK model (Table 1). In particular, the User-centered sub-criteria appears to contribute most to this alignment.

Comparing the correlation between MARuL and MARS scores with Subjective quality shows important differences. Subjective quality reflects evaluators' qualitative opinions on the usefulness of mobile apps for medical education. The MARuL tool shows a stronger alignment with these subjective evaluations. This alignment is mainly due to the User-centered sub-criterion of the MARuL tool.

In contrast, the Teaching and learning sub-criterion of the MARuL tool show weak alignment with Subjective quality sub-criterion. This is demonstrated by sub-criteria a low correlation value ($r = 0.24$). This indicates that this sub-criterion did not effectively reflect evaluators' perceptions of teaching and learning quality in the mobile apps.

The comparison of rankings also shows differences between tools. The three mobile apps with the highest Subjective quality scores in MARS tool were also among the top three mobile apps according to the MARuL tool. However, only two of the top three mobile apps based on the MARS tool were included in the top three when ranked by Subjective quality. This shows a partial overlap, but not full agreement, between the two evaluation approaches.
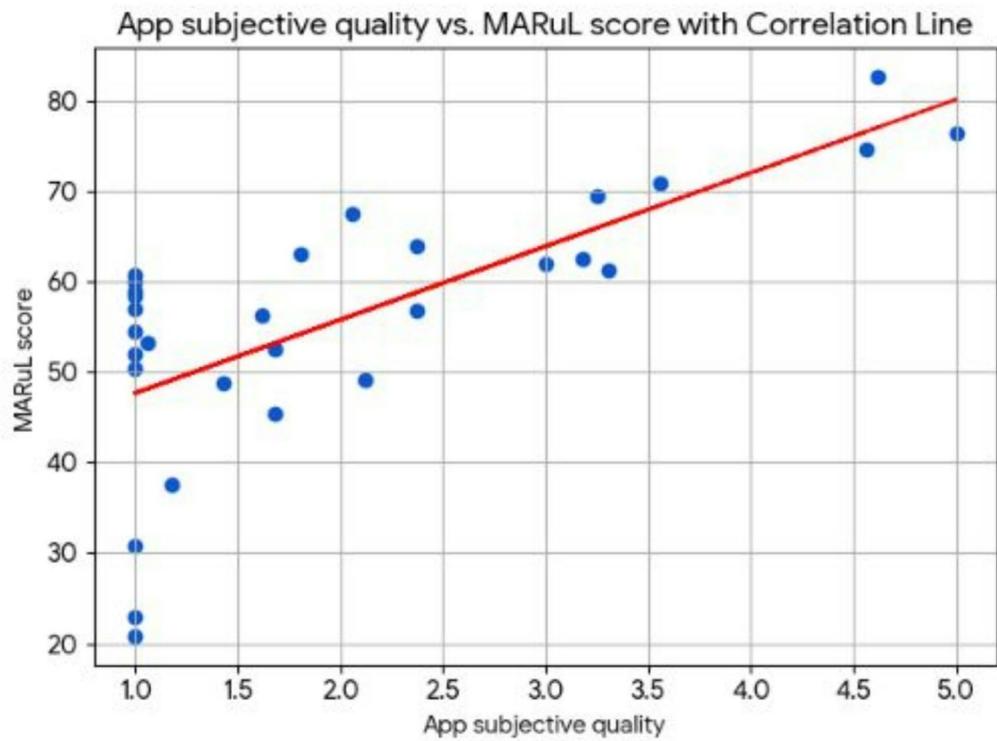
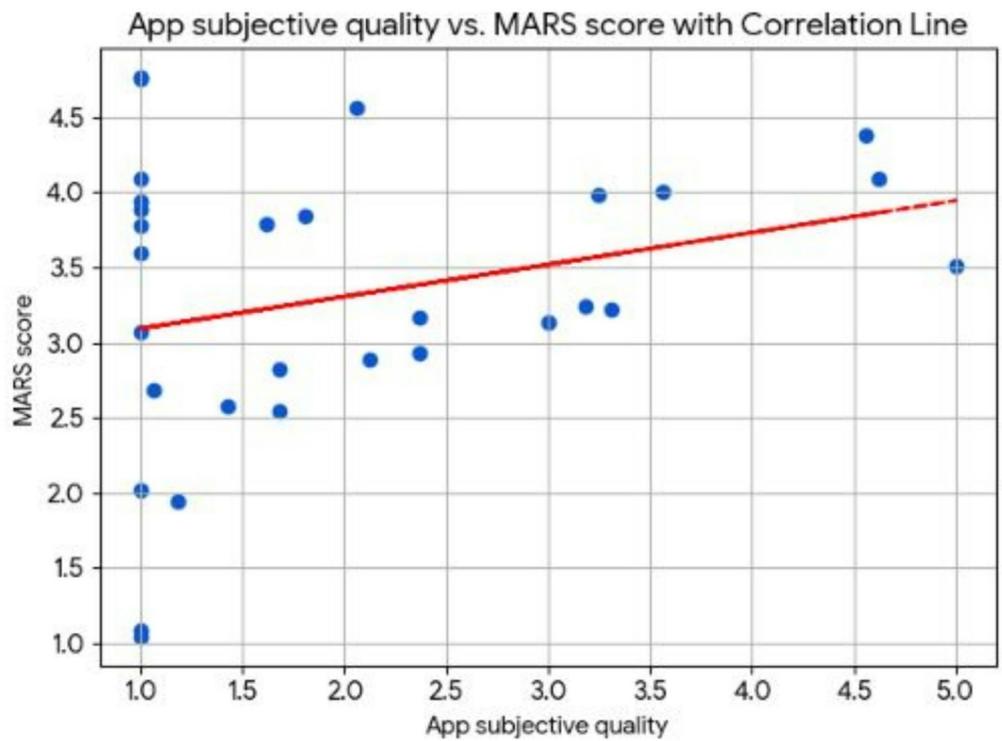**Fig. 8**. Correlation between subject quality and MARuL scores.

Fig. 9. Correlation between subject quality and MARS scores.

Discussion

Medical education mobile app evaluation via MARS and marul: implications for developers

Existing medical education mobile apps showed the lowest performance in the Engagement sub-criterion in MARS and the User-centered sub-criterion in MARuL. Half of the mobile apps (15 out of 30) did not meet acceptable Engagement scores and 73% (22 out of 30) fell below acceptable User-centered scores, with mean values of 3.01/5 (SD = 1.23) and 10.27/28 (SD = 6.80), respectively. Key Weaknesses were noted in Entertainment and Interest sub-criteria (MARS) and in Engagement and Subjective quality sub-criteria (MARuL). Overall, the results indicate inadequate user-engagement mechanisms, highlighting the need to strengthen interactive and engagement-focused features.

Visual design also emerged as a concern. A substantial proportion of mobile apps failed to meet aesthetics standards (MARS: 13 out of 30; MARuL: 10 out of 30) with mean scores of 3.16/5 (SD = 0.96) and 15.87/28 (SD = 3.58), respectively. Weaknesses were noted in page layout, graphics, and overall visual coherence, including clarity, color schemes, and format compatibility. These finding highlight the need for more improved user interface design to enhance both the aesthetic appeal and usability of medical education mobile apps.

The MARS Information sub-criteria also revealed significant limitations. 13 out of 30 mobile apps scored below acceptable levels (Mean = 3.29, SD = 1.10), particularly in providing believable, evidence-based content. Most mobile apps relied predominantly on text, with minimal use of visual or multimedia elements, thereby reducing engagement and learning impact. While MARS evaluates content accuracy, quantity, credibility, and alignment with mobile app goals, it does not capture key pedagogical factors such as scaffolding, prerequisite alignment, difficulty level, or appropriate learning contexts. Consequently, mobile apps containing accurate and credible information may still fail to support deeper understanding or meaningful behavior change.

Performance was generally adequate, with 25 out of 30 mobile apps achieving acceptable scores (Mean = 3.8, SD = 1.08). This criterion evaluates ease of use, efficiency, navigation, and motion design. While mobile apps performed well in ease of use and efficiency, navigation and motion design remained suboptimal.

For the MARuL Teaching and learning sub-criterion, mobile apps obtained an average score of 22.09 (SD = 6.71), with only seven mobile apps scoring below acceptable levels. This sub-criterion relies on qualitative judgment of learning suitability. However, no correlation was observed between these scores and evaluators' subjective evaluation under the MARS Subjective quality sub-criterion, indicating that MARuL may insufficiently capture key pedagogical dimensions.

MARS and MARuL performance in medical education mobile app evaluation: strengths and weaknesses within TPACK

The central hypothesis—that existing evaluation tools do not adequately capture the qualitative features of medical education mobile apps—was supported. Previous evaluations using MARS, conducted by experts in medical and educational technology, showed that mobile apps with acceptable MARS scores were often rated poorly in subjective expert evaluations. Conversely, mobile apps receiving acceptable subjective quality ratings consistently achieved satisfactory MARS scores[8], indicating that MARS alone is insufficient for comprehensive qualitative evaluation. Extending this analysis using correlation tests between MARuL scores and subjective quality ratings reduced but did not eliminate these discrepancies. So, MARuL, similar to MARS, remains inadequate for a complete and standardized evaluation of medical education mobile apps.

Evaluators' responses to general criteria—willingness to recommend the mobile app, willingness to pay, and overall satisfaction—showed a very high correlation between MARS Subjective quality and MARuL User-centered sub-criteria ($r = 0.92$) indicating a shared understanding overall mobile app quality. However, weak correlations between the Teaching and learning sub-criterion and Subjective quality sub-criterion ($r = 0.24$) and the User-centered sub-criterion ($r = 0.11$) confirm that both tools insufficiently capture pedagogical effectiveness.

Although MARS and MARuL provide valuable insights into mobile app functionality and basic usability, both tools remain inadequate for evaluating the pedagogical effectiveness of medical education mobile apps. In the comparative analysis of the MARS and MARuL tools against the TPACK based educational mobile app evaluation framework[43,50], the shortcomings of both tools were identified (Table 1). MARuL covers a greater proportion of pedagogy-related criteria within the TPACK model (MARuL = 45%, MARS = 18%; Table 1), it still does not provide comprehensive coverage. Both tools require refinement to address essential pedagogical criteria. Incorporating additional dimensions—such as cognitive development (e.g., providing case-based learning modules that require students to analyze clinical symptoms and formulate differential diagnoses), practice-based learning (e.g., timed quizzes and competency-oriented checklists), motivation (e.g., progress dashboards, achievement badges, and adaptive feedback), goal orientation (e.g., structured pathways progressing from foundational concepts to advanced reasoning), effective scaffolding (e.g., gradual progression from guided to complex, minimally supported clinical cases), and self-directed learning (e.g. enabling users to select learning topics, establish personal study goals, and review materials at their own pace.), curriculum alignment (e.g., mapping to medical education standards), critical thinking (e.g., interactive simulations requiring clinical decision-making), and authenticity (e.g., virtual patients, real-world case repositories, and evidence-based guidelines)- would enable these tools to provide a more comprehensive, evidence-based pedagogical evaluation aligned with the TPACK model. These enhancements directly addressing the limitations identified in the present study.

Limitation

The primary limitation of this study was the difficulty in recruiting experts with combined expertise in medical knowledge, medical education, and mobile app evaluation. The extensive training required to familiarize evaluators with the evaluation tools further added to this challenge. The evaluation process was time-intensive, and limited experts' availability prolonged the overall timeline. Another constraint was the lack of access to iOS

devices, due to high cost and sanctions restricting certain features in Iran, which prevented inclusion of iOS mobile apps. An additional limitation was the variability in mobile app performance across different hardware models and operating system versions, which extended both the mobile app selection and evaluation processes.

Conclusion

In this research, the quality of medical education mobile apps was evaluated using the MARS and MARuL tools. Four evaluators, all holding medical doctorates, conducted the evaluation. Approximately 73% of the reviewed mobile apps achieved acceptable scores on the MARuL tool, whereas 56% met the acceptable threshold on the MARS tool. (Figures 5 and 6). Statistical analysis demonstrated excellent inter-rater agreement (Tables 4 and 5).

Among the sub-criteria, Functionality and Professional received the highest scores, followed by Information and Aesthetics in MARS, and Teaching and learning, and Usability in MARuL. However, many medical education mobile apps did not achieve acceptable scores in the MARS Engagement and Subjective quality sub-criterion, nor in the MARuL User-centered sub-criterion (Figs. 5 and 6). In MARS, the primary weaknesses were low scores for Entertainment and Interest, while in MARuL, the weakest dimensions were Engagement and Subjective quality. These findings show that most medical education mobile apps lack effective mechanisms to engage users or provide an enjoyable learning experience.

While MARS and MARuL provide useful insights, they inadequately address pedagogical dimensions outlined in the TPACK model (Table 1). MARS overlooks educational objectives, and MARuL omits key components such as cognitive development, practice targets, motivation, goal orientation, effective scaffolding, self-directed learning, curriculum oriented, critical thinking and authenticity (Table 1). Addressing these gaps is essential to enhance both tools for comprehensive, evidence-based evaluation of medical education mobile apps. As a next step, we recommend developing new evaluation frameworks derived from MARS and MARuL but extended to incorporate these pedagogical criteria, thereby enabling a more thorough and education-centered evaluation of medical education mobile apps.

Data availability

All data generated or analyzed during this study are included in this published article.

References

1. Lin, B. & Wu, S. Digital transformation in personalized medicine with artificial intelligence and the internet of medical things. *Omics J. Integr. Biol.* **26** (2), 77–81 (2022).
2. Singh, S., Bhatt, P., Sharma, S. K. & Rabiu, S. Digital transformation in healthcare: innovation and technologies. In: Blockchain for Healthcare Systems. *CRC.* 61–79. (2021).
3. Moran, J., Briscoe, G. & Peglow, S. Current technology in advancing medical education: perspectives for learning and providing care. *Acad. Psychiatry.* **42** (6), 796–799 (2018).
4. Chandran, V. P. et al. Mobile applications in medical education: A systematic review and meta-analysis. *PLoS One.* **17** (3), e0265927 (2022).
5. Latif, M. Z., Hussain, I., Saeed, R., Qureshi, M. A. & Maqsood, U. Use of smart phones and social media in medical education: trends, advantages, challenges and barriers. *Acta Inf. Med.* **27** (2), 133 (2019).
6. Aldekhyyel, R. N. & Almulhem, J. A. Workshop on evaluating medical apps using the mobile app rubric for learning. *Health Inf. J.* **31** (2), 14604582251338663 (2025).
7. Gladman, T. et al. A tool for rating the value of health education mobile apps to enhance student learning (MARuL): development and usability study. *JMIR MHealth UHealth.* **8** (7), e18015 (2020).
8. Owrangi, S., Zarifsanaiey, N. & Rezaei, M. Mobile apps to enhance student learning in medical education: A systematic search in app stores and evaluation using the mobile app rating scale. *Comput. Biol. Med.* **196**, 110740 (2025).
9. Wallace, L. S. & Dhingra, L. K. A systematic review of smartphone applications for chronic pain available for download in the united States. *J. Opioid Manag.* **10** (1), 63–68 (2014).
10. de la Vega, R. & Miró, J. mHealth: A strategic field without A solid scientific soul. A systematic review of pain-related apps. *PloS One.* **9** (7), e101312 (2014).
11. Salazar, A., de Sola, H., Failde, I. & Moral-Munoz, J. A. Measuring the quality of mobile apps for the management of pain: systematic search and evaluation using the mobile app rating scale. *JMIR MHealth UHealth.* **6** (10), e10718 (2018).
12. Trépanier, L. C. et al. Smartphone apps for menstrual pain and symptom management: A scoping review. *Internet Interv.* **31**, 100605 (2023).
13. Lalloo, C. et al. Commercially available smartphone apps to support postoperative pain self-management: scoping review. *JMIR MHealth UHealth.* **5** (10), e8230 (2017).
14. Shen, N. et al. Finding a depression app: a review and content analysis of the depression app marketplace. *JMIR MHealth UHealth.* **3** (1), e3713 (2015).
15. Powell, A. C. et al. Interrater reliability of mHealth app rating measures: analysis of top depression and smoking cessation apps. *JMIR MHealth UHealth.* **4** (1), e5176 (2016).
16. Bardus, M., van Beurden, S. B., Smith, J. R. & Abraham, C. A review and content analysis of engagement, functionality, aesthetics, information quality, and change techniques in the most popular commercial apps for weight management. *Int. J. Behav. Nutr. Phys. Act.* **13** (1), 1–9 (2016).
17. Rivera, J. et al. Mobile apps for weight management: a scoping review. *JMIR MHealth UHealth.* **4** (3), e5115 (2016).
18. Gladman, T., Tylee, G., Gallagher, S., Mair, J. & Grainger, R. Measuring the quality of clinical skills mobile apps for student learning: systematic search, analysis, and comparison of two measurement scales. *JMIR MHealth UHealth.* **9** (4), e25377 (2021).
19. Park, J. Y. E., Tracy, C. S. & Gray, C. S. Mobile phone apps for family caregivers: a scoping review and qualitative content analysis. *Digit. Health.* **8**, 20552076221076672 (2022).
20. Jannati, N., Salehinejad, S., Kuenzig, M. E. & Pena-Sanchez, J. N. Review and content analysis of mobile apps for inflammatory bowel disease management using the mobile application rating scale (MARS): systematic search in app stores. *Int. J. Med. Inf.* **180**, 105249 (2023).

21. Naranjo-Rojas, A., Perula-de Torres, L. Á., Cruz-Mosquera, F. E. & Molina-Recio, G. Usability of a mobile application for the clinical follow-up of patients with chronic obstructive pulmonary disease and home oxygen therapy. *Int. J. Med. Inf.* **175**, 105089 (2023).

22. do Nascimento, V. S., Rodrigues, A. T., Rotta, I., de Mendonça Lima, T. & Aguiar, P. M. Evaluation of mobile applications focused on the care of patients with anxiety disorders: A systematic review in app stores in Brazil. *Int. J. Med. Inf.* 175,105087. (2023).

23. Payo, R. M. et al. Spanish adaptation and validation of the mobile application rating scale questionnaire. *Int. J. Med. Inf.* **129**, 95–99 (2019).

24. Stevens, D. J., Jackson, J. A., Howes, N. & Morgan, J. Obesity surgery smartphone apps: a review. *Obes. Surg.* **24**, 32–36 (2014).

25. Nicholas, J., Larsen, M. E., Proudfoot, J. & Christensen, H. Mobile apps for bipolar disorder: a systematic review of features and content quality. *J. Med. Internet Res.* **17** (8), e198 (2015).

26. Bender, J. L., Yue, R. Y. K., To, M. J., Deacken, L. & Jadad, A. R. A lot of action, but not in the right direction: systematic review and content analysis of smartphone applications for the prevention, detection, and management of cancer. *J. Med. Internet Res.* **15** (12), e2661 (2013).

27. Martínez-Pérez, B., De La Torre-Díez, I., López-Coronado, M. & Herreros-González, J. Mobile apps in cardiology. *JMIR MHealth UHealth.* **1** (2), e2737 (2013).

28. Arnhold, M., Quade, M. & Kirch, W. Mobile applications for diabetics: a systematic review and expert-based usability evaluation considering the special requirements of diabetes patients age 50 years or older. *J. Med. Internet Res.* **16** (4), e104 (2014).

29. Schnall, R. & Iribarren, S. J. Review and analysis of existing mobile phone applications for health care–associated infection prevention. *Am. J. Infect. Control.* **43** (6), 572–576 (2015).

30. Muessig, K. E., Pike, E. C., LeGrand, S. & Hightow-Weidman, L. B. Mobile phone applications for the care and prevention of HIV and other sexually transmitted diseases: a review. *J. Med. Internet Res.* **15** (1), e2301 (2013).

31. Schnall, R. et al. Comparison of a user-centered design, self-management app to existing mHealth apps for persons living with HIV. *JMIR MHealth UHealth.* **3** (3), e4882 (2015).

32. Roy, A., Nayak, P. P., Shenoy, P. R. & Somayaji, K. Availability and content analysis of smartphone applications on Conservative dentistry and endodontics using mobile application rating scale (MARS). *Curr. Oral Health Rep.* **9** (4), 215–225 (2022).

33. Bonetti, R., Rod, B., Sabourin, C. & Hauw, D. Quality of mobile apps for psychological skills training in sport: a MARS-based study. *Comput. Hum. Behav. Rep.* 18, 100635. (2025).

34. Firth, J. & Torous, J. Smartphone apps for schizophrenia: a systematic review. *JMIR MHealth UHealth.* **3** (4), e4930 (2015).

35. Koehler, M. MishraWhat is technological pedagogical content knowledge (TPACK)? *Contemp. Issues Technol. Teach. Educ.* **9** (1), 60–70 (2009).

36. Alamoodi, A. H. et al. A systematic review into the assessment of medical apps: Motivations, challenges, recommendations and methodological aspect. *Health Technol.* 10, 1045-1061. (2020).

37. Kim, P., Eng, T. R., Deering, M. J. & Maxfield, A. Published criteria for evaluating health related web sites. *Bmj* **318** (7184), 647–649 (1999).

38. Nasser, S., Mullan, J. & Bajorek, B. Assessing the quality, suitability and readability of internet-based health information about warfarin for patients. *Australas Med. J.* **5** (3), 194–203. https://doi.org/10.4066/AMJ (2012).

39. Stoyanov, S. R. et al. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. *JMIR MHealth UHealth.* **3** (1), e3422 (2015).

40. Taki, S. et al. Infant feeding websites and apps: a systematic assessment of quality and content. *Interact. J. Med. Res.* **4** (3), e18 (2015).

41. Moher, D. et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst. Rev.* **4** (1), 1–9 (2015).

42. Creber, R. M. M. et al. Review and analysis of existing mobile phone apps to support heart failure symptom monitoring and self-care management using the mobile application rating scale (MARS). *JMIR MHealth UHealth.* **4** (2), e5882 (2016).

43. Shahjad, M. K. A systematic literature review on learning apps evaluation. *J. Inf. Technol. Educ. Res.* **21**, 663–700 (2022).

44. Rosell-Aguilar, F. State of the app: A taxonomy and framework for evaluating Language learning mobile applications. *CALICO J.* **34** (2), 243–258 (2017).

45. Tahir, R. & Arif, F. Framework for evaluating the usability of mobile educational applications for children. In: 2014 3th International Conference on Informatics Engineering and Information Science (ICIEIS2014). Lodz University of Technology, Lodz, Poland. 156–170. (2018).

46. Cherner, T., Lee, C. Y., Fegely, A. & Santaniello, L. A detailed rubric for assessing the quality of teacher resource apps. *J. Inf. Technol. Educ. Innov. Pract.* 15, 117-143, (2016).

47. Fernández-Pampillón Cesteros, A. M., Domínguez Romero, E. & de Armas Ranero, I. *Herramienta Para La revisión De La Calidad De Objetos De Aprendizaje Universitarios (COdA): guía Del usuario. v. 1.1* (Universidad Complutense, 2011).

48. Lubniewski, K. L., McArthur, C. L. & Harriott, W. Evaluating instructional apps using the app checklist for educators (ACE). *Int. Electron. J. Elem. Educ.* **10** (3), 323–329 (2018).

49. Baran, E., Uygun, E. & Altan, T. Examining preservice teachers' criteria for evaluating educational mobile apps. *J. Educ. Comput. Res.* **54** (8), 1117–1141 (2017).

50. Lee, J. S. & Kim, S. W. Validation of a tool evaluating educational apps for smart education. *J. Educ. Comput. Res.* **52** (3), 435–450 (2015).

51. Israelson, M. H. A tool for systematic evaluation of apps for early literacy learning. *Read. Teach.* **69** (3), 339–349 (2015).

52. Bentrop, S. M. Creating an educational app rubric for teachers of students who are deaf and hard of hearing [Internet] [Masters of Science]. Available from: https://digitalcommons.wustl.edu/cgi/viewcontent.cgi?article=1681&context=pacs_capstones (Washington University, 2014).

53. Chen, X. Evaluating language-learning mobile apps for second-language learners. *J. Educ. Technol. Dev. Exch. JETDE.* **9** (2), 3 (2016).

54. Green, L. S., Hechter, R. P., Tysinger, P. D. & Chassereau, K. D. Mobile app selection for 5th through 12th grade science: the development of the MASS rubric. *Comput. Educ.* **75**, 65–71 (2014).

55. Kay, R. Creating a framework for selecting and evaluating educational apps. In: INTED2018 Proceedings. *IATED.* 374–82. (2018).

56. Chen, T., Hsu, H. M., Stamm, S. W. & Yeh, R. Creating an instrument for evaluating critical thinking apps for college students. *E-Learn Digit. Media.* **16** (6), 433–454 (2019).

57. Koo, T. K. & Li, M. Y. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J. Chiropr. Med.* **15** (2), 155–163 (2016).

58. Larco, A., Enríquez, F. & Luján-Mora, S. Review and evaluation of special education iOS Apps using MARS. In: 2018 IEEE World Engineering Education Conference (EDUNINE). *IEEE.* 1–6. (2018).

59. Knitza, J. et al. German mobile apps in rheumatology: review and analysis using the mobile application rating scale (MARS). *JMIR MHealth UHealth.* **7** (8), e14991 (2019).

Author contributions

Mohammadsadegh Rezaei: Conceptualization, Methodology development, Review and editing, Discussion, SupervisionSara Owrangi: Writing, Visualization, Formal analysis, Validation, InvestigationSahar Jafari: Experi-

ment, InvestigationAll authors reviewed the manuscript.

Declarations

Competing interests
The authors declare no competing interests.

Ethical considerations
The study protocol was reviewed and approved in 2023 by the Ethics and Research Committee of Shiraz University of Medical Sciences. Given that no human or animal subjects were directly involved in this research, the study received ethical approval. The research was conducted in accordance with the ethical principles outlined in the revised Declaration of Helsinki, ensuring the protection of participants' rights, dignity, and confidentiality.

Informed consent
was obtained from all human participants who took part in the study as experts in the evaluation of the medical education mobile app.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-31973-4.

Correspondence and requests for materials should be addressed to M.R. or S.O.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.