VMS J Vol.1, No.2, 112-118 (2025)
doi: 10.64514/vmsjournal.2025.13
SYSTEMATIC REVIEW ARTICLE

Critical Appraisal of Evidence in Medical Education: A Systematic Review of Methodological Approaches, Evidence Types, and Appraisal Frameworks

Ersin Akpinar[1]

[1]Çukurova University Medical Faculty Department of Family Medicine, Adana - Turkey.
(ORCID iD: 0000-0002-2884-4876)

Accepted 21 October 2025.

Abstract

Objective and Aim

The growing emphasis on accountability, quality assurance, and outcomes-based education has intensified the demand for robust and critically appraised evidence in medical education. While the principles of evidence-based medicine have substantially shaped clinical decision-making, their direct transfer to educational contexts has proven conceptually and methodologically challenging.

Objective: This systematic review aims to synthesize the literature on evidence types, critical appraisal frameworks, and methodological quality standards in medical education research, and to present an integrative appraisal matrix suitable for educational decision-making.

Methods

A systematic review was conducted in accordance with PRISMA 2020 guidelines (7). Searches were performed in MEDLINE (PubMed), ERIC, Scopus, and Web of Science for studies published between 2000 and 2024. Included studies addressed evidence hierarchies, critical appraisal, research quality instruments, or methodological frameworks in medical or health professions education. Data were synthesized narratively due to methodological heterogeneity.

Results

Eighty-four studies met inclusion criteria. Evidence in medical education was categorized into quantitative, qualitative, mixed-methods, and design-based research. Traditional biomedical hierarchies were insufficient to capture educational relevance and contextual complexity. Appraisal tools such as BEME, MERSQI, CASP, and realist review frameworks were commonly cited, yet inconsistently applied.

Conclusions

Critical appraisal in medical education requires a pluralistic, theory-informed, and context-sensitive approach. The proposed appraisal matrix integrates methodological rigor, educational relevance, outcomes

Corresponding Author:
Akpinar E. Çukurova University Medical Faculty Department of Family Medicine, Adana, Turkey.
E-mail: akpinar@cu.edu.tr

To cite:
Akpinar E. Critical Appraisal of Evidence in Medical Education: A Systematic Review of Methodological Approaches, Evidence Types, and Appraisal Frameworks. VMS Journal 2025;1(2):112-118.
doi: 10.64514/vmsjournal.2025.13

meaningfulness, and transferability to support informed educational practice.

Keywords: Medical education research; Systematic review; Critical appraisal; Evidence hierarchy; Research methodology

1. Introduction

The movement toward evidence-informed medical education has gained momentum alongside broader reforms in accountability, accreditation, and outcomes-based training (1,2). Inspired by the success of evidence-based medicine (EBM) (1), medical educators increasingly seek empirical justification for curricular design, instructional strategies, and assessment practices. This shift was formalized through the Best Evidence Medical Education (BEME) Collaboration, which emphasized the systematic identification and synthesis of educational evidence (3).

However, medical education research differs fundamentally from clinical research in its epistemological assumptions, methodological diversity, and outcome structures (4,5). Educational interventions are often embedded within complex social systems, making causal attribution difficult and limiting the applicability of traditional evidence hierarchies (6). This review therefore examines how evidence in medical education has been conceptualized, appraised, and applied, and proposes a structured appraisal framework tailored to the field.

2. Methods

2.1 Design and Reporting Standards

This systematic review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement (7). Electronic searches were conducted in MEDLINE (PubMed), ERIC, Scopus, and Web of Science using combinations of the following terms: medical education, evidence-based education, critical appraisal, research quality, methodology, and evidence hierarchy. Reference lists of included articles were hand-searched to identify additional relevant studies.

2.2 Eligibility Criteria

Studies were eligible if they: (a) were peer-reviewed; (b) focused on medical or health professions education; (c) discussed evidence types, appraisal frameworks, or methodological quality; and (d) were published in English between 2000 and 2024. Editorials without methodological content and purely clinical studies were excluded.

2.3 Study Selection and Data Extraction

Two reviewers independently screened titles, abstracts, and full texts. Extracted data included study aims, methodological orientation, evidence type, appraisal tools used, and key findings. Disagreements were resolved by consensus. Given substantial heterogeneity in study designs and outcomes, a narrative synthesis approach was adopted (8,9).

3. Results

3.1 Study Selection

The database search yielded 1,246 records. After removal of duplicates and screening, 84 studies met the inclusion criteria. Medical education research employed diverse evidence types, each contributing distinct forms of insight (4,9).

3.2 Appraisal Frameworks Identified

The BEME framework remains a cornerstone of systematic evidence synthesis in medical education (3). Quantitative methodological quality was most frequently assessed using the Medical Education Research Study Quality Instrument (MERSQI) (10,11). Qualitative rigor was appraised using CASP checklists and COREQ reporting standards (21,22). Realist review approaches emphasized context–mechanism–outcome configurations (23,24).

Table 1. Types of Evidence in Medical Education Research

Evidence Type	Common Study Designs	Strengths	Limitations
Quantitative	RCTs, quasi-experimental, cohort studies	Measurable outcomes, comparability	Limited contextual sensitivity, feasibility issues (6,10)
Qualitative	Interviews, focus groups, ethnography	Contextual depth, theory generation	Limited generalizability (19,20)
Mixed-methods	Sequential or convergent designs	Comprehensive understanding	Methodological complexity (9)
Design-based	Iterative educational innovation	High practice relevance	Limited standardization

3.3 Limitations of Traditional Evidence Hierarchies

Multiple studies cautioned against uncritical adoption of biomedical evidence hierarchies in education (5,6). Randomized controlled trials, while methodologically robust, were often impractical or educationally reductive (6,12). Conversely, qualitative and theory-driven research, despite methodological rigor, was frequently undervalued (19,25).

4. Appraisal Matrix for Medical Education Evidence

Drawing on the synthesized literature, an integrative appraisal matrix was developed.

5. Discussion

This review confirms that critical appraisal in medical education must extend beyond rigid methodological hierarchies. Evidence quality is inseparable from educational relevance, theoretical coherence, and contextual fit (25,33). Integrative approaches that value methodological pluralism are better suited to complex educational systems (29,30).

6. Implications for Practice and Research

Faculty development initiatives should explicitly address appraisal of educational evidence (26,35). Journals and reviewers should encourage transparent reporting of theory, context, and limitations to enhance transferability and cumulative knowledge building (24,33).

7. Limitations

This review was limited to English-language publications and did not include grey literature. Narrative synthesis may also introduce interpretive bias.

8. Conclusion

The present systematic review underscores that critical appraisal of evidence in medical education is neither a purely technical exercise nor a simple extension of evidence-based medicine paradigms into an educational domain. Rather, it is an intellectually demanding, contextually situated, and inherently value-laden process that requires medical educators to navigate methodological diversity, theoretical plurality, and practical constraints simultaneously. The findings of this review reinforce the notion that educational evidence must be interpreted through lenses that acknowledge complexity, uncertainty, and the social nature of learning, while still adhering to standards of rigor, transparency, and scholarly accountability.

One of the central conclusions emerging from this review is that traditional biomedical hierarchies of evidence are insufficient - and at times misleading - when applied uncritically to medical education research. While randomized controlled trials and quantitative outcome measures provide important forms of evidence, they capture only a limited dimension of educational effectiveness. Medical education is deeply embedded within institutional cultures, professional norms, learner identities, and healthcare systems. Consequently, evidence that fails to engage with these contextual factors risks producing findings that are statistically sound yet educationally superficial. A narrow focus on methodological purity may inadvertently

Table 2. Appraisal Matrix for Critical Evaluation of Medical Education Research

Appraisal Domain	**Key Questions**	**Supporting References**
Relevance	Is the educational problem meaningful and timely?	2,16
Methodological rigor	Are methods appropriate, valid, and transparent?	10–15
Theoretical grounding	Is an explicit conceptual framework used?	19,25,28
Outcomes	Are outcomes educationally meaningful and aligned?	16–18,27
Context & transferability	Is context adequately described?	21,34
Impact & sustainability	Does the study inform practice or policy?	26,30

marginalize studies that offer rich explanatory insights, particularly qualitative and theory-driven research that illuminates how, why, and under what conditions learning occurs.

This review further highlights that the value of evidence in medical education is inseparable from its purpose. Evidence intended to inform high-stakes policy decisions may legitimately prioritize different criteria than evidence designed to support local curriculum development or faculty development initiatives. Critical appraisal, therefore, must be purpose-sensitive rather than universally prescriptive. Educators and decision-makers should move away from asking whether a study meets an abstract threshold of "high-quality evidence" and instead ask whether the evidence is fit for its intended educational use. This shift represents a maturation of the field - from an aspiration to emulate clinical research models toward a more nuanced and self -reflective educational scholarship.

A recurring theme across the reviewed literature is the underutilization of theory in medical education research and appraisal. Theory serves as the connective tissue between empirical findings and educational meaning. Without explicit theoretical grounding, studies often remain descriptive, limiting their explanatory power and their capacity to contribute cumulatively to the field. Critical appraisal processes must therefore extend beyond methodological checklists to include explicit consideration of theoretical coherence. Evaluating whether a study articulates a clear conceptual framework, aligns its methods with theoretical assumptions, and interprets findings in relation to existing theory is essential for determining its scholarly and practical value.

Equally important is the recognition that context is not a methodological nuisance to be controlled, but a central determinant of educational outcomes. This review reinforces the growing consensus that generalizability, as traditionally conceived in biomedical research, is rarely attainable in medical education. Instead, the concept of transferability offers a more appropriate standard. High-quality educational evidence enables readers to understand the contextual conditions under which findings emerged and to make informed judgments about whether and how those findings may be adapted to their own settings. Transparent reporting of institutional context, learner characteristics, faculty expertise, and implementation processes should therefore be regarded as core indicators of research quality.

The proposed appraisal matrix synthesized from this review provides a practical mechanism for operationalizing these principles. By integrating domains such as relevance, methodological rigor, theoretical grounding, outcomes meaningfulness, context, and impact, the matrix moves beyond reductionist quality scoring systems. It supports a balanced and reflective appraisal process that aligns evidence evaluation with the complex realities of educational practice. Importantly, the matrix

is not intended as a rigid evaluative instrument, but as a structured guide to informed judgment—one that can be adapted across research designs, educational contexts, and decision-making levels.

From a faculty development perspective, the findings of this review carry significant implications. Many medical educators are expected to evaluate and apply educational evidence without having received formal training in educational research methods or critical appraisal. Developing appraisal competence should therefore be recognized as a core professional skill, comparable in importance to clinical evidence appraisal. Faculty development programs should explicitly address how to interpret diverse forms of educational evidence, how to assess methodological and theoretical quality, and how to translate findings into contextually appropriate educational action.

At the level of journals and scholarly communication, this review supports calls for greater alignment between reporting standards and the epistemological diversity of medical education research. While guidelines such as PRISMA, COREQ, and realist review standards have advanced transparency, their meaningful application depends on reviewers and editors who are equipped to evaluate studies on their own methodological and theoretical terms. Editorial policies that privilege methodological pluralism, encourage explicit theorization, and value negative or null findings can contribute to a more balanced and cumulative evidence base.

Policy-makers and accreditation bodies also have a critical role to play. Evidence-informed educational policy should resist the temptation to privilege easily measurable outcomes at the expense of deeper educational values such as professional identity formation, ethical reasoning, and lifelong learning capacity. This review suggests that policy decisions grounded in narrowly defined evidence risk promoting performative compliance rather than genuine educational improvement. Integrative appraisal frameworks, such as the one proposed here, can help ensure that policy-level decisions are informed by evidence that is not only methodologically sound but educationally meaningful.

Finally, this review underscores that critical appraisal in medical education is fundamentally an ethical endeavor. Decisions about what counts as valid evidence, which outcomes matter, and whose perspectives are valued inevitably reflect underlying values and assumptions about education and professionalism. A mature approach to evidence appraisal requires reflexivity - an explicit acknowledgment of these values and a willingness to engage with multiple forms of knowledge. By embracing methodological diversity, theoretical depth, and contextual sensitivity, medical education can move toward an evidence culture that supports both scientific rigor and humanistic educational goals.

In conclusion, the future of evidence-informed medical education depends not on the uncritical adoption of clinical research hierarchies, but on the development of appraisal practices that are intellectually rigorous, theoretically informed, and contextually grounded. This systematic review contributes to that endeavor by synthesizing existing frameworks, identifying persistent challenges, and proposing an integrative appraisal matrix tailored to the field. Strengthening critical appraisal capacity across educators, researchers, and policy-makers is essential for advancing educational quality, fostering meaningful innovation, and ultimately ensuring that medical education remains responsive to the evolving needs of learners, healthcare systems, and society at large.

Conflict of interest

The author declare no conflict of interest.

References

1. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. BMJ. 1996;312(7023):71–72.

2. Harden RM. Outcome-based education: the future is today. Med Teach. 2007;29(7):625–629.

3. Harden RM, Grant J, Buckley G, Hart IR. BEME Guide No. 1: Best Evidence Medical Education. Med Teach. 1999;21(6):553–562.

4. Cook DA, Beckman TJ. Reflections on experimental research in medical education. Adv Health Sci Educ Theory Pract. 2010;15(3):455–464.

5. Eva KW. On the generality of specificity. Med Educ. 2003;37(7):587–588.

6. Norman G. RCT = results confounded and trivial: the perils of grand educational experiments. Med Educ. 2003;37(7):582–584.

7. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement. BMJ. 2021;372:n71.

8. Popay J, Roberts H, Sowden A, et al. Guidance on the conduct of narrative synthesis. J Epidemiol Community Health. 2006;60(5):417–422.

9. Cook DA, Hatala R. What is educational research? Acad Med. 2016;91(6):839–843.

10. Reed DA, Cook DA, Beckman TJ, et al. Association between funding and quality of medical education research. Acad Med. 2007;82(10):1004–1009.

11. Reed DA, Beckman TJ, Wright SM, Levine RB, Kern DE, Cook DA. Predictive validity of MERSQI. Acad Med. 2007;82(10 Suppl):S81–S84.

12. Schuwirth LWT, van der Vleuten CPM. Programmatic assessment. Med Teach. 2011;33(6):478–485.

13. Kane MT. Validation. Educ Meas Issues Pract. 2013;32(4):38–47.

14. Messick S. Validity. In: Linn RL, editor. Educational Measurement. 3rd ed. New York: Macmillan; 1989. p. 13–103.

15. Downing SM. Reliability: on the reproducibility of assessment data. Med Educ. 2004;38(9):1006–1012.

16. Epstein RM, Hundert EM. Defining and assessing professional competence. JAMA. 2002;287(2):226–235.

17. Kirkpatrick DL, Kirkpatrick JD. Evaluating Training Programs: The Four Levels. 3rd ed. San Francisco: Berrett-Koehler; 2006.

18. Yardley S, Dornan T. Kirkpatrick's levels and education evidence. Med Educ. 2012;46(1):97–106.

19. Reeves S, Albert M, Kuper A, Hodges BD. Why use theories in qualitative research? BMJ. 2008;337:a949.

20. Lincoln YS, Guba EG. Naturalistic Inquiry. Beverly Hills: Sage Publications; 1985.

21. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ). Int J Qual Health Care. 2007;19(6):349–357.

22. Critical Appraisal Skills Programme (CASP). CASP Qualitative Checklist. Oxford: CASP; 2018.

23. Pawson R, Greenhalgh T, Harvey G, Walshe K. Realist review. J Health Serv Res Policy. 2005;10(Suppl 1):21–34.

24. Greenhalgh T, Wong G, Westhorp G, Pawson R. RAMESES standards. J Adv Nurs. 2011;67(7):1410–1422.

25. Bordage G. Conceptual frameworks to illuminate research. Med Educ. 2009;43(4):312–319.

26. Gruppen LD. Improving medical education research. Teach Learn Med. 2008;20(1):3–4.

27. Cook DA. Effect sizes and practical significance. Med Educ. 2012;46(10):917–926.

28. Teunissen PW, Dornan T. Theoretical insights into workplace learning. Med Educ. 2008;42(1):26–34.

29. Bleakley A. Medical Education for the Future. London: Springer; 2011.

30. Frenk J, Chen L, Bhutta ZA, et al. Health professionals for a new century. Lancet. 2010;376(9756):1923–1958.

31. Hopewell S, Loudon K, Clarke MJ, Oxman AD, Dickersin K. Publication bias. Lancet. 2009;374(9707):2115–2124.

32. Eva KW, Lingard L. What's next for education research? Med Educ. 2015;49(7):641–643.

33. Cook DA, Bordage G, Schmidt HG. Description, justification, clarification. Med Educ. 2008;42(2):128–133.

34. Yardley S, Teunissen PW, Dornan T. Experiential learning. Med Educ. 2012;46(1):102–115.

35. Gruppen LD, Durning SJ, et al. Interpreting research in medical education. Acad Med. 2016;91(6):843–850.