

Metacognition-related education in medical residency: a scoping review

Andres Felipe Yepes-Velasco

andres.yepes@fsfb.edu.co

Fundación Santa Fe de Bogotá

Dormar David Barrios

Hospital Universitario de San Vicente Fundación

Leonardo Andres Gomez

Hospital Universitario Mayor, Mederi

Research Article

Keywords: metacognition, self-regulated learning, reflective practice, cognitive bias, clinical reasoning, graduate medical education, residency, scoping review

Posted Date: October 6th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10217839/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background.

Metacognition — the knowledge and regulation of one's own cognitive processes — is increasingly recognized as a core component of clinical reasoning and reflective practice. In postgraduate medical education, a heterogeneous body of educational interventions, validated instruments, and observational studies has emerged around metacognitive constructs, but the breadth and nature of this literature has not been systematically mapped.

Methods.

A scoping review was conducted following the JBI Manual for Evidence Synthesis and reported in accordance with PRISMA-ScR. A Population–Concept–Context (PCC) framework was used: residents/fellows in postgraduate medical training; metacognition and related concepts (self-regulated learning, reflective practice, cognitive debiasing, calibration, self-assessment, structured debriefing); any educational or clinical setting. Electronic searches were conducted in MEDLINE (PubMed), Embase, ERIC, Web of Science, and Cochrane CENTRAL during March and April 2026. Each record was independently screened by two of the three reviewers; conflicts were resolved by the third reviewer. Data charting used a pre-piloted form (initial pilot with 15 articles) organising findings by metacognitive regulation versus metacognitive knowledge per Flavell (1979); extraction was verified by consensus among the three authors. As per JBI guidance for scoping reviews, no formal risk-of-bias appraisal was performed.

Results.

Of 4,985 records identified through database searching, 3,723 remained after deduplication; 386 were sought as full text, 356 were assessed for eligibility, and 92 studies were included. Studies were stratified into Tier 1 (n = 80, core metacognitive concept) and Tier 2 (n = 12, multiple adjacent concepts). When mapped to the Zimmerman SRL cycle, included studies addressed: planning interventions (40/92, 43%); monitoring interventions (47/92, 51%); reflection interventions (40/92, 43%); calibration interventions (57/92, 62%); and measurement/observation studies (5/92, 5%) (categories are non-mutually-exclusive). The most represented specialties were Internal Medicine (33, 35.9%), Pediatrics (10, 10.9%), Family Medicine (8, 8.7%), Surgery (7, 7.6%), Emergency Medicine (7, 7.6%) and Anesthesiology (6, 6.5%). Geographic distribution was Norteamerican-dominated: USA (51, 55.4%) and Canada (17, 18.5%), followed by the Netherlands (6), United Kingdom (4) and France (4). Predominant designs were RCT (13, 20%), mixed-methods (14, 17.4%), pre-post without control (9, 12.0%), qualitative (12, 10.9%), and educational innovation reports (6, 9.8%). The median sample size was 33 participants (range 4–963, reported in 41/92 studies, 45%).

Conclusions.

The empirical literature on metacognition in postgraduate medical training is moderately extensive but conceptually and methodologically heterogeneous. Studies cluster around self-assessment, clinical reasoning, and self-regulated learning, with substantial concentration in Internal Medicine and Norteamerican institutions. Notable evidence gaps include longitudinal randomized trials with objective outcomes, validated instruments for clinical metacognition, standardized reporting of complex interventions (TIDieR), and representation of underrepresented specialties and geographies. This map of the field can orient targeted systematic reviews in specific sub-domains and guide future primary research.

Background

Clinical reasoning is the central cognitive competency that distinguishes expert physicians [1]. Four decades of research in cognitive psychology and medical education have established that expert clinical performance depends not only on the volume of knowledge but also on the capacity to actively monitor and regulate one's own cognitive processes during decision-making – the phenomenon known as metacognition [2, 3]. Flavell (1979) originally distinguished metacognitive knowledge (what an individual knows about their own cognitive processes, including knowledge of person, task and strategy) from metacognitive regulation (the actions of planning, monitoring and evaluation applied during learning and problem-solving) [4].

In medicine, this distinction translates into capacities such as recognizing one's diagnostic uncertainty, calibrating confidence with actual accuracy, detecting cognitive biases in decision-making and applying deliberate mitigation strategies [5, 6]. Kahneman's dual-process model [7], applied to clinical cognition by Croskerry [8, 9], proposes alternation between System 1 (intuitive, fast) and System 2 (analytical, deliberate); the former is efficient but vulnerable to systematic biases (anchoring, availability, framing, premature closure) that contribute to diagnostic error, which accounts for 6–17% of preventable adverse events [10, 11].

In parallel, the self-regulated learning (SRL) theory developed by Zimmerman [12] and applied to medical education by Brydges, Sandars, Cleary and others [13–15] emphasises that effective acquisition of clinical competencies requires active performance monitoring, goal setting, strategy selection and critical reflection. Under this conceptual umbrella, numerous educational interventions have been proposed for postgraduate trainees: exam wrappers [16], deliberate reflection [17], WOOP (Wish-Outcome-Obstacle-Plan) [18], technology-augmented debriefing [19], cognitive debiasing training [20] and cognitive apprenticeship [21].

This literature presents three challenges for educators and researchers. First, terminological proliferation with inconsistent operational definitions [22] – concepts such as "metacognition", "self-regulation", "reflection", "self-assessment" and "debiasing" are used interchangeably or with variable definitions.

Second, marked heterogeneity in designs (from randomized trials to qualitative content analyses), interventions (from single sessions to longitudinal curricula), populations (different specialties, PGY levels and countries) and outcomes (from self-reported instruments to objective diagnostic performance). Third, a substantial portion of the literature is descriptive, observational or qualitative – study types typically excluded from traditional systematic reviews focused on effectiveness.

A scoping review according to the Joanna Briggs Institute (JBI) methodological framework [23, 24] is the appropriate design to address these challenges. Unlike systematic reviews that evaluate effectiveness through quantitative synthesis of trials, scoping reviews systematically map the extent, nature and characteristics of a body of literature, identify key concepts and knowledge gaps, and provide a structured foundation for future research [25, 26]. To our knowledge, no comprehensive scoping review has previously mapped the breadth of metacognition-related interventions, measures and empirical studies in postgraduate medical training.

Objective and review questions

The aim of this scoping review was to map the extent, nature and characteristics of the empirical literature on metacognition-related educational interventions, instruments and observational studies in residents and fellows of postgraduate medical training programs. Specific review questions were:

1. What types of metacognition-related interventions, instruments and observational studies have been reported in postgraduate medical training, and how do they distribute by Zimmerman's SRL phases (planning, monitoring, reflection, calibration) plus measurement/observation studies?
2. What populations (specialties, PGY levels, countries) and study designs predominate?
3. What outcomes and measurement instruments are used, and how objective are these outcomes?
4. What knowledge gaps prevent the field from advancing toward effectiveness trials or focused systematic reviews?

Methods

Protocol

This scoping review was designed following the JBI Manual for Evidence Synthesis [24] and is reported in accordance with the PRISMA Extension for Scoping Reviews (PRISMA-ScR) [27]. The protocol was developed a priori by consensus among the three authors. The completed PRISMA-ScR checklist is provided in Additional file 1.

Eligibility criteria

Following the Population–Concept–Context (PCC) framework recommended by JBI:

Population: residents or fellows in postgraduate medical training programs (any specialty, any PGY level, any country).

Concept: empirical studies addressing one or more of the following metacognition-related concepts: metacognition (per Flavell), self-regulated learning, reflective practice, cognitive bias/debiasing, calibration, self-assessment, structured debriefing, cognitive apprenticeship/scaffolding, deliberate practice, goal-setting/learning plans, WOOP/mental contrasting, exam wrappers, metareasoning, cognitive load theory.

Context: any educational or clinical setting (classroom, simulation, ward, ambulatory, OSCE, virtual).

Eligible study types: empirical studies of any design, including experimental, quasi-experimental, observational, qualitative, mixed-methods, descriptive (educational innovation reports), and psychometric/validation studies. Languages: English and Spanish. No date restriction.

Exclusion criteria: editorials, commentaries, letters and perspective pieces without primary data; conference supplements without full text; protocols of ongoing studies; previous systematic, scoping or narrative reviews; studies in undergraduate medical students only, faculty only, or non-medical professions; studies without an identifiable metacognitive component.

Information sources and search strategy

Electronic searches were conducted in five databases: MEDLINE (PubMed), Embase, ERIC (ProQuest), Web of Science Core Collection, and Cochrane CENTRAL. Searches were performed during March and April 2026. The search strategy combined two thematic blocks using Boolean operators (Population AND Concept), implemented with controlled vocabulary (MeSH, Emtree, ERIC descriptors) and truncated free-text terms in title/abstract/keyword fields. Cochrane Library was also checked for existing Cochrane Reviews and Protocols (0 found, confirming the absence of prior systematic reviews on this topic). Reference lists of all included full-text studies were manually screened to identify additional relevant studies. The complete search strategy by database is provided in Additional file 2. Supplementary figures showing the temporal distribution, tier stratification, specialty distribution, design distribution and geographic distribution are provided in Additional file 5.

Study selection

All retrieved records were imported into Rayyan QCRI [28] for automated and manual deduplication. Each record was screened independently by two of the three authors (title/abstract phase, then full-text phase); conflicts at any stage were resolved by the third author. No fourth reviewer was required. The two-reviewer-plus-tiebreaker model was applied to both title/abstract and full-text screening.

Data charting

A standardised charting form was discussed and refined in the research group. The form captured, per study: year of publication, first author, country, journal, specialty, PGY level, sample size, study design, intervention or instrument description, outcomes and measurement instruments, and metacognitive findings. Metacognitive findings were organised according to Flavell's original taxonomy (1979) [4] as relating to metacognitive regulation (planning, monitoring, evaluation, control of cognition) or metacognitive knowledge (knowledge of person, task, strategy). An initial pilot test was carried out with the first 15 included articles, and through iterative discussion and feedback among the three authors the charting form was refined to its final version. Data extraction was verified by consensus among the three authors; no independent duplicate extraction was performed beyond the pilot phase. As per JBI guidance for scoping reviews [24], no formal risk-of-bias appraisal of individual studies was performed; the rationale and implications of this decision are discussed in Limitations.

Synthesis and presentation

Results were synthesised descriptively and visually as recommended by JBI [24]. We present: (a) the PRISMA-ScR flow diagram (Fig. 1); (b) frequency distributions of populations (specialty, country), study designs and PCC concepts (Table 2; Additional file 5 (Supplementary Figures S3–S5)); (c) a typology of interventions and studies mapped to Zimmerman's SRL cycle phases (planning, monitoring, reflection, calibration) plus a measurement/observation category (Table 3; Fig. 3); (d) a tier stratification by concept density (Tier 1 = core metacognitive concept present; Tier 2 = multiple adjacent concepts; Additional file 5, Supplementary Figure S2). Coding definitions are provided in Table 1 (coding dictionary). No quantitative meta-analysis was undertaken, as is standard for scoping reviews. The complete charted dataset for all 92 included studies – including bibliographic data, intervention characterisation, outcomes and concept tagging – is provided in Additional file 4.

Results

Study selection

Database searching identified 4,985 records (MEDLINE/PubMed n = 1,792; Embase n = 1,209; ERIC n = 1,152; Web of Science n = 716; Cochrane CENTRAL n = 116). After deduplication in Rayyan, 3,723 unique records remained. After title/abstract screening 3,337 were excluded, leaving 386 reports sought for full text. Of these, 30 were not retrievable, and 356 were assessed for eligibility. After applying PCC criteria and resolving conflicts via the third reviewer, 92 studies met inclusion criteria. The PRISMA-ScR flow diagram (Fig. 1) summarises the selection process and reasons for exclusion at each stage.

Coding dictionary

To ensure transparency in category assignment, Table 1 presents the coding dictionary used to classify each study by specialty, design, and metacognitive phase.

Table 1
Coding dictionary.

Variable	Categories used	Operational definition
Specialty	Internal Medicine, Pediatrics, Family Medicine, Surgery, Emergency Medicine, Anesthesiology, Obstetrics and Gynecology, Radiology, Otolaryngology, Neurology, Critical Care, Hematology/Oncology, Multiple specialties, Not classifiable	Primary specialty of residency programme reported by authors. When trainees from multiple programmes participated, classified as "Multiple specialties"; when not clearly stated, "Not classifiable".
Study design	RCT, Quasi-experimental, Pre-post (no control), Pre-post (with follow-up), Cross-sectional, Cohort (prospective), Cohort (retrospective), Case-control, Mixed-methods, Qualitative, Educational innovation report, Pilot/feasibility, Psychometric/validation, Case series/report, Observational, Descriptive	Highest-level design label per author description, normalised to standard taxonomy. Cluster RCTs grouped with RCT. Educational innovation reports = descriptive accounts of curricular implementations with empirical evaluation.
Metacognitive phase (Zimmerman cycle)	Planning interventions, Monitoring interventions, Reflection interventions, Calibration interventions, Measurement/observation studies	Planning = pre-task processes (goal-setting, learning plans, WOOP, exam wrappers pre-task). Monitoring = during-task processes (debiasing, cognitive forcing, scaffolding, deliberate practice). Reflection = post-task (reflective writing, debriefing, error analysis). Calibration = confidence-accuracy alignment, self-assessment accuracy. Measurement/observation = studies whose primary purpose is to assess or describe metacognitive constructs rather than intervene. Categories are non-mutually-exclusive.
PCC concept tier	Tier 1, Tier 2	Tier 1 = at least one "core" metacognitive concept present (metacognition, metareasoning, SRL, reflective practice, cognitive bias/debiasing, calibration, WOOP, exam wrapper). Tier 2 = no core concept but multiple (≥ 3) adjacent concepts present.

General characteristics of the cohort

The 92 included studies were published between 1997 and 2026, with concentration in the last decade. Sample size was reported in 41/92 studies (45%); median = 33 participants (range 4–963). Tier 1 (core metacognitive concept) comprised 80 studies (87%) and Tier 2 (multiple adjacent concepts) 12 studies (13%). Table 2 summarises general characteristics. The distribution of included studies by medical specialty was strongly skewed towards Internal Medicine (34/92, 37%), Pediatrics (10/92, 11%), Family Medicine (8/92, 9%), Emergency Medicine (7/92, 8%), Surgery (7/92, 8%) and Anesthesiology (6/92, 7%);

the remainder was distributed across Obstetrics and Gynecology, Radiology, Multiple specialties and other subspecialties (Neurology, Pathology, Psychiatry, Otolaryngology, Critical Care, Geriatrics, each with one or two studies). Predominant study designs were mixed-methods (14, 15%), qualitative (13, 14%), randomised controlled trials (13, 14%), cross-sectional (10, 11%), pre-post without control (9, 10%) and educational innovation reports (6, 7%); the remainder consisted of quasi-experimental, pilot/feasibility, cohort, case-based and other designs. Geographic distribution was strongly concentrated in North America: the United States (51, 55%) and Canada (17, 18%) together accounted for 74% of the cohort, followed by the Netherlands (6), the United Kingdom (4), France (4) and a small number of studies from Denmark, Australia, Germany, Ireland, South Korea, Singapore, China, India and the Philippines. No included study originated from Latin America or sub-Saharan Africa. Detailed frequency tables (Supplementary Tables ST1–ST3) and the corresponding visualisations (Supplementary Figures S3–S5) are provided in Additional file 5.

Table 2
General characteristics of included studies.

Characteristic	Category	# Studies	%
PCC concept tier	Tier 1 (core metacognitive concept)	80	87.0%
	Tier 2 (multiple adjacent concepts)	12	13.0%
Sample size	Reported in publication	41	45%
	Not reported	51	55%
	Median sample size	33	—
	Range	4–963	—

PCC concept frequency

The most frequently represented PCC concepts across the 92 included studies (concepts are non-mutually-exclusive) were: self-assessment (41, 45%), clinical reasoning (31, 34%), metacognition (explicit term, 27, 29%), self-regulated learning (26, 28%), goal-setting/learning plans (22, 24%), cognitive bias/debiasing (22, 24%), reflective practice (20, 22%), and diagnostic accuracy/error (17, 18%). Less common were calibration (10, 11%), WOOP/mental contrasting (2), and cognitive load theory (0). Frequencies are visualised in Fig. 2.

Typology by metacognitive phase (Zimmerman SRL cycle)

To address Review question 1 — what types of interventions exist and what metacognitive processes they target — we classified each of the 92 included studies into one or more of five categories: planning interventions, monitoring interventions, reflection interventions, calibration interventions, and measurement/observation studies. Categories are non-mutually-exclusive (a single study may target multiple phases). The distribution is summarised in Fig. 3 and detailed in Table 3.

Table 3

Typology of interventions and studies by metacognitive phase, with verified examples from the included corpus (n = 92).

Phase / Category	Cognitive processes targeted	Interventions or study types found	Verified examples from the included corpus	Studies in cohort (n, %)
Planning interventions (pre-task)	Task analysis, goal setting, strategic planning, motivational appraisal, anticipation of obstacles	Individualised learning plans (ILPs); goal-setting curricula; WOOP / mental contrasting; exam wrappers (pre-task component); self-regulated learning frameworks	• Ding et al. (2025) [35]: The Metacognitive Enhancement Triad model enhances learning outcomes in fir • Go et al. (2024) [37]: Measuring Chief Resident Skill and Entrustment Progression in An Operative • Awad et al. (2024) [32]: Exploring perspectives of personal learning plans in a residency programme • Nishii A et al. (2021) [40]: Identification of promotion and prevention associated surgeon behaviors in	40 (43%)
Monitoring interventions (during-task)	Real-time self-monitoring, attention/strategy control, error detection, Type 1 ↔ Type 2 switching, cognitive load management	Debiasing/cognitive forcing strategies; cognitive apprenticeship; deliberate practice with monitoring; think-aloud protocols; scaffolded simulation; eye-tracking during reasoning	• Roumeliotis et al. (2026) [41]: Competency-based simulation and evaluation in medical education: clinical d • Abernethy et al. (2026) [31]: Integrating AI Scribes into Medical Education: Guardrails for Preserving CI • Choi et al. (2023) [33]: Residents need competence	47 (51%)

Phase / Category	Cognitive processes targeted	Interventions or study types found	Verified examples from the included corpus	Studies in cohort (n, %)
			not confidence: A retrospective evaluation of the • Farhat et al. (2017) [36]: Development of a longitudinal curriculum for a residency medical education	
Reflection interventions (post-task)	Self-judgement of performance, causal attribution, identification of learning points, plan adjustment, transfer to next episode	Reflective writing/journaling; structured debriefing (PEARLS, eye-tracking-augmented); group narrative discussion of errors; coaching with feedback; exam wrappers (post-task component)	• Helm et al. (2026) [38]: Diagnostic accuracy and metacognition in dermatology: A cross-sectional ana • Abernethy et al. (2026) [31]: Integrating AI Scribes into Medical Education: Guardrails for Preserving CI • Shelke et al. (2025) [42]: IMPACT OF FEEDBACK DURING OSCE AS A LEARNING TOOL IN POSTGRADUATE STUDENTS • Go et al. (2024) [37]: Measuring Chief Resident Skill and Entrustment Progression in An Operative	40 (43%)
Calibration interventions (integrative meta-process)	Confidence-accuracy alignment, correction of over-/under-confidence, updating of self-competence models	Explicit calibration training; self-assessment accuracy assessment; confidence-accuracy correlation tasks; multi-source feedback with objective data	• Helm et al. (2026) [38]: Diagnostic accuracy and metacognition in dermatology: A cross-sectional ana • Roumeliotis et al. (2026) [41]: Competency-based simulation and	57 (62%)

Phase / Category	Cognitive processes targeted	Interventions or study types found	Verified examples from the included corpus	Studies in cohort (n, %)
			evaluation in medical education: clinical d • Choi et al. (2023) [33]: Residents need competence not confidence: A retrospective evaluation of the • Farhat et al. (2017) [36]: Development of a longitudinal curriculum for a residency medical education	
Measurement / observation studies	Description, assessment or modelling of metacognitive constructs without an explicit interventional component	Psychometric studies of metacognition instruments; cross-sectional studies of SRL trajectories; observational studies of clinician reasoning processes; correlational designs	• Torti JMI et al. (2026) [44]: In Support of Meaningful Assessment and Feedback: A Study of Clinical Reaso • Tallentire et al. (2011) [43]: Understanding the behaviour of newly qualified doctors in acute care contex • Mitchell et al. (2009) [39]: A New Measure of the Cognitive, Metacognitive, and Experiential Aspects of • Crowley et al. (2007) [34]: Evaluation of an intelligent tutoring system in pathology: Effects of exter	5 (5%)

Categories are non-mutually-exclusive; the sum of "Studies in cohort" exceeds N because a single study can target multiple phases.

Patterns and evidence gaps

Several descriptive patterns emerge from the cohort:

1. North American predominance: 68/92 studies (74%) originate from the USA or Canada. The Netherlands, United Kingdom and France contribute the next largest blocks.
2. Specialty concentration: Internal Medicine accounts for 34/92 (37%) of studies, followed by Pediatrics, Family Medicine, Surgery, Emergency Medicine and Anesthesiology. Specialties such as Psychiatry, Pathology and procedural subspecialties are underrepresented or absent.
3. Predominance of non-randomised designs: RCTs comprise 13 of 92 studies (14%); mixed-methods, pre-post, qualitative and educational innovation reports together account for the majority.
4. Conceptual overlap: self-assessment, self-regulated learning, reflective practice and metacognition are frequently used together within single studies without clearly distinct operational definitions.
5. Asymmetric phase coverage: calibration concepts appear in more than half of the studies, while measurement/observation studies are scarce despite their importance for psychometric advancement of the field.

Knowledge gaps identified through this mapping include:

1. a. Longitudinal randomised trials (> 1 year follow-up) with objective clinical reasoning outcomes are scarce.
2. b. Validated, clinically-specific instruments for metacognitive measurement are lacking; many studies rely on the Metacognitive Awareness Inventory (MAI), originally validated in academic settings.
3. c. Reporting of complex interventions is inconsistent; few studies follow TIDieR guidance [29].
4. d. Procedural and diagnostic subspecialties (Anesthesiology, Radiology, Pathology, Psychiatry) are underrepresented.
5. e. Few studies document transfer to actual clinical practice or patient outcomes.

Discussion

Summary of findings

This scoping review mapped a moderately extensive and conceptually heterogeneous literature on metacognition-related interventions, measurement instruments and observational studies in postgraduate medical training. From 3,723 unique records screened, 92 studies met inclusion criteria. The cohort is dominated by non-randomised designs (RCTs only 14%), Norteamerican institutions (74%), and the specialties of Internal Medicine, Pediatrics, Family Medicine, Surgery, Emergency Medicine and Anesthesiology. When mapped to the Zimmerman SRL cycle, the largest single phase category is

calibration (57/92, 62%), followed by monitoring (47/92), reflection (40/92) and planning (40/92); measurement/observation studies are the smallest category (5/92).

Comparison with previous reviews

Our findings extend and complement previous focused reviews. Lambe [30] systematically reviewed dual-process cognitive interventions for diagnostic reasoning and reported a heterogeneous literature with limited methodological quality, focused mainly on medical students. Brydges [13] meta-analysed self-regulated learning in simulation-based training and reported substantial heterogeneity with moderate but inconsistent effects. The present scoping review confirms these observations and extends them to the specific subfield of postgraduate training through quantitative mapping of specialties, countries, designs and phase-typology – none of which has, to our knowledge, been previously mapped in this manner.

Implications for educational practice

The descriptive map nonetheless suggests considerations for educators: (a) the incorporation of metacognitive components into residency curricula rests on a theoretically rich but empirically heterogeneous evidence base; (b) self-reported instruments such as the MAI are widely used but require complementary objective measures; (c) interventions span multiple phases of the metacognitive cycle, suggesting that programmes seeking to develop metacognitive competence may need coordinated approaches addressing planning, monitoring, reflection and calibration rather than isolated activities.

Implications for research

Six priorities for future research emerge from this map:

1. Randomised controlled trials with adequate power and longitudinal follow-up, using TIDieR-compliant intervention descriptions [29].
2. Development and validation of clinically-specific metacognition instruments that complement or extend the MAI for postgraduate medical contexts.
3. Terminological standardisation through international Delphi consensus, to facilitate comparability across studies and enable future quantitative syntheses.
4. Longitudinal designs (> 1 year) and outcomes related to transfer to clinical practice and patient care.
5. Geographic and specialty diversification: studies from Latin America, Africa, Asia and from underrepresented specialties (Psychiatry, Anesthesiology, Pathology, Radiology).
6. Focused systematic reviews in sub-domains where the evidence base is sufficiently dense (e.g., debiasing interventions for diagnostic reasoning; exam wrappers for resident learning; calibration training).

Strengths and limitations

Strengths:

- • Comprehensive search of five electronic databases (MEDLINE, Embase, ERIC, Web of Science, Cochrane CENTRAL).
- • Dual independent screening at title/abstract and full-text stages, with third-reviewer arbitration for conflicts.
- • Pre-piloted charting form (pilot with first 15 studies) refined through iterative team discussion.
- • Explicit dual classification by Flavell's metacognitive knowledge/regulation dimensions and by Zimmerman's SRL phases, enabling triangulation against established psychology theory.
- • Transparent reporting per PRISMA-ScR.

Limitations:

- • Restriction to English and Spanish publications; grey literature was not exhaustively searched.
- • Per JBI guidance for scoping reviews, no formal risk-of-bias appraisal of individual studies was performed; readers seeking effect-size estimates should consult focused systematic reviews on specific sub-domains.
- • Typology categorisation by metacognitive phase was made by consensus of the three authors based on stated intervention components; some studies could plausibly be reclassified.
- • A small number of studies (Tier 2) were included on the basis of multiple adjacent concepts; researchers focused strictly on "metacognition" as defined by Flavell may prefer to consider only Tier 1.

Conclusions

The empirical literature on metacognition-related educational interventions, measures and observational studies in postgraduate medical training is moderately extensive but conceptually fragmented. Our scoping review identified 92 studies spanning a broad typology of interventions and observational research, with strong concentration in Norteamerican Internal Medicine settings. The map identifies clear gaps to be addressed by future primary research and focused systematic reviews. Our typology by Zimmerman's SRL cycle phases provides a structured framework that may help organise future research efforts and educational programme design in this field.

Declarations

Ethics approval and consent to participate

Not applicable. This scoping review used publicly available published literature only and did not involve human participants directly. All research activities involving the retrieval, screening, extraction, and

analysis of published sources adhered to the ethical principles for research integrity outlined in the Singapore Statement on Research Integrity (2010) and, where applicable to the treatment of authorship and citation of prior work, to the general principles of the Declaration of Helsinki

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Funding

This scoping review received no specific external funding.

Author Contribution

All Four authors (AFY-V, DDB, LG, JLM-A) contributed jointly to study conceptualisation, protocol design, literature search, screening, data charting, analysis, manuscript drafting and final approval. All authors read and approved the final manuscript.

Availability of data and materials

All data generated and analysed during this scoping review are included as additional files with this manuscript: the PRISMA-ScR checklist (Additional file 1), the complete search strategies per database (Additional file 2), the full list of included studies with Vancouver-formatted references (Additional file 3), and the complete charted dataset of the 92 included studies (Additional file 4, Excel format). Further details are available from the corresponding author on reasonable request.

References

1. Bordage G. Elaborated knowledge: a key to successful diagnostic thinking. *Acad Med.* 1994;69(11):883–5.
2. Norman GR, Monteiro SD, Sherbino J, Ilgen JS, Schmidt HG, Mamede S. The Causes of Errors in Clinical Reasoning. *Acad Med.* 2017;92(1):23–30.
3. Schmidt HG, Mamede S. How to improve the teaching of clinical reasoning: a narrative review and a proposal. *Med Educ.* 2015;49(10):961–73.
4. Flavell JH. Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *Am Psychol.* 1979;34(10):906–11.

5. Eva KW, Regehr G. Self-assessment in the health professions: a reformulation and research agenda. *Acad Med.* 2005;80(10 Suppl):S46–54.
6. Hayes MM, Chatterjee S, Schwartzstein RM. Critical Thinking in Critical Care. *Ann Am Thorac Soc.* 2017;14(4):569–75.
7. Kahneman D, Thinking. Fast and Slow. New York: Farrar, Straus and Giroux; 2011.
8. Croskerry P. The importance of cognitive errors in diagnosis and strategies to minimize them. *Acad Med.* 2003;78(8):775–80.
9. Croskerry P, Singhal G, Mamede S. Cognitive debiasing 1: origins of bias and theory of debiasing. *BMJ Qual Saf.* 2013;22(Suppl 2):ii58–64.
10. Singh H, Meyer AND, Thomas EJ. The frequency of diagnostic errors in outpatient care. *BMJ Qual Saf.* 2014;23(9):727–31.
11. Newman-Toker DE, Schaffer AC, Yu-Moe CW, et al. Serious misdiagnosis-related harms in malpractice claims. *Diagnosis (Berl).* 2019;6(3):227–40.
12. Zimmerman BJ. Becoming a self-regulated learner: An overview. *Theory Pract.* 2002;41(2):64–70.
13. Brydges R, Manzone J, Shanks D, et al. Self-regulated learning in simulation-based training: a systematic review and meta-analysis. *Med Educ.* 2015;49(4):368–78.
14. Sandars J, Cleary TJ. Self-regulation theory: applications to medical education: AMEE Guide 58. *Med Teach.* 2011;33(11):875–86.
15. Cleary TJ, Durning SJ, Gruppen LD, Hemmer PA, Artino AR. Self-regulated learning in medical education. *Acad Med.* 2013;88(8):1090–100.
16. Lovett MC. Make exams worth more than the grade: Using exam wrappers to promote metacognition. In: Kaplan M, et al. editors. *Using Reflection and Metacognition to Improve Student Learning.* Sterling, VA: Stylus; 2013.
17. Mamede S, Schmidt HG, Penaforte JC. Effects of reflective practice on the accuracy of medical diagnoses. *Med Educ.* 2008;42(5):468–75.
18. Oettingen G. Future thought and behaviour change. *Eur Rev Soc Psychol.* 2012;23:1–63.
19. Cheng A, Eppich W, Grant V, Sherbino J, Zendejas B, Cook DA. Debriefing for technology-enhanced simulation: a systematic review and meta-analysis. *Med Educ.* 2014;48(7):657–66.
20. Sherbino J, Dore KL, Wood TJ, et al. The relationship between response time and diagnostic accuracy. *Acad Med.* 2012;87(6):785–91.
21. Stalmeijer RE. When I say... cognitive apprenticeship. *Med Educ.* 2015;49(4):355–6.
22. Veenman MV, Van Hout-Wolters BH, Afflerbach P. Metacognition and learning: conceptual and methodological considerations. *Metacognition Learn.* 2006;1(1):3–14.
23. Munn Z, Peters MDJ, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Med Res Methodol.* 2018;18(1):143.

24. Peters MDJ, Marnie C, Tricco AC, et al. Updated methodological guidance for the conduct of scoping reviews. *JBIM Evid Synth.* 2020;18(10):2119–26.
25. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. *Int J Soc Res Methodol.* 2005;8(1):19–32.
26. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. *Implement Sci.* 2010;5:69.
27. Tricco AC, Lillie E, Zarin W, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Ann Intern Med.* 2018;169(7):467–73.
28. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan—a web and mobile app for systematic reviews. *Syst Rev.* 2016;5:210.
29. Hoffmann TC, Glasziou PP, Boutron I, et al. Better reporting of interventions: TIDieR checklist and guide. *BMJ.* 2014;348:g1687.
30. Lambe KA, O'Reilly G, Kelly BD, Curristan S. Dual-process cognitive interventions to enhance diagnostic reasoning: a systematic review. *BMJ Qual Saf.* 2016;25(10):808–20.
31. Abernethy J, Shah A, Chen B, Reynolds S, Wright SM, et al. Integrating AI Scribes into Medical Education: Guardrails for Preserving Clinical Reasoning. *J Gen Intern Med.* 2026. 10.1007/s11606-025-10149-w.
32. Awad S, Turnnidge J, Cheung JJH, Taylor D, Dalgarno N, et al. Exploring perspectives of personal learning plans in a residency programme. *Clin Teach.* 2024. 10.1111/tct.13677.
33. Choi H, Oh J, Kim CK, Ryu H, Ryu Y. Residents need competence not confidence: A retrospective evaluation of the new competency education program for Korean neurology residents. *PLoS ONE.* 2023. 10.1371/journal.pone.0290503.
34. Crowley RS, Legowski E, Medvedeva O, Tseytlin E, Roh E, JOURNAL OF THE AMERICAN MEDICAL INFORMATICS ASSOCIATION. Evaluation of an intelligent tutoring system in pathology: Effects of external representation on performance gains, metacognition, and acceptance. 2007. doi:10.1197/jamia.M2241 WE - Science Citation Index Expanded (SCI-EXPANDED) WE - Social Science Citation Index (SSCI).
35. Ding W, Jiang L, Qian A, Xu S. The Metacognitive Enhancement Triad model enhances learning outcomes in first-year emergency medicine residents. *Front Med.* 2025. 10.3389/fmed.2025.1720071.
36. Farhat A, Gellin C, Dadiz R. Development of a longitudinal curriculum for a residency medical education track. *Acad Pediatr.* 2017. 10.1016/j.acap.2017.04.068.
37. Go MR, Traugott AL, Ejaz A, Collins C, Harzman AE, et al. Measuring Chief Resident Skill and Entrustment Progression in An Operative Coaching Program: Four Years' Experience. *J Surg Educ.* 2024. 10.1016/j.jsurg.2023.12.014.
38. Helm M, Ray Baroz A, Dhengre S, Rothrock L, Ackerman R. Diagnostic accuracy and metacognition in dermatology: A cross-sectional analysis of confidence and decision-making. *JAAD Int.* 2026. 10.1016/j.jdin.2026.01.001.

39. Mitchell R, Regan-Smith M, Fisher MA, Knox I, Lambert DR. A New Measure of the Cognitive, Metacognitive, and Experiential Aspects of Residents' Learning. *ACADEMIC MEDICINE*. 2009. doi:10.1097/ACM.0b013e3181a85609 WE - Science Citation Index Expanded (SCI-EXPANDED) WE - Conference Proceedings Citation Index - Science (CPCI-S).
40. Nishii A, Burdine J, Matusko N, Wang T, De Roo A, et al. Identification of promotion and prevention associated surgeon behaviors in the operating room to facilitate resident self-regulated learning. *Am J Surg*. 2021. 10.1016/j.amjsurg.2020.10.025.
41. Roumeliotis M, Warkentin H, Buckley L, Babic S, Becker N, et al. Competency-based simulation and evaluation in medical education: clinical decision making and reference dosimetry. *Int J Radiat Oncol Biol Phys*. 2026. 10.1016/j.ijrobp.2026.01.033.
42. Shelke SP, Tayade SA, Shelke YP, Shishodiya M, Alaspurkar SS, IMPACT OF FEEDBACK DURING OSCE AS A LEARNING, TOOL IN POSTGRADUATE STUDENTS OF OBSTETRICS AND GYNECOLOGY IN MAHATMA GANDHI INSTITUTE OF MEDICAL SCIENCES. *Int J Med Public Health*. 2025. 10.70034/ijmedph.2025.4.177. SEWAGRAM WARDHA.
43. Tallentire VR, Smith SE, Skinner J, Cameron HS. Understanding the behaviour of newly qualified doctors in acute care contexts. *MEDICAL EDUCATION*. 2011. 10.1111/j.1365-2923.2011.04024.x WE - Science Citation Index Expanded (SCI-EXPANDED).
44. Torti JMI, Murto SH, Bishop KA, Ahrari A, Goldszmidt M. In Support of Meaningful Assessment and Feedback: A Study of Clinical Reasoning Tasks Used in Ambulatory Case Reviews. *Perspectives on medical education*. 2026. 10.5334/pme.2294

Figures

Figure 1. PRISMA-ScR flow diagram

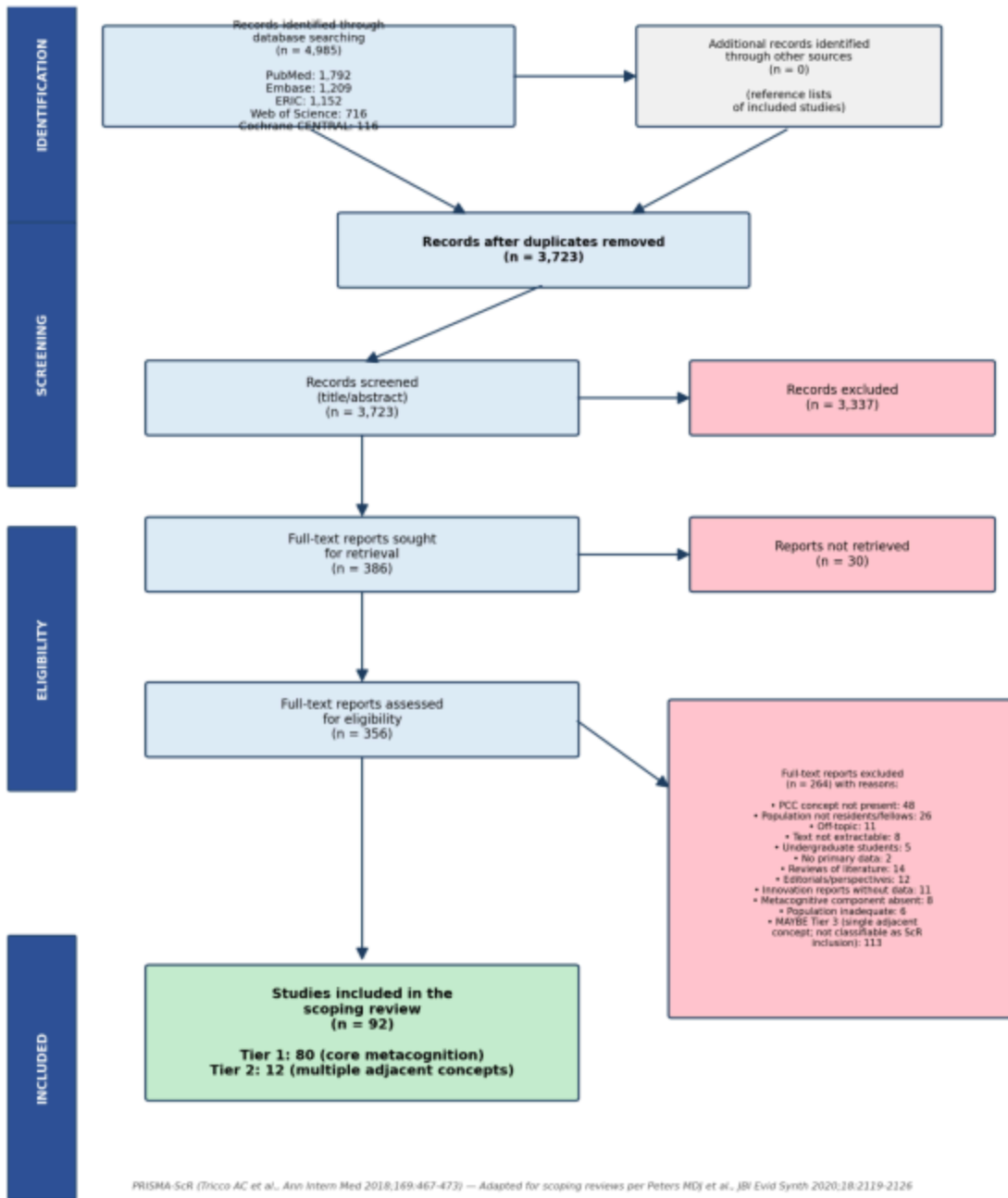


Figure 1

PRISMA-ScR flow diagram.

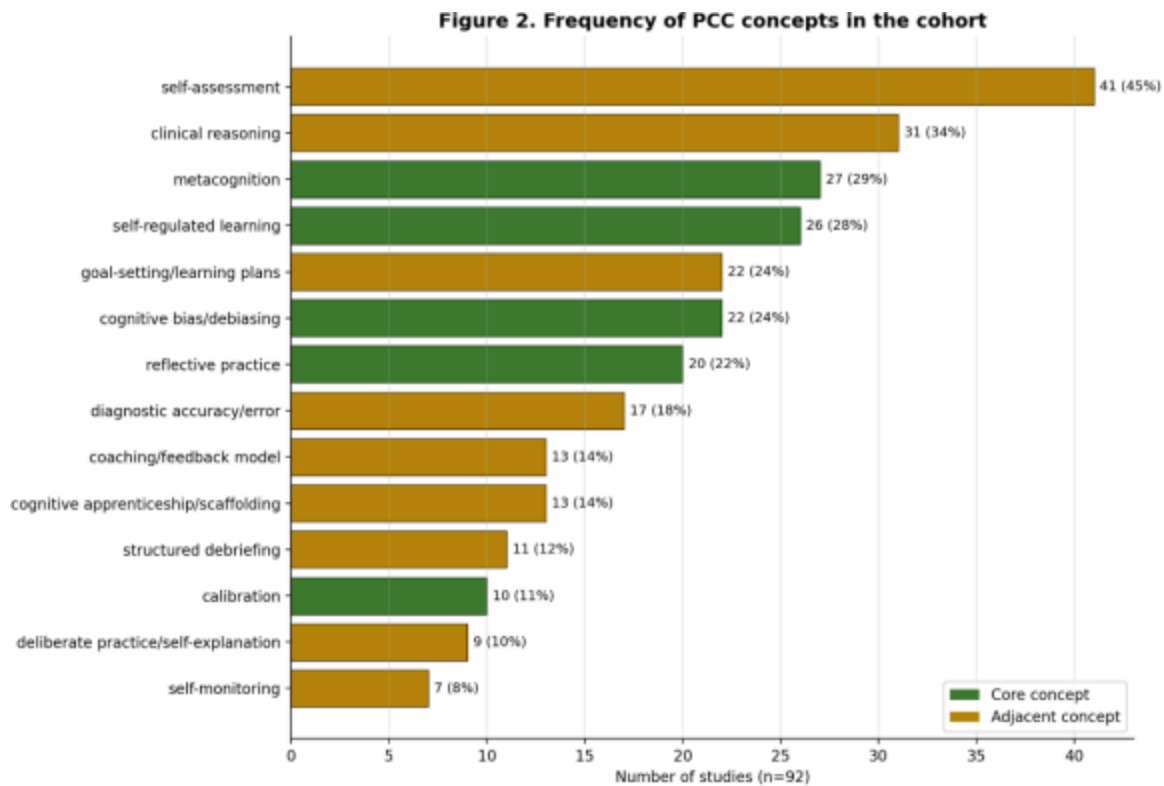


Figure 2

Frequency of PCC concepts in the cohort.

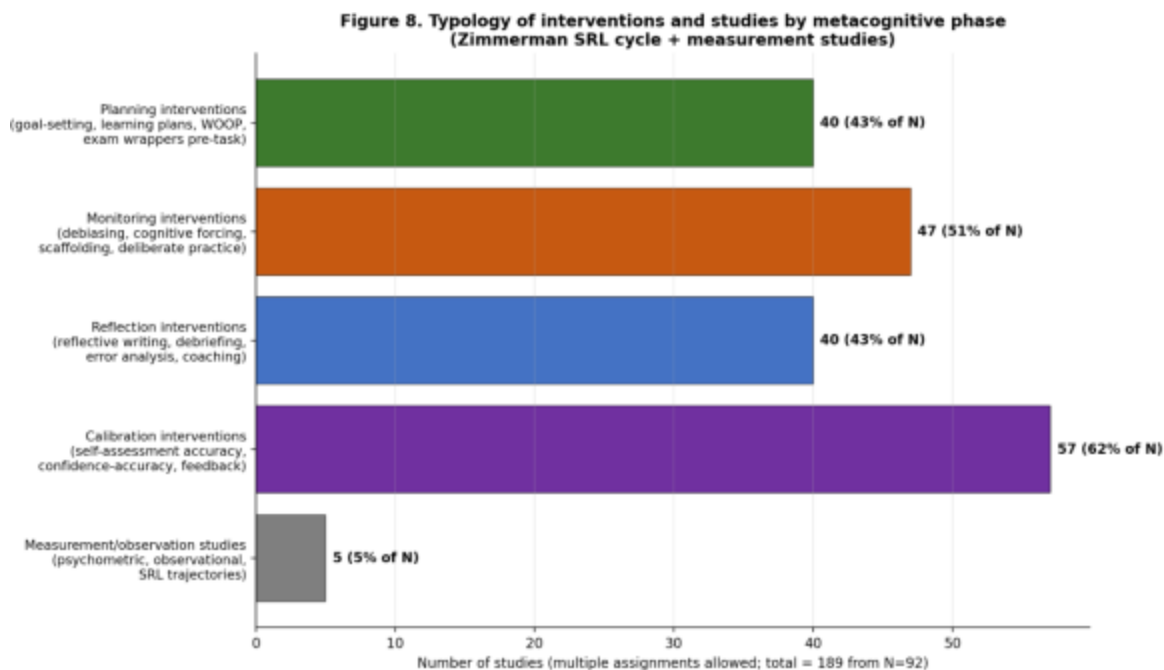


Figure 3

Typology of interventions and studies by metacognitive phase.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile5SupplementaryTablesandFigures.docx](#)
- [Additionalfile1PRISMA ScR Checklist.docx](#)
- [Additionalfile4ChartedData.xlsx](#)
- [Additionalfile2SearchStrategy.docx](#)
- [Additionalfile3IncludedStudiesReferences.docx](#)