






Medical education research quality (MERSQ) checklist development: Are searches of BEME and non-BEME reviews standard?

A mixed method study

Maryam Alizadeh, PhD^a, Rahem Rahmati, MD^{b,c}, Fatemeh Zarimeidani, MD^{b,c}, Fatemeh Hasani, MD^d, Arshin Ghaedi, MD^e, Aida Bazrgar, MD^f, Reza Hosseini Doalame, MD^{c,g}, Hojat Vahedi, MD^h, Hamidreza Hekmat, MDⁱ, Negar Omid, MD^{i,*}

Abstract

Even though there has been a lot of research in medical education, the quality of it has not increased similarly. This study aimed to provide a valid and reliable user-friendly tool for evaluating search strategies in medical education systematic reviews. This mixed study was conducted in 2019 to 2021, including 3 phases: systematic search, developing a medical education research quality (MERSQ) checklist, and evaluation of the search quality of best evidence in medical education collaboration (BEME) and non-BEME reviews. Three hundred nineteen items were retrieved from the systematic search of PubMed, Embase, Scopus, Psycinfo, ERIC, and Google Scholar. Following ensuring acceptable criteria, 30 items were included in comprehensiveness or reproducibility guarantees. The results showed that the instrument had an the intra-class correlation coefficient of 0.922 ($P = .002$), the reproducibility guarantee had 0.903 ($P = .003$), and the comprehensiveness guarantee had 0.926 ($P = .006$). We also calculated inter-rater reliability and internal consistency using Cronbach alpha of 0.827 ($P < .001$) and an instrument the intra-class correlation coefficient of 0.978. Using MERSQ, the overall search quality (41.75 vs 31.25, $P = .009$), reproducibility (22 vs 14.50, $P = .004$), and comprehensive score (18.75 vs 15.75, $P = .880$) of BEME studies were higher than non-BEME ones. Moreover, we found only 30% of studies completed searching documents. The search strategy query concerning the selection of synonym terms received the lowest score among studies. This study led to the development of a valid and reliable checklist for evaluating the search quality of medical education systematic reviews. Utilizing the MERSQ checklist, we found that BEME studies had higher quality than non-BEME ones, making the results from BEME studies more reliable.

Abbreviations: BEME = best evidence in medical education collaboration, CVI = content validity index, ICC = the intra-class correlation coefficient, IRR = inter-rater reliability, MERSQ = medical education research quality.

Keywords: best evidence medical education, medical education, search quality, search strategy, systematic review

This work was supported by the Tehran University of Medical Sciences under Grant 41822.

The authors have no conflicts of interest to disclose.

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

This study has been approved by the Ethical Committee of Tehran University of Medical Sciences (IR.TUMS.VCR.REC.1398.111) and conducted in accordance with the declaration of Helsinki.

Supplemental Digital Content is available for this article.

^a Department of Medical Education, School of Medicine, Health Professions Education Research Center, Tehran University of Medical Sciences, Tehran, Iran,

^b Students Research Committee, Shahrekord University of Medical Sciences, Shahrekord, Iran, ^c Tehran Heart Center, Cardiovascular Diseases Research Institute, Tehran University of Medical Sciences, Tehran, Iran, ^d Golestan Research Center of Gastroenterology and Hepatology, Golestan University of Medical Sciences, Gorgan, Iran, ^e Trauma Research Center, Shahid Rajaei (Emtiaz) Trauma Hospital, Shiraz University of Medical Sciences, Shiraz, Iran, ^f Student Research Committee, School of Medicine, Shiraz University of Medical Sciences, Shiraz, Iran, ^g Students' Scientific Research Center, Tehran University of Medical Sciences, Tehran, Iran, ^h Department of Medical Education, Health Professions Education Research Center, Education Development Center, Tehran University of Medical Sciences, Tehran, Iran, ⁱ School of Medicine, Baharloo Hospital,

International Campus, Tehran University of Medical Sciences, Tehran, Iran,

^j Cardiac Primary Prevention Research Center, Cardiovascular Diseases Research Institute, Tehran University of Medical Sciences, Tehran, Iran.

* Correspondence: Negar Omid, Cardiac Primary Prevention Research Center, Cardiovascular Diseases Research Institute, Tehran University of Medical Sciences, Kargar St. Jalal Al-Ahmad Cross, Tehran 1411713138, Iran (e-mail: negar.omidi@gmail.com).

Copyright © 2025 the Author(s). Published by Wolters Kluwer Health, Inc. This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial License 4.0 (CCBY-NC), where it is permissible to download, share, remix, transform, and buildup the work provided it is properly cited. The work cannot be used commercially without permission from the journal.

How to cite this article: Alizadeh M, Rahmati R, Zarimeidani F, Hasani F, Ghaedi A, Bazrgar A, Hosseini Doalame R, Vahedi H, Hekmat H, Omid N. Medical education research quality (MERSQ) checklist development: Are searches of BEME and non-BEME reviews standard?: A mixed method study. *Medicine* 2025;104:18(e42316).

Received: 9 December 2024 / Received in final form: 10 April 2025 / Accepted: 15 April 2025

<http://dx.doi.org/10.1097/MD.00000000000042316>

1. Introduction

In the last decade, multiple individuals and organizations, such as the best evidence in medical education collaboration (BEME), have put effort into guiding approaches and conducting systematic reviews in medical education. Checklists could help researchers develop search strategies and achieve reviews with more homogenous qualities.^[1–3] Several guidelines have been published in this regard.^[4–7] Furthermore, multiple checklists and peer-reviewing search strategies, including AMSTAR,^[8] MOOSE,^[9] and STARLITE^[10] have been designed. However, these checklists are not designed for the medical education research context and tend to include complicated items and criteria that medical education researchers may find challenging to understand and assess.

The medical education research field has rapidly developed over the last decade. The enormous bulk of controversial articles causes medical educators to be overloaded, leading to more confusion.^[11] Nevertheless, the quality of research in medical education has not improved at the same pace as its quantity.^[12–14] All these factors highlight the need for high-quality systematic reviews to synthesize and summarize published research on a specific topic and stress their strengths and weaknesses.^[15,16]

Developing search strategies for systematic reviews, determining the optimal balance between sensitivity and specificity, and translating search strategies across databases are challenging.^[17] For example, a study showed that 90.5% of the analyzed systematic reviews contained more than one error, including missing mesh terms, failing to adjust the search strategy to different databases, and logical use of Boolean operators.^[18] A meta-research study also indicated that many systematic reviews fail to conduct reproducible searches.^[19]

Therefore, the present study aimed to (a) develop a user-friendly instrument for assessing search strategies in medical education systematic reviews, and (b) assess and compare search quality between BEME reviews and non-BEME reviews published in Medical Teacher Journal.

2. Methods

This study employed the mixed method^[20] approach. It was conducted in Tehran, Iran (2019–2021), including 3 phases: systematic search, developing a checklist, and evaluation of BEME and non-BEME reviews. The present study has been approved by the Ethical Committee of Tehran University of Medical Sciences (Ethics code: IR.TUMS.VCR.REC.1398.111).

The process of this study consisted of 4 main phases. Initially, based on the AMEE Guide NO.87,^[21] authors reviewed literature published between January 2000 and 2020 about the quality of studies, standards, and checklists for systematic review protocols, particularly medical education ones. The authors searched multiple online databases, including PubMed, Embase, Scopus, Psycinfo, and ERIC, using the following search strategy: (“systematic review” OR “meta-analysis” OR “review protocol” OR “search strategy”) AND (“evaluation instrument” or “checklist” and “standard” or “statement”) AND (“medical education”). Google Scholar was also used to identify relevant gray literature. Only the documentation in English was included. Following, the research team analyzed the full text of available publications and extracted the items related to search strategies from Higgins and Green (Cochrane handbook), PRISMA, RAMESES, QUOROM, STORIES, MOOSE, and PRESS checklists, resulting in preliminary subjects of medical education research quality (MERSQ) checklist draft (Table S1, Supplemental Digital Content, <https://links.lww.com/MD/O781>).

As shown in Fig. 1, in the second phase, the authors independently reviewed all extracted items ($n = 319$). They screened them for duplication ($n = 215$) and then for relevance and appropriateness ($n = 104$) regarding the evaluation of search strategies in the context of medical education. Inappropriate

items were excluded ($n = 61$), resulting in the primary MERSQ checklist of 43 items.

As part of the second phase, the study included a panel of experts in the field of medical education. The primary checklist was sent to ten experts in medical education, medicine, statistics, and epidemiology with extensive research experience. They were asked to rate the items based on their necessity, relevancy, clarity, and applicability, and all comments were recorded. Based on comments, similar items were merged ($n = 13$) into more comprehensive items and organized into 7 major categories. Following, the content validity ratio and content validity index (CVI) were calculated using a nominal group technique for each item. Items with CVI $< .62$ were excluded, and proposed changes were considered. Thus, the final instrument with 30 items was established.

After acceptable CVI, items were divided into 2 categories: 1. Reproducibility-related items, and 2. Comprehensiveness-related items. To establish the frame of reference, authors independently score the 5 random BEME systematic reviews published in the Medical Teacher Journal during 2010 to 2020. The Intra-class Correlation Coefficient (ICC) was .903, .926, and .922 for the reproducibility guarantee ($P = .003$), comprehensiveness guarantee ($P = .006$), and the instrument ($P = .002$), respectively. Then, the authors discussed challenges and solved some issues. They also included blank parts for comments to improve cooperation between raters. Further adjustments were made considering any complications encountered.

After achieving the frame of reference, 61 BEME systematic reviews published in Medical Teacher Journal between 2000 to 2020 have been found. However, 3 of them did not meet the inclusion criteria. In the third phase, all other 58 BEME systematic reviews were scored with the MERSQ checklist by 2 researchers in the field of medical education. Two-way random ICC was computed for the exact match rate.

After that, 10 non-BEME systematic reviews published in the same journal during 2018 to 2021 were also included in the study to compare with BEME systematic reviews. Hence, these ten non-BEME systematic reviews were also assessed by those 2 researchers scoring BEME studies. Internal consistency of the instrument with Cronbach alpha and inter-rater reliability (IRR) with two-way mixed ICC was computed (Table 1). Besides, Cronbach alpha for the MERSQ was calculated at .827 with an ICC of .978 ($P < .001$). Cohen Kappa for each item IRR was also calculated (Table 2).

2.1. Statistical analysis

In this study, descriptive statistics were used to summarize the dataset. For continuous variables, data are presented as means and standard deviations if normally distributed, or as medians with interquartile ranges (IQR) for non-normally distributed variables. Categorical variables are denoted as frequencies and percentages. The content validity ratio and CVI were calculated using a nominal group technique for each item of questionnaires. Also, internal consistency of the instrument with Cronbach alpha, IRR, Cohen Kappa, and two-way mixed ICC was computed. Besides, the Mann–Whitney U test was used to compare search quality in BEME and non-BEME studies. Additionally, to control for the probability of a Type-I error, the significance level (α) was set at 0.05 for all statistical tests. All hypothesis testing was two-tailed, and P -values $< .05$ were considered statistically significant. All statistical analysis was computed using IBM SPSS Statistics for Windows, version 16.0.

3. Results

3.1. Checklist development

The MERSQ is designed to evaluate the complete search strategy of a single systematic review. This checklist has 30 items,

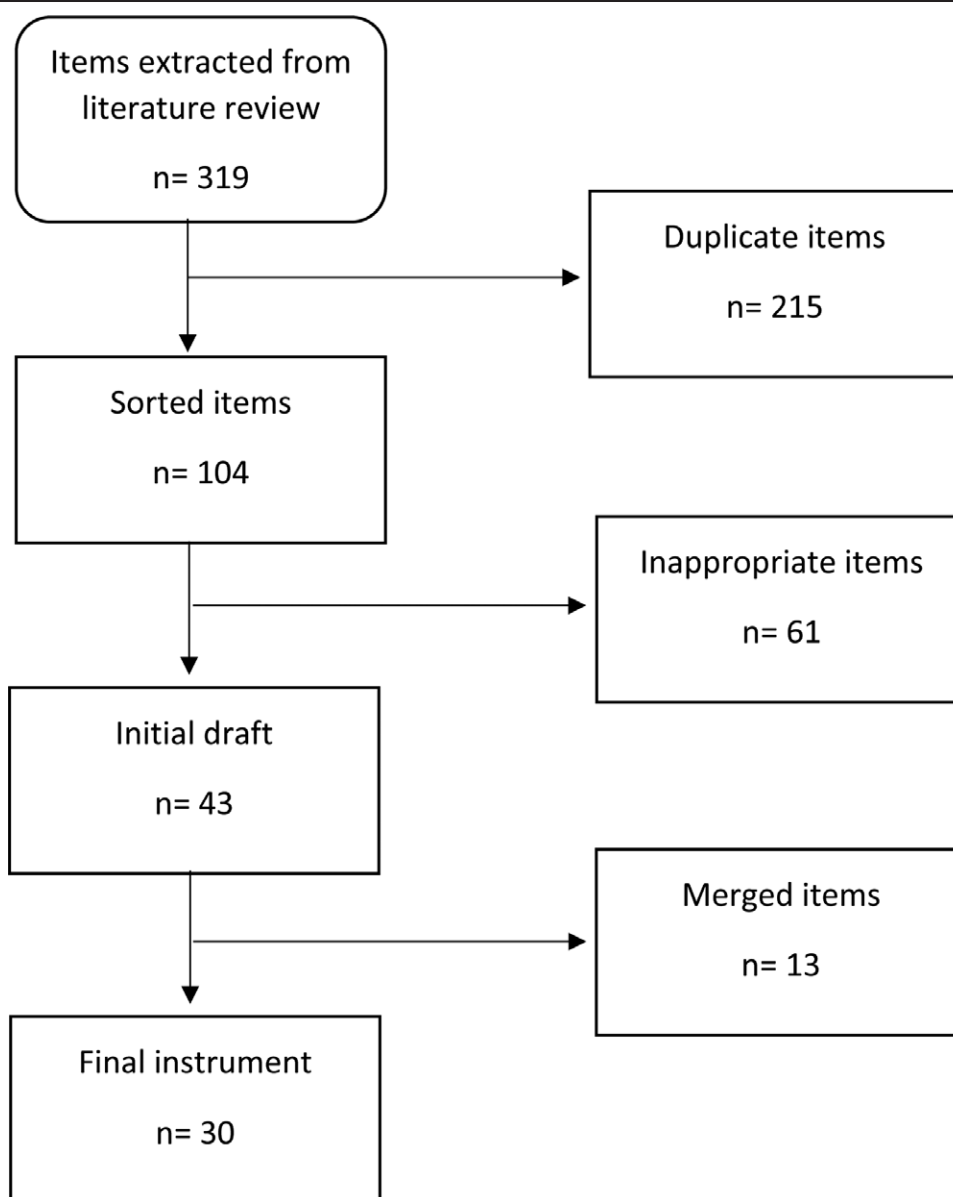


Figure 1. Data collection.

Table 1			
Internal consistency and intraclass correlation.			
	Cronbach alpha	Intraclass correlation	P-value
Reproducibility guarantee	0.715	0.971 (CI 95% [0.953–0.982])	<.001*
Comprehensiveness guarantee	0.673	0.972 (CI 95% [0.955–0.983])	<.001*
Instrument	0.827	0.978 (CI 95% [0.965–0.986])	<.001*

*Statistically significant.

and the score of each item ranged from 0 (No/Unclear) to 2 (Yes). This means that the maximum systematic score for all items was 60:36 for reproducibility and 24 for comprehensiveness of the search. Table 3 shows the MERSQ checklist. A blank part has also been provided for the rater's description regarding each item.

The first section of the tool includes questions regarding publication details, availability of study protocol and Supplementary

Material (Supplemental Digital Content, <https://links.lww.com/MD/O781>), and study questions. The second section consists of 18 items to report search in a systematic review of medical education. This section mainly emphasizes the reproducibility of the search and covers details such as the electronic sources and other search methods used, the complete search strategy, limits, and the number of retrieved records. The third section includes 12 items ensuring the search's quality and guaranteeing search comprehensiveness. It includes items that consider the diversity of search methods and databases used, the rationale for their selection, the appropriateness of the search strategy, and whether a librarian or information specialist developed the search or was consulted for searching.

3.2. Search quality in BEME studies

We assessed BEME systematic reviews from 2000 to 2020 using the MERSQ checklist and scored. Among comprehensiveness-related items, the highest score was about alternative search strategy items, including manual search, reference checking, and gray literature assessments, to which all BEME studies had

Table 2
Inter-rater reliability.

Section	Item category	Item	Cohen kappa	P-value*
Reproducibility guarantee	Search sources	Names of electronic databases	0.406	<.001
		Name of and their platforms (interfaces)	0.789	<.001
		Electronic databases' dates of coverage	0.790	<.001
		Name of sources for gray literature (if any)	0.839	<.001
		Name of journals chosen for hand-searching (if any)	0.861	<.001
		Descriptions of contacts with experts/authors (if any)	0.936	<.001
		Checking references	0.809	<.001
		Tracking citations	0.815	<.001
		The search terms and keywords have been mentioned.	0.711	<.001
		Boolean and/or proximity operators, truncations, and wildcards used	0.792	<.001
	Search strategy	Fields in which the terms were searched (e.g., title, abstract, etc)	0.882	<.001
		Limits used for the time, language, study design, or publication type	0.820	<.001
		Database special features (e.g., explosion)	1.000	<.001
		The complete search strategies of other databases have been reported, or the modifications made for their use have been mentioned.	0.894	<.001
		Rationales for limits on study type, time, etc, have been stated.	0.973	<.001
		The dates when the searches were conducted/updated have been reported for each source.	0.898	<.001
		The person(s) who searched and their role (e.g., writing, advising on, executing search strategy) has been indicated.	0.920	<.001
		Findings	0.975	<.001
	Search sources	The number of records retrieved from each source has been mentioned.	0.975	<.001
		At least 2 medical databases (PubMed, EMBASE, ...) have been searched.	0.901	<.001
Comprehensiveness guarantee	Search sources	At least one educational database (ERIC, Psych INFO, ...) has been searched.	0.926	<.001
		At least one alternative search approach has been adopted (e.g., hand searching, checking references and citations, searching for gray literature, contacting key persons, etc.)	0.805	<.001
	Search strategy	The rationale for the type and number of selected resources are justified and defensible.	0.690	<.001
		The keywords (free and controlled) have been selected appropriately according to the review question.	1.000	<.001
		The potential variants, likely synonyms, and related terms have been identified appropriately.	0.765	<.001
		Boolean and proximity operators have been used correctly.	1.000	<.001
		The search has not been restricted by language or publication status.	0.891	<.001
		Rationales for limits on study type, time, etc, are appropriate.	1.000	<.001
		The search strategies have been translated for different databases appropriately.	0.919	<.001
		Search has been run or updated within 6 months before submission date	0.888	<.001
		Searcher	0.803	<.001
		A librarian or information specialist developed or was consulted to conduct the search.	0.803	<.001

*Statistically significant.

complied. The second one was choosing keywords appropriate to the topic. On the other hand, the lowest score was about choosing synonym words in the search strategy. In half of the studies, the search was done by a nonmedical-education expert. Also, only 27% of studies updated their research 6 months before submission. Besides, the search documents were completed in only 30% of the cases.

3.3. BEME and non-BEME systematic reviews

In comparing search quality by MERSQ, the median score of 41.75 (IQR = 11.00) for BEME and 31.25 (IQR = 15.25) for non-BEME studies was calculated. In detail, the median reproducibility score was 22.00 (IQR = 8.13) and 14.50 (IQR = 9.50), and the median comprehensive score was 18.75 (IQR = 3.63) and 15.75 (IQR = 6.63) for BEME and non-BEME studies, respectively. The overall search quality of BEME studies was significantly higher than non-BEME studies ($U = 140, P = .009$). In addition, the reproducibility guarantee was significantly higher in BEME studies ($U = 124, P = .004$). However, no significant difference in comprehensive guarantee was seen ($U = 192, P = .880$).

4. Discussion

The present study designed the MERSQ checklist to determine the quality of medical education systematic reviews. This is also utilized to compare BEME and non-BEME systematic reviews and the BEME systematic reviews trend. The findings showed

that BEME ones were significantly more reproducible and had higher search quality than non-BEME ones.

In contrast to PRISMA, which is limited to reporting only on each item, it is important to note that in the MERSQ, the quality of each item was also included, reviewed, and scored. In other words, in each item, in addition to the report (reproducibility criterion), the quality is also examined (search comprehensiveness criterion). For instance, in evaluating "keywords," the topic's relevance is evaluated in addition to checking if the study reports its keywords. It is also considered for Boolean operators. In all BEME systematic reviews, the reproducibility-related score was higher than the comprehensiveness score, which indicates the compliance quality with each item. This problem is closely related to the decline in the quality of systematic reviews, the lack of a checklist specific to the area of medical education, and the justification for using the PRISMA checklist, which analyzes only the report of each item.

Regarding the quality of studies, according to search quality evaluation in HIV behavioral interventions systematic reviews over ten years, merely 6% reported their database for search, and only 35% reported search intervals among them, indicating not using common checklists in writing.^[22] Considering medical education reviews, Maggio et al^[23] reported that 56% of the literature revealed the search strategy vividly, and 41% also searched in nonmedical databases, including Web of Science or ERIC. In agreement with the authors' criteria, no search strategies reported in 34 studies were reproducible. Maggio et al did not evaluate studies published in the Medical Teacher Journal; therefore, BEME studies were not included.

In the present study, BEME reviews, compared to non-BEME reviews, had a significantly higher quality in terms of overall search and reproducibility score. The median search quality score, reproducibility, and comprehensiveness were all higher

in BEME studies than in non-BEME systematic review studies (41.75 vs 31.25, 22 vs 14.50, and 18.75 vs 15.75, respectively). Considering that a single checklist and 2 fixed authors were used in both studies, the reason for this discrepancy can

Table 3**The MERSQ checklist.****The MERSQ checklist**

Title:

Corresponding author: Date of publication: Journal:

Study protocol: Yes ☐ No ☐ Supplementary material: Yes ☐ No ☐

Study questions:

Section	Item category	No.	Item	Score	Comments
Reproducibility guarantee	Search sources	1.	Names of electronic databases	Yes <input type="checkbox"/> To some extent <input type="checkbox"/> No/Unclear <input type="checkbox"/>	
		2.	Name of and their platforms (interfaces)	Yes <input type="checkbox"/> To some extent <input type="checkbox"/> No/Unclear <input type="checkbox"/>	
		3.	Electronic databases' dates of coverage	Yes <input type="checkbox"/> To some extent <input type="checkbox"/> No/Unclear <input type="checkbox"/>	
		4.	Name of sources for gray literature (if any)	Yes <input type="checkbox"/> To some extent <input type="checkbox"/> No/Unclear <input type="checkbox"/>	
		5.	Name of journals chosen for hand-searching (if any)	Yes <input type="checkbox"/> To some extent <input type="checkbox"/> No/Unclear <input type="checkbox"/>	
		6.	Descriptions of contacts with experts/authors (if any)	Yes <input type="checkbox"/> To some extent <input type="checkbox"/> No/Unclear <input type="checkbox"/>	
		7.	Checking references	Yes <input type="checkbox"/> To some extent <input type="checkbox"/> No/Unclear <input type="checkbox"/>	
		8.	Tracking citations	Yes <input type="checkbox"/> To some extent <input type="checkbox"/> No/Unclear <input type="checkbox"/>	
Reproducibility guarantee	Search strategy	9.	The search terms and keywords have been mentioned.	Yes <input type="checkbox"/> To some extent <input type="checkbox"/> No/Unclear <input type="checkbox"/>	
		10.	Boolean and/or proximity operators, truncations, and wildcards used	Yes <input type="checkbox"/> To some extent <input type="checkbox"/> No/Unclear <input type="checkbox"/>	
		11.	Fields in which the terms were searched (e.g., title, abstract, etc)	Yes <input type="checkbox"/> To some extent <input type="checkbox"/> No/Unclear <input type="checkbox"/>	
		12.	Limits used for the time, language, study design, or publication type	Yes <input type="checkbox"/> To some extent <input type="checkbox"/> No/Unclear <input type="checkbox"/>	
		13.	Database special features (e.g., explosion)	Yes <input type="checkbox"/> To some extent <input type="checkbox"/> No/Unclear <input type="checkbox"/>	
		14.	The complete search strategies of other databases have been reported, or the modifications made for their use have been mentioned.	Yes <input type="checkbox"/> To some extent <input type="checkbox"/> No/Unclear <input type="checkbox"/>	
		15.	Rationales for limits on study type, time, etc have been stated.	Yes <input type="checkbox"/> To some extent <input type="checkbox"/> No/Unclear <input type="checkbox"/>	
		16.	The dates when the searches were conducted/updated have been reported for each source.	Yes <input type="checkbox"/> To some extent <input type="checkbox"/> No/Unclear <input type="checkbox"/>	
		17.	The person(s) who searched and their role (e.g., writing, advising on, executing search strategy) has been indicated.	Yes <input type="checkbox"/> To some extent <input type="checkbox"/> No/Unclear <input type="checkbox"/>	
		18.	The number of records retrieved from each source has been mentioned.	Yes <input type="checkbox"/> To some extent <input type="checkbox"/> No/Unclear <input type="checkbox"/>	

(Continued)

Table 3
(Continued)**The MERSQ checklist**

Comprehensiveness guarantee	Search sources	19.	At least 2 medical databases (PubMed, EMBASE, ...) have been searched.	Yes <input type="checkbox"/> To some extent <input type="checkbox"/> No/Unclear <input type="checkbox"/>
		20.	At least one educational database (ERIC, Psych INFO, ...) has been searched.	Yes <input type="checkbox"/> To some extent <input type="checkbox"/> No/Unclear <input type="checkbox"/>
		21.	At least one alternative search approach has been adopted (e.g., hand searching, checking references and citations, searching for gray literature, contacting key persons, ...)	Yes <input type="checkbox"/> To some extent <input type="checkbox"/> No/Unclear <input type="checkbox"/>
		22.	The rationale for the type and number of selected resources are justified and defensible.	Yes <input type="checkbox"/> To some extent <input type="checkbox"/> No/Unclear <input type="checkbox"/>
	Search strategy	23.	The keywords (free and controlled) have been selected appropriately according to the review question.	Yes <input type="checkbox"/> To some extent <input type="checkbox"/> No/Unclear <input type="checkbox"/>
		24.	The potential variants, likely synonyms, and related terms have been identified appropriately.	Yes <input type="checkbox"/> To some extent <input type="checkbox"/> No/Unclear <input type="checkbox"/>
		25.	Boolean and proximity operators have been used correctly.	Yes <input type="checkbox"/> To some extent <input type="checkbox"/> No/Unclear <input type="checkbox"/>
		26.	The search has not been restricted by language or publication status.	Yes <input type="checkbox"/> To some extent <input type="checkbox"/> No/Unclear <input type="checkbox"/>
		27.	Rationales for limits on study type, time, etc are appropriate.	Yes <input type="checkbox"/> To some extent <input type="checkbox"/> No/Unclear <input type="checkbox"/>
		28.	The search strategies have been translated for different databases appropriately.	Yes <input type="checkbox"/> To some extent <input type="checkbox"/> No/Unclear <input type="checkbox"/>
Comprehensiveness guarantee	Search strategy	29.	Search has been run or updated within 6 months before submission date	Yes <input type="checkbox"/> To some extent <input type="checkbox"/> No/Unclear <input type="checkbox"/>
	Searcher	30.	A librarian or information specialist developed or was consulted to conduct the search.	Yes <input type="checkbox"/> To some extent <input type="checkbox"/> No/Unclear <input type="checkbox"/>

be categorized as follows: I. Existence of protocol writing process and its judging using standard checklists, including review of protocol written in BEME and editorial committees; II. Existence of a protocol monitoring process based on the BEME systematic review by BEME member BICC committees; III. Holding training courses by BEME to increase the quality of arbitration and conduct a systematic review of BEME; and IV. Judging the reviews done by a group of experts before printing in the form of a BEME Guide.

4.1. Limitations

One of the primary constraints lies in the inadequate consideration of pragmatism, which is a crucial aspect overlooked in our research. While our current focus centers on systematic reviews from a technical standpoint, it is important to acknowledge that education encompasses intricate elements that require pragmatic considerations. Future investigations should explore whether the exclusion of certain studies poses a risk of incompleteness and potentially alters the overall ideas or conclusions. Furthermore, it appears that studies should take into account individual needs to facilitate the advancement of more comprehensive personalized medicine across various fields of healthcare.^[24] Another limitation of the present research is the systematic evaluation of reviews by only few authors. If it is possible to increase the number of experts to evaluate and score, they will have more validity

points. Also, only English-written articles were assessed in our literature review. In addition, since non-BEME studies were all selected from the same journal, the possibility of bias in selecting articles should be considered. Moreover, since no study has been done so far, comparing the MERSQ with others was impossible.

5. Conclusion

In implementing the MERSQ checklist, it is noteworthy that according to the recommendations made in the PRISMA design, journals and sources should be reviewed continuously to keep the checklist up to date and based on evidence.^[25] Considering MERSQ in the submission stages of the systematic reviews on medical education could improve the quality of these studies and more precise findings. Using the MERSQ to evaluate all non-BEME systematic reviews published so far and the BEME systematic reviews published later will significantly improve the quality of these studies and update the checklist.

Author contributions

Conceptualization: Maryam Alizadeh, Negar Omid.

Data curation: Rahem Rahmati, Fatemeh Zarimeidani, Reza Hosseini Doalame.

Formal analysis: Reza Hosseini Doalame.

Funding acquisition: Negar Omidi.

Investigation: Rahem Rahmati, Fatemeh Zarimeidani, Fatemeh Hasani, Arshin Ghaedi, Aida Bazrgar.

Methodology: Maryam Alizadeh, Rahem Rahmati, Fatemeh Zarimeidani, Reza Hosseini Doalame, Hojat Vahedi, Hamidreza Hekmat, Negar Omidi.

Project administration: Negar Omidi.

Resources: Rahem Rahmati, Fatemeh Zarimeidani.

Supervision: Maryam Alizadeh, Hojat Vahedi, Hamidreza Hekmat, Negar Omidi.

Validation: Maryam Alizadeh, Hojat Vahedi, Hamidreza Hekmat, Negar Omidi.

Writing – review & editing: Maryam Alizadeh, Rahem Rahmati, Negar Omidi

Writing – original draft: Rahem Rahmati, Fatemeh Hasani, Arshin Ghaedi, Aida Bazrgar.

References

- [1] Al-Khafaji J, Townsend RF, Townsend W, Chopra V, Gupta A. Checklists to reduce diagnostic error: a systematic review of the literature using a human factors framework. *BMJ Open*. 2022;12:e058219.
- [2] Wang J, Zhao S, Xu DR, et al. Developing evidenced-based quality assessment checklist for real practice in primary health care using standardized patients: a systematic review. *Ann Palliative Med*. 2021;10:8232–41.
- [3] Makram AM, Wang J, Vaghela G, et al. Is checklist an effective tool for teaching research students? A survey-based study. *BMC Med Educ*. 2022;22:561.
- [4] Reeves S, Koppel I, Barr H, Freeth D, Hammick M. Twelve tips for undertaking a systematic review. *Med Teach*. 2002;24:358–63.
- [5] Phillips AC, Lewis LK, McEvoy MP, et al. Development and validation of the guideline for reporting evidence-based practice educational interventions and teaching (GREET). *BMC Med Educ*. 2016;16:237.
- [6] Hammick M, Dornan T, Steinert Y. Conducting a best evidence systematic review. Part 1: From idea to data coding. *BEME Guide No. 13*. *Med Teach*. 2010;32:3–15.
- [7] Haig A, Dozier M. BEME Guide no 3: systematic searching for evidence in medical education--Part 1: sources of information. *Med Teach*. 2003;25:352–63.
- [8] Shea BJ, Grimshaw JM, Wells GA, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol*. 2007;7:10.
- [9] Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. *JAMA*. 2000;283:2008–12.
- [10] Booth A. "Brimful of STARLITE": toward standards for reporting literature searches. *J Med Libr Assoc*. 2006;94:421.
- [11] Chan T, Sebok-Syer S, Thoma B, Wise A, Sherbino J, Pusic M. Learning analytics in medical education assessment: the past, the present, and the future. *AEM Educ Train*. 2018;2:178–87.
- [12] Horsley T, Galipeau J, Petkovic J, Zeiter J, Hamstra SJ, Cook DA. Reporting quality and risk of bias in randomised trials in health professions education. *Med Educ*. 2017;51:61–71.
- [13] Majumder MAA, Haque M, Razzaque MS. Editorial: Trends and challenges of medical education in the changing academic and public health environment of the 21st century. Editorial. *Front Commun*. 2023;8:1153764.
- [14] Buja LM. Medical education today: all that glitters is not gold. *BMC Med Educ*. 2019;19:110.
- [15] Maggio LA, Samuel A, Stellrecht E. Systematic reviews in medical education. *J Grad Med Educ*. 2022;14:171–5.
- [16] Cook DA, West CP. Conducting systematic reviews in medical education: a stepwise approach. *Med Educ*. 2012;46:943–52.
- [17] Bramer WM, de Jonge GB, Rethlefsen ML, Mast F, Kleijnen J. A systematic approach to searching: an efficient and complete method to develop literature searches. *J Med Libr Assoc*. 2018;106:531–41.
- [18] Sampson M, McGowan J. Errors in search strategies were identified by type and frequency. *J Clin Epidemiol*. 2006;59:1057.e1–9.
- [19] Rethlefsen ML, Brigham TJ, Price C, et al. Systematic review search strategies are poorly described and not reproducible: a cross-sectional meta-research study. *medRxiv*. 2023;2023:23289873.
- [20] Creswell JW. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. 4th ed. SAGE; 2009.
- [21] ArtinoAR Jr, La Rochelle JS, Dezee KJ, Gehlbach H. Developing questionnaires for educational research: AMEE Guide No. 87. *Med Teach*. 2014;36:463–74.
- [22] Mullins MM, DeLuca JB, Crepaz N, Lyles CM. Reporting quality of search methods in systematic reviews of HIV behavioral interventions (2000–2010): are the searches clearly explained, systematic and reproducible? *Res Synthesis Methods*. 2014;5:116–30.
- [23] Maggio LA, Tannery NH, Kanter SL. Reproducibility of literature search reporting in medical education reviews. *Academic Med*. 2011;86:1049–54.
- [24] Ramezannezhad P, Rahmati R, Zarimeidani F, et al. Personalized patient education: patient's perspective. *Iran Biomed J*. 2024;28:66.
- [25] Pati D, Lorusso LN. How to write a systematic review of the literature. *HERD*. 2018;11:15–30.