frontiers in Education

TYPE Hypothesis and Theory
PUBLISHED 22 October 2026
DOI 10.3389/feduc.2026.1861062

Check for updates

OPEN ACCESS

EDITED BY
Shashidhar Venkatesh Murthy,
James Cook University, Australia

REVIEWED BY
Rita Payan Carreira,
University of Evora, Portugal
Caitlin Hanlon,
Quinnipiac University, United States
Ching-Chung Lin,
Mackay Medical College, Taiwan

*CORRESPONDENCE
Eulho Jung
eulho.jung@usuhs.edu

RECEIVED 20 April 2026
REVISED 05 September 2026
ACCEPTED 27 September 2026
PUBLISHED 22 October 2026

CITATION
Jung E, Reigeluth CM, Zheng B, Durning SJ and Merkebu J (2026) A five-function theory-informed instructional design model for developing clinical reasoning.
Front. Educ. 11:1861062.
doi: 10.3389/feduc.2026.1861062

COPYRIGHT
© 2026 Jung, Reigeluth, Zheng, Durning and Merkebu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

A five-function theory-informed instructional design model for developing clinical reasoning

Eulho Jung[1]*, Charles M. Reigeluth[2], Binbin Zheng[1], Steven J. Durning[1] and Jerusalem Merkebu[1]

[1]Department of Health Professions Education, School of Medicine, Uniformed Services University of the Health Sciences, Bethesda, MD, United States, [2]Department of Instructional Systems Technology, School of Education, Indiana University Bloomington, Bloomington, IN, United States

Clinical reasoning is an iterative, context-sensitive phenomenon. Its instruction and assessment require distinct forms of educational design. Existing reviews, consensus guidance, and longitudinal curricula provide substantial content and teaching strategies. Further specification is needed to coordinate conceptual preparation, expert modeling, scaffolded whole-case practice, feedback, adaptive fading, and transfer across learner development. We present CRAFT, a theory-informed instructional design model organized around five functions: Conceptualize and Frame, Reveal and Model Clinical Reasoning, Apply Through Scaffolded Practice, Feedback and Reflect, and Transfer Toward Independence. CRAFT provides a curricular architecture organized around proposed readiness gates, in which the five functions recur in progressively more demanding cycles as knowledge, contextual variation, uncertainty, learner responsibility, and assessment stakes increase while instructional support decreases. These gates are design criteria for local testing, not validated progression thresholds. We compare CRAFT directly with established instructional approaches and longitudinal clinical-reasoning curricula, define function-specific transition criteria, and formulate five testable propositions linking model components to mechanisms, proximal outcomes, clinical-reasoning performance, and transfer. A multilevel sepsis example illustrates how the same case family could be adapted for learners at different stages. No learner cohort data are reported; the example is an instructional design demonstration, and the evaluation plan is prospective. The model's feasibility, mechanisms, incremental value, and boundary conditions require empirical testing.

KEYWORDS
clinical reasoning, curriculum, instructional design, medical education, scaffolding, transfer

1 Introduction

Clinical reasoning is the iterative, context-sensitive work through which clinicians gather and interpret information, construct problem representations, generate and prioritize diagnostic and management possibilities, act, and revise decisions as new evidence emerges (Bowen, 2006; Schmidt and Mamede, 2015). Its processes vary across clinical contexts and tasks. Performance depends on organized knowledge, the features of the case, prior experience, uncertainty, team and system conditions, and the learner's capacity to monitor and adapt reasoning.

Substantial guidance addresses both the content and methods of clinical-reasoning instruction. Case-based coaching makes expert reasoning accessible for discussion (Kassirer, 2010); whole-case and knowledge-oriented approaches emphasize knowledge organization and stage-appropriate practice (Schmidt and Mamede, 2015); the UK Clinical Reasoning in Medical Education consensus identifies curricular content and teaching strategies (Cooper et al., 2021); and several institutions have implemented longitudinal or spiral curricula (Connor et al., 2022; Duca and Glod, 2019; Rowat and Suneja, 2022; Singh et al., 2022). The remaining challenge is to integrate and operationalize this guidance across the full course of learner development.

Available evidence indicates a gap in implementation and curricular coordination. In a national survey of internal medicine clerkship directors, 57% of responding institutions lacked sessions dedicated to clinical-reasoning concepts; the most frequently reported barriers were limited curricular time (88%) and faculty expertise (69%) (Rencic et al., 2017). A recent scoping review found wide variation in curricular theories, teaching methods, and assessments and no all-encompassing design across the included health-professions curricula (Elvén et al., 2023). A 2025 systematic review and meta-analysis found that explicit instruction was generally beneficial, especially during clinical years, while also identifying a shortage of multiyear studies that deliberately scaffold the same cohort across a full program (Wark et al., 2025). These findings support further design guidance that builds on existing approaches to sequencing, scaffolding, and longitudinal development.

The coordination problem can be located at three related levels. At the activity level, educators may use effective cases, think-alouds, worksheets, debriefs, and workplace observations as isolated events. At the course level, those activities may be sequenced without an explicit account of which reasoning demands are being increased, which supports are being withdrawn, or what evidence would justify that change. At the program level, learners may encounter different terminology, expectations, and assessment practices across courses and clinical sites. Across these levels, the developmental rationale for instruction may remain partly implicit. CRAFT is intended to make coordination decisions explicit and open to review within existing activities and curricula.

This coordination focus defines the criteria for evaluating CRAFT. Demonstrating that learners improve after repeated cases would not establish CRAFT's specific contribution, because practice, feedback, clinical exposure, and maturation are plausible explanations. The proposed contribution is the explicit coordination of instructional functions, readiness gates, developmental recursion, and transition criteria. The relevant test is whether this architecture helps educators make more coherent sequencing, support, transition, and transfer decisions than a credible alternative. Evaluation should first establish whether CRAFT improves design decisions and coordination, then examine effects on learner outcomes.

CRAFT is proposed as a theory-informed instructional design model. Here, a model is a representation of design functions and hypothesized relations that can guide curriculum decisions. Its scope is instructional design; theories of clinical reasoning and assessment frameworks address related, distinct questions. Its proposed contribution is the explicit coordination of established approaches within an integrative architecture: five recurring instructional functions defined by the locus of cognitive work; readiness gates that align support and progression with performance evidence; developmental recursion as complexity, contextual variation, uncertainty, and learner responsibility increase; and transition criteria and boundary conditions that can be tested against existing instruction.

Accordingly, this Hypothesis and Theory article specifies the proposed architecture, mechanisms, comparisons, and testable claims for critical appraisal and future study. It reports no learner cohort results; the worked example demonstrates the design process.

The intended unit of analysis is a designed sequence of reasoning experiences across instructional encounters. A brief session may incorporate several functions. The model guides progression when evidence from one encounter informs the support, case demand, or responsibility assigned next. CRAFT's scope is limited to reasoning-focused instruction. Foundational science instruction, communication, procedural skills, professional formation, and patient-safety systems retain their own aims and design requirements. The model addresses how reasoning-focused experiences are coordinated within and across those structures. This bounded scope is necessary both for feasible implementation and for interpretable evaluation.

2 Clinical reasoning, instruction, and assessment as distinct levels

Clinical reasoning as a phenomenon refers to what a learner or clinician does when interpreting and acting on a patient problem. It includes knowledge use, problem representation, hypothesis revision, diagnostic and management decisions, metacognitive monitoring, and adaptation to context (Charlin et al., 2000; Konopasky et al., 2020).

Teaching clinical reasoning is a separate instructional design task. It concerns the selection and sequencing of cases, the visibility of expert performance, the distribution and withdrawal of support, the timing of feedback, and opportunities for increasingly independent performance. CRAFT is located at this instructional level.

Assessing clinical reasoning is an evidence and inference task. Assessment samples performance so educators can judge learning, guide feedback, make progression decisions, or evaluate a curriculum (Daniel et al., 2019). Assessment evidence may inform readiness gates and evaluation of CRAFT. A score alone provides neither an instructional sequence nor evidence that the model caused learning. Accordingly, learner outcomes concern acquisition and transfer of clinical-reasoning competencies; implementation outcomes concern feasibility, acceptability, fidelity, resources, and equity.

Clear distinctions among these interacting levels are necessary for interpreting evidence. A learner may reason well on one case yet lack stable competence across cases; an assessment may identify a weak problem representation without specifying how to teach it; and a carefully implemented function may fail because the cases do not sample the intended construct. An instructional activity can also be feasible and acceptable without

improving reasoning. CRAFT therefore treats assessment evidence as an input to design decisions and an outcome of evaluation while preserving the separate warrants required for claims about reasoning, instruction, and measurement.

For curriculum design, this distinction determines what counts as actionable evidence. Concept inventories and explanations can inform entry into conceptual work; annotations of expert demonstrations can show whether consequential cues are being noticed; committed case responses can reveal how learners perform with support; feedback uptake can be examined in a subsequent case; and reduced-support performance can inform transfer and supervision decisions. None of these observations should be interpreted outside its sampling conditions. The instructional question is how evidence changes the next design decision, whereas the assessment question is whether the evidence supports the intended inference.

3 Theoretical foundations

CRAFT uses elaboration theory and the Four-Component Instructional Design model (4C/ID) as primary prescriptive resources because they address the sequence and support of complex whole tasks (Reigeluth, 1999; Van Merriënboer et al., 2003). Elaboration theory contributes an epitome that preserves the structure of a task while simplifying its conditions. 4C/ID contributes whole-task classes, supportive information, just-in-time procedural information, varied practice, and fading. Situated cognition adds purposeful variation in patient, team, resource, emotional, and workflow conditions; cognitive apprenticeship adds modeling, coaching, articulation, and fading; self-regulated learning and deliberate practice add goal setting, monitoring, repeated performance, feedback, and revision (Collins et al., 1991; Ericsson, 2004; Zimmerman, 2002).

These resources were selected for their relevance to specific instructional design decisions. Elaboration theory and 4C/ID address curriculum-scale decisions about where to begin, which dimensions of a whole task to simplify, how to group tasks into increasingly demanding classes, and how support should change across those classes. Cognitive apprenticeship and situated cognition then specify how performance becomes visible and how context is introduced. Self-regulated learning and deliberate practice specify how learners use criteria, monitoring, feedback, and repeated attempts. CRAFT assigns each tradition a specific design function and identifies tensions among their prescriptions.

Integrating these traditions requires attention to competing instructional demands. Authenticity and contextual variation can raise element interactivity beyond a novice's capacity (Sweller, 1988); extensive scaffolding can conflict with learner independence (Van de Pol et al., 2010); reflection prompts can support monitoring but also create avoidable processing demands (Mann et al., 2009); and assessment can guide learning while high-stakes use can narrow it. CRAFT therefore applies four precedence rules. First, safety and prerequisite knowledge take precedence over realism. Second, evidence of performance takes precedence over a calendar-based schedule for fading. Third, task-relevant complexity takes precedence over realism for its own sake. Fourth, human oversight takes precedence over automated instructional convenience.

These precedence rules guide specific design decisions. If a realistic case contains so many interacting data that a novice cannot construct an interpretable representation, the designer should simplify the case while preserving its clinical structure. If a learner completes the scheduled number of supported cases but still omits a safety-critical action, the support should not be withdrawn merely to preserve pacing. If a reflection prompt produces lengthy narration without a change in subsequent performance, it should be shortened, retargeted, or removed. If an automated tutor offers rapid feedback but cannot explain the provenance of its reference path, human review should supersede convenience. The resulting decisions provide observable targets for evaluating the synthesis.

CRAFT's treatment of cognitive load follows current distinctions among task complexity, extraneous demands, and learning-relevant processing. Intrinsic load varies with element interactivity and learner expertise, while extraneous load arises from avoidable demands created by the design (Sweller, 1988). Germane processing refers to working-memory resources directed toward learning-relevant aspects of the task; it is not an independent load that should simply be increased. Instruction should therefore reduce extraneous demands, calibrate intrinsic complexity to prior knowledge, and direct available resources toward schema construction and automation within capacity limits (Sweller et al., 2019). The same support can reverse from helpful to redundant as expertise grows (Kalyuga, 2007), making learner knowledge and performance central to fading decisions.

Readiness varies across tasks and domains. A learner may require substantial support for an unfamiliar endocrine presentation while functioning independently in a well-practiced cardiopulmonary domain. Readiness evidence should therefore be interpreted locally, across more than one observation, and in relation to the next proposed increase in difficulty. The expertise-reversal effect further indicates that a scaffold's value depends on learner expertise: it may reduce an otherwise unmanageable demand or compete with an already organized routine. CRAFT uses this conditional view to connect theory selection with transition decisions.

Table 1 sets out each theoretical resource, the design question it answers, the tension or limit it carries, and CRAFT's proposed response to that tension.

The synthesis draws on these theories to preserve the whole reasoning task, make consequential cognition visible, allocate meaningful work to the learner, adjust support to observed performance, and sample adaptation under changing conditions. No single source supplies all five commitments, and their combination does not prove that the resulting model is effective. The synthesis provides an explicit rationale for why each function is present, what mechanism it is expected to activate, and where the design could fail. Table 1 links the theoretical resources to design choices that can be evaluated empirically.

Designers can apply the synthesis through three questions. What feature of the next task is expected to become more demanding? What temporary support makes that demand interpretable without performing the reasoning for the learner? What observable pattern would justify withdrawing or changing the support? These questions connect whole-task sequencing, cognitive load, apprenticeship, and self-regulation without assuming that one theory resolves the others. They also guide

TABLE 1 Theoretical resources, tensions, and their resolution in CRAFT.

Theoretical resource	Design question	Principal contribution	Tension or limit	CRAFT resolution
Elaboration theory (Reigeluth, 1999)	How should whole tasks be sequenced?	Begin with an epitome that preserves task structure; add complexity by changing simplifying conditions.	Over-simplification can remove authentic relations.	Preserve the whole reasoning task while limiting interacting elements.
4C/ID (Van Merriënboer et al., 2003)	How should whole-task practice and support be organized?	Task classes, supportive information, just-in-time procedural guidance, varied practice, fading.	Generic model; clinical-reasoning constructs and readiness evidence require specification.	Supplies the primary whole-task and fading architecture.
Cognitive apprenticeship (Collins et al., 1991)	How can tacit performance become learnable?	Modeling, coaching, articulation, reflection, exploration, gradual withdrawal.	Modeling may promote passive imitation; coaching can persist too long.	Model dynamically, require learner explanation, then fade on evidence.
Situated cognition (Konopasky et al., 2020)	How should context enter instruction?	Vary patient, team, resource, emotional, and workflow conditions.	Authenticity can overload novices and reduce interpretability.	Introduce contextual variation in calibrated layers across cycles.
Cognitive load theory (Sweller, 1988; Sweller et al., 2019)	How should complexity and information be calibrated?	Reduce extraneous demands; match intrinsic complexity to expertise; direct resources to learning-relevant processing.	Support can become redundant as expertise grows.	Use readiness and performance to select complexity and withdraw support.
Self-regulated learning and deliberate practice (Ericsson, 2004; Zimmerman, 2002)	How do learners monitor and improve?	Goals, repeated performance, feedback, reflection, revision, planning.	Reflection may be ritualized; self-regulation cannot be assumed in novices.	Use brief criterion-linked reflection and verify change in later performance.

adaptation: a change in technology or delivery format is relevant only insofar as it changes the task, support, learner work, or evidence. Theoretical principles thus inform the selection and adaptation of instructional activities.

4 Positioning CRAFT among existing approaches

Table 2 positions CRAFT against the approaches most relevant to its claimed contribution and documents substantial conceptual overlap. Established models and curricula already address case sequencing, coaching, longitudinal integration, and assessment. CRAFT's proposed contribution lies in explicitly coordinating these elements through five recurring instructional functions, proposed readiness gates, developmental cycle parameters, and testable component-to-outcome propositions. Whether this integration offers incremental value remains an empirical question.

The approaches can also be distinguished by the decision level at which they are most informative. Cognitive apprenticeship offers a strong account of guided participation within an activity; 4C/ID offers a broad architecture for complex whole-task learning; case-based coaching and whole-case proposals specify clinically meaningful teaching encounters; consensus statements identify curricular content and strategies; and published longitudinal themes show how programs have coordinated reasoning instruction locally. CRAFT draws from all of these levels. Its five functions are best understood as a common planning language that connects an individual encounter to a course sequence and then to a program-level progression.

The comparison clarifies the scope of the implementation gap. Cognitive apprenticeship and 4C/ID are comprehensive generic instructional designs; clinical-reasoning curricula at Manchester, Iowa, UCSF, and other institutions demonstrate longitudinal implementation; and consensus and review papers provide detailed content and strategy guidance. What remains unevenly specified across these sources is a compact decision architecture that identifies the dominant instructional function, the locus of cognitive work, the evidence needed to fade support, and the way the same functions change across successive levels of expertise. Whether CRAFT improves decisions beyond careful use of these sources is an empirical question.

Theoretical fidelity to CRAFT depends on how instructional activities and progression decisions are coordinated. A curriculum that integrates whole-case reasoning, cognitive modeling, contingent coaching, feedback, and transfer can align with the model's commitments regardless of terminology. Assigning CRAFT labels to disconnected activities does not establish that the hypothesized mechanisms are operating. The critical unit of analysis is the relationship among task complexity, learner agency, scaffolding, assessment evidence, and responsive sequencing. Readiness gates and developmental cycles make this theoretical logic explicit and auditable.

This overlap guides comparator selection. One testable specification concerns the locus of cognitive work and evidence for transitions among case-based activities; another concerns how recurring functions and independently variable cycle parameters are coordinated across a curriculum. Both are design

TABLE 2 Comparison with relevant instructional approaches and clinical-reasoning curricula.

Approach	Primary purpose or unit	Established strength	Less explicit or locally variable	CRAFT relationship
Cognitive apprenticeship (Collins et al., 1991)	Generic learning through guided participation	Makes expert performance visible and links modeling, coaching, articulation, reflection, and fading.	Does not itself specify a clinical-reasoning curriculum map or readiness measures.	CRAFT assigns these activities to recurring functions and ties fading to cross-case evidence.
4C/ID (Van Merriënboer et al., 2003)	Generic design of complex learning and whole tasks	Provides task classes, supportive/procedural information, varied practice, and systematic fading across courses.	Requires domain-specific task classes, constructs, and criteria.	CRAFT applies whole-task logic to clinical reasoning and adds CR-specific functions and progression evidence.
Case-based and coached reasoning (Kassirer, 2010)	Case-level reasoning instruction	Emphasizes repeated authentic cases and coaching across diagnostic and therapeutic decisions.	Less explicit about cross-course governance, developmental cycles, and assessment architecture.	CRAFT retains cases and coaching while specifying curriculum-level coordination and transitions.
Whole-case developmental proposal (Schmidt and Mamede, 2015)	Stage-sensitive selection of CR teaching approaches	Distinguishes serial-cue/whole-case and knowledge-/process-oriented designs; emphasizes knowledge and learner stage.	Does not provide a full implementation and governance architecture.	CRAFT adopts stage sensitivity and whole cases, then adds recurring functions and transition criteria.
CReME consensus (Cooper et al., 2021)	What, how, and when to teach CR in undergraduate medicine	Defines content domains and supports horizontal and vertical integration.	Consensus guidance permits varied local sequencing and fading decisions.	CRAFT proposes a sequence for coordinating these decisions while retaining the consensus content guidance.
Longitudinal course examples (Connor et al., 2022; Duca and Glod, 2019)	Implemented course/thread within local programs	Demonstrate repeated cases, explicit CR language, worksheets, scaffolding, and systems/team content.	Local designs and outcomes do not provide a general comparative model or common transition rules.	CRAFT offers a cross-context hypothesis about how local activities could be coordinated and tested.
Spiral curriculum themes (Rowat and Suneja, 2022; Singh et al., 2022)	Program-level longitudinal integration	Demonstrate progressive themes, curriculum mapping, faculty development, assessment, and feasibility.	The published implementations do not use CRAFT's five-function readiness gates or propositions.	CRAFT overlaps substantially; its possible added value is explicit function-level and readiness logic, which remains untested.

claims. Table 2 supports interpretation of null findings, including the possibility that CRAFT offers no incremental benefit.

5 CRAFT as a stage-adaptive instructional design model

The five CRAFT functions are Conceptualize and Frame; Reveal and Model Clinical Reasoning; Apply Through Scaffolded Practice; Feedback and Reflect; and Transfer Toward Independence. Support is the broad category of designed assistance. Scaffolding is temporary, contingent support that enables performance not yet possible independently (Van de Pol et al., 2010; Vygotsky, 1978). Procedural guidance is just-in-time information for recurrent actions (Van Merriënboer et al., 2003). Coaching is interactive observation, questioning, cueing, and adjustment during performance (Collins et al., 1991). Feedback is information comparing observed performance with criteria or a reference and may occur during or after performance (Lefroy et al., 2015). The functions are distinguished by their dominant purpose and by who performs the principal cognitive work, not by a rigid chronology. Feedback may occur during application, and reflection may be embedded within practice; the label identifies the function foregrounded at a given moment.

Figure 1 shows how readiness evidence guides the five functions and how they recur across developmental cycles.

In this article, a readiness gate is a proposed design checkpoint for deciding whether to retain, change, or withdraw support and whether to increase a specified task demand. The criteria in Table 3 are not validated thresholds or established assessment standards. Programs must define and validate local decision rules, match evidence requirements to the stakes, and account for placements where observations are sparse or course schedules limit recursive support.

5.1 Conceptualize and frame

The educator establishes an interpretable task, prerequisite knowledge, performance criteria, and a simple knowledge structure before asking for full reasoning (Charlin et al., 2000; Gobet, 2005). Early cases should be epitomes: authentic enough to preserve the whole task but simplified in data volume, uncertainty, and contextual disruption (Reigeluth, 1999). The learner organizes relevant biomedical and clinical knowledge, constructs an initial problem representation, and connects it to

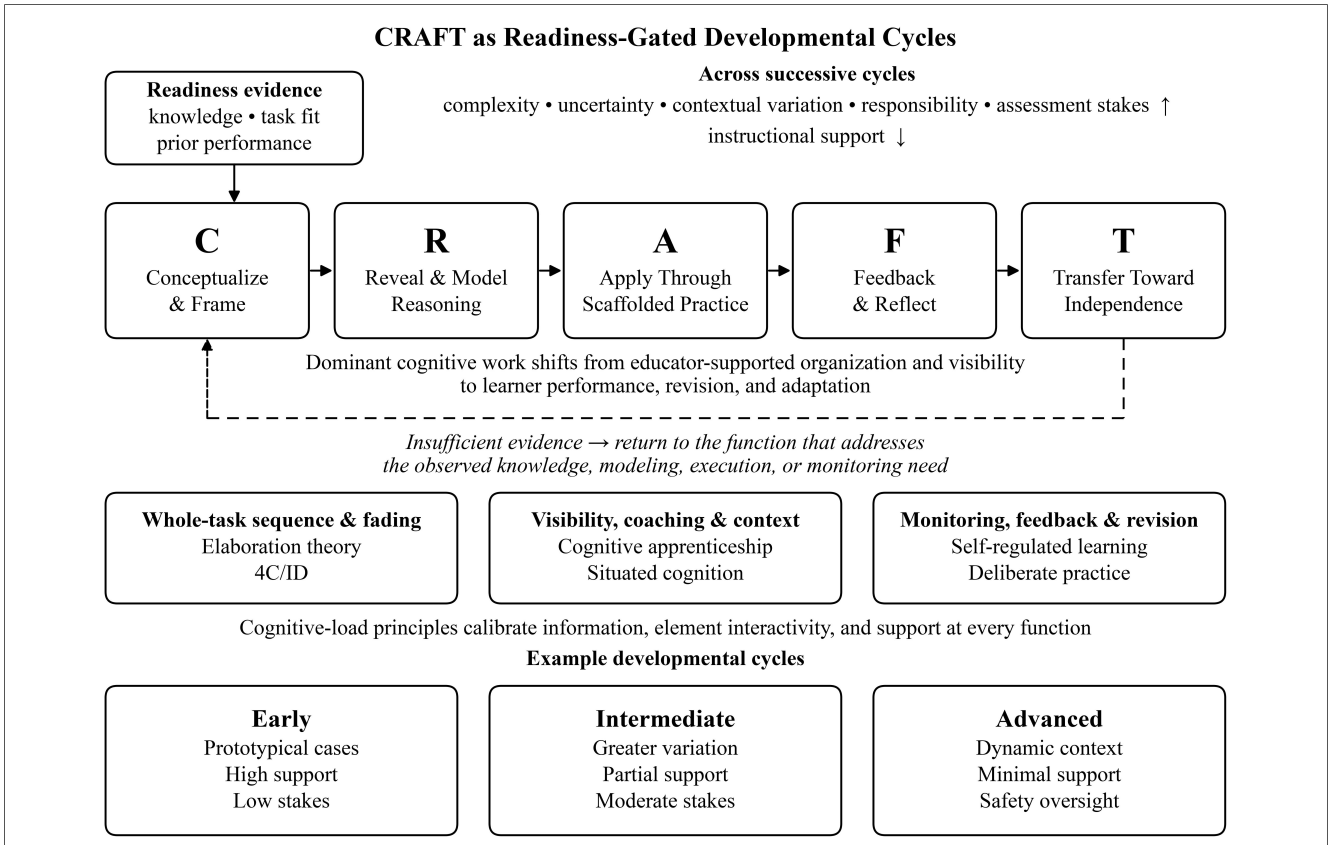


FIGURE 1
CRAFT as readiness-gated developmental cycles. The five functions recur within and across developmental cycles and do not map onto specific curricular years. Entry into a cycle requires foundational knowledge and an interpretable task; progression to a more demanding cycle requires consistent performance across varied cases and occasions. Across successive cycles, case complexity, uncertainty, contextual variation, learner responsibility, and assessment stakes increase while instructional support decreases. The dashed path indicates recursion: when evidence is insufficient, the learner returns to the function that addresses the observed knowledge, modeling, execution, or monitoring need.

TABLE 3 Operational differentiation and transition criteria for the five CRAFT functions.

Function	Dominant purpose	Learner responsibility	Instructor support	Mechanism	Evidence for progression
C - Conceptualize and Frame	Knowledge architecture and interpretable task	Organize knowledge; construct a concise representation	High; schemas, epitome case, criteria, prior-knowledge prompts	Schema construction and activation of relevant knowledge	Adequate prerequisite explanation and coherent representations on multiple low-complexity cases
R - Reveal and Model Clinical Reasoning	Visibility of dynamic expert performance	Explain cue significance; compare and anticipate revisions	High to moderate; staged think-alouds and contrastive examples	Observational learning, explanation, discrimination of consequential features	Explains probability/priority shifts and anticipates selected revisions across examples
A - Apply Through Scaffolded Practice	Learner performance with contingent support	Commit to representation, differential, requests, and plan	Moderate and adaptive; coaching, completion tasks, temporary tools	Whole-task practice, retrieval, feedback preparation, automation	Consistent performance across varied cases with fewer prompts and no safety-critical omissions
F - Feedback and Reflect	Error analysis and planned revision	Interpret feedback; identify error pattern; plan next action	Moderate to low; rubric, debrief, reference path, targeted prompts	Calibration, metacognitive monitoring, error correction	Planned adjustment is evident in subsequent cases; feedback dependence declines
T - Transfer Toward Independence	Generalization and readiness under reduced support	Lead reasoning, adapt to context, seek help appropriately	Low but safety-preserving; oversight and access to trusted resources	Variable practice, transfer, adaptive expertise, entrustment	Triangulated consistency across cases, occasions, contexts, and assessors, including delayed transfer

prior knowledge; biomedical knowledge remains consequential for later performance on difficult cases (Woods, 2007; Woods et al., 2007). Progression requires the learner to demonstrate an adequate foundational explanation, recognize the task's major features, and produce a coherent representation on more than one low-complexity case. Framing should specify the clinical decision to be made, the time horizon, the expected level of explanation, and the resources a learner may use. It should also surface misconceptions early enough to change the case or support. A concept map is useful only if it helps the learner compress and retrieve relations that matter for the whole task; it should not become an isolated recall exercise. When several learners reach the same incorrect representation, educators should review the prerequisite instruction and case wording before assigning further practice. Transition decisions therefore depend on evidence of learner readiness and the interpretability of the task.

5.2 Reveal and model clinical reasoning

An experienced clinician or calibrated near-peer makes dynamic reasoning visible as information changes (Collins et al., 1991; Jagannath et al., 2023). Think-aloud demonstrations should explain how problem representations, probability estimates, uncertainty, management priorities, and help-seeking inform the final decision (Jagannath et al., 2023). Contrastive examples can show how premature closure, weak weighting of evidence, or context changes the path (Eva and Norman, 2005). The learner's responsibility is to annotate and explain why the representation or plan changed; worked examples paired with prompted self-explanation and interleaved problem solving support this transition (Chamberland et al., 2015; Stark et al., 2011; Van Gog et al., 2011). Progression is appropriate when the learner can identify consequential cues, compare reasoning paths, and anticipate at least some revisions before they are revealed. The model should be selective: narrating every thought can overload attention and create the impression that expertise is exhaustive verbalization. Educators should foreground the moments at which evidence changes probability, urgency, or management and should identify which knowledge structures support those shifts. Near-peers may make recently learned strategies accessible, but their explanations require calibration to the same criteria and safety standards. Learners should alternate observation with prediction, explanation, or partial completion to prepare for subsequent performance. Repeated inability to anticipate or explain a consequential revision signals a need for further conceptual work or a different contrastive example.

5.3 Apply through scaffolded practice

The learner now performs the whole reasoning task and commits to a representation, prioritized differential, information requests, and plan before receiving feedback. Cases are sequenced by purposeful changes in diagnosis, uncertainty, comorbidity, information quality, and context. Temporary scaffolds may include completion problems, worksheets, cueing questions, and coaching (Braun et al., 2017; Levin et al., 2016; Nelson and Rhoney, 2024; Theobald and Ramsbotham, 2019). Their intensity should be adjusted to error pattern and prior knowledge, because the same scaffold can help a novice and hinder a more knowledgeable learner (Braun et al., 2019; Kiesewetter et al., 2020). Progression requires consistent performance across varied cases, decreasing reliance on prompts, and the absence of safety-critical omissions. Completing a fixed number of exercises is not sufficient. Commitment is important because feedback delivered before a learner externalizes a representation or plan can conceal the reasoning that requires support. A concise committed response should show how evidence is weighted and what action follows. Scaffolds should target the observed bottleneck: a data organizer for information management, a comparison prompt for weak discrimination, a partial representation for knowledge organization, or coaching for execution under pressure. Adding every available scaffold at once obscures which support matters and increases extraneous demand. Fading can involve removing prompts, delaying hints, broadening case variation, or shifting from instructor to learner control. A single adjustment may be sufficient to assess readiness for greater independence.

5.4 Feedback and reflect

The dominant function shifts from task performance to analysis and revision. Faculty, peers, or locally validated digital tools compare reasoning with explicit criteria and reference paths (Lefroy et al., 2015; Ruczynski et al., 2025). Learners explain how their reasoning changed, distinguish knowledge gaps from process or contextual failures, and state a specific adjustment for the next case (Mamede and Schmidt, 2023). Reflection should be brief, criterion-referenced, and connected to subsequent action (Choi et al., 2020; Mann et al., 2009). Progression requires evidence that learners understand the feedback and apply the planned adjustment in subsequent performance. Effective feedback specifies an action for the next attempt. A learner who identifies premature closure should specify which evidence threshold or comparison strategy will be used on the next case; a learner with a knowledge gap should identify what must be reorganized before reattempting. Faculty should distinguish errors that require immediate correction from productive differences in a defensible reasoning path. Peer feedback can broaden exposure to alternatives when peers use common criteria and the educator retains responsibility for disputed or safety-critical points. The next case serves as the test of uptake. If the same pattern persists, targeted modeling or supported practice should address the unresolved difficulty.

5.5 Transfer toward independence

Learners reason with substantially reduced support across unfamiliar, variable, or workplace-like cases while supervisors preserve patient safety. Complexity may involve atypical presentation, competing diagnoses, multiple problems, time pressure, interrupted information, resource constraints, and team coordination (Konopasky et al., 2020). Independence

includes appropriate escalation, deliberate slowing when a situation demands it, and use of trusted resources under appropriate supervision (Moulton et al., 2007; Ten Cate, 2013). Progression is justified by triangulated evidence across cases, occasions, contexts, and assessors, with particular attention to critical omissions, calibration, consistency, and delayed transfer. Transfer tasks should vary along specified, educationally relevant dimensions.

Educators might preserve the underlying pathophysiology while changing presentation, setting, resource constraints, or competing priorities, allowing interpretation of what generalized and what did not. Ordinary clinical resources should remain available as instructional support decreases; independent practice includes knowing when and how to consult references, teammates, and supervisors. The evidence should distinguish efficient adaptation from reckless speed and appropriate help-seeking from avoidable dependence. Failure in one contextual dimension may justify a targeted return to an earlier function without assuming that the learner has lost all prior competence.

Table 3 differentiates the five functions by dominant purpose, learner responsibility, instructor support, hypothesized mechanism, and the evidence required before a learner progresses.

Coherence across functions depends on using evidence from each function to guide the next design decision. The representation developed during conceptual framing should organize what the learner notices during modeling; the modeled comparison should become a demand during supported practice; the learner's committed response should provide evidence for feedback; and the feedback plan should be observable in a subsequent, increasingly dissimilar case. When these links are absent, the same activity can lose its intended function. A worksheet completed after the answer is known does not prepare feedback in the same way as a committed response, and a transfer case preceded by stepwise prompts does not sample reduced-support adaptation. The function labels identify dominant design purposes and may overlap or recur within the same instructional encounter.

6 Developmental logic of repeated CRAFT cycles

The timing and fidelity of an initial CRAFT cycle should reflect learner readiness. Before entering a clinical-reasoning cycle, learners need enough foundational knowledge and patient-language familiarity to interpret a simplified problem meaningfully. Programs may prepare these prerequisites through basic science, history and examination skills, simple patient narratives, and vocabulary work as preparation for later encounters with clinical uncertainty.

Prerequisite preparation should permit meaningful early contact with patients. The entry threshold is the learner's capacity to interpret a simplified task sufficiently to organize relevant knowledge. Early patient narratives, basic data interpretation, and guided explanations may contribute to that preparation. Because readiness is domain-specific, a learner can enter an early CRAFT cycle in one content area while continuing prerequisite work in another. Programs should document the evidence used for entry and revisit the threshold if it systematically delays participation or creates inequitable access.

Once the threshold is met, the five functions recur in developmental cycles consistent with a spiral curriculum (Harden, 1999). An early cycle uses common, prototypical, low-uncertainty cases with high support and low stakes. An intermediate cycle adds atypical features, comorbidity, incomplete data, and moderate contextual variability while reducing prompts. An advanced cycle uses dynamic multi-problem cases, competing demands, broader contextual variation, and workplace responsibility with supervision for safety. Programs can place these cycles within their own schedules while retaining readiness-based progression.

The dimensions of development should not all be increased at once. A subsequent cycle might retain a familiar diagnosis while adding competing priorities, or keep the clinical pattern prototypical while introducing interruptions, team communication, or limited resources. Another cycle might increase diagnostic ambiguity but preserve a worksheet so that the effect of ambiguity remains interpretable. Separating complexity, uncertainty, contextual variation, responsibility, support, and stakes allows educators to identify which change precipitated difficulty and to avoid equating high fidelity with educational value. The sequence uses planned variation in specific dimensions of case difficulty.

Between cycles, educators review evidence across multiple cases. If performance is unstable, learners return to the function that addresses the observed need: conceptual framing for a knowledge-organization gap, modeling for an invisible strategy, supported practice for an execution gap, or feedback and reflection for weak monitoring. If performance is stable, the next cycle increases selected dimensions of difficulty while withdrawing only the support that has become unnecessary (Kalyuga, 2007). This recursive decision rule governs progression within the course schedule.

At program level, repeated cycles require coordinated handoffs among courses. A preclerkship team may establish shared representations and model comparison; a transition course may increase whole-case commitment and feedback; clerkships may sample reduced-support reasoning in distinct clinical contexts. The receiving course needs a completion record and a concise account of the demands already sampled, supports still required, recurrent error patterns, and the evidence used to justify greater responsibility. Common terminology, a small set of shared criteria, and scheduled review of learner and curriculum evidence can support this coordination across different platforms.

One illustrative four-year arrangement could revisit the sepsis case family in three program-level cycles without assigning one CRAFT function to a particular year. In Year 1, learners interpret a prototypical deterioration with a schema, a clinician model, and low-stakes feedback. Near the transition to clerkships, they manage atypical presentations and comorbidities with fewer prompts, commit to a plan, and complete moderate-stakes case-based assessments. During core clerkships and the final year, they lead reasoning in dynamic cases with interruptions and resource constraints, using routine clinical resources and supervision for safety; progression decisions draw on higher-stakes, triangulated evidence across cases, contexts, and assessors. The same functions recur at each level, but

contextual complexity, learner responsibility, assessment stakes, and faculty support change.

Table 4 is an illustrative design demonstration, not empirical evidence. It traces a single sepsis case family from prerequisite preparation through an advanced cycle, showing what changes in the case, the learner's responsibility, the support provided, and the evidence used to justify progression.

CRAFT operationalizes situativity perspectives on clinical reasoning through the design of case families (Merkebu et al., 2020). These perspectives locate reasoning performance in dynamic interactions between practitioners and their contexts. Using a case family across developmental cycles allows educators to vary contextual conditions while retaining core pathophysiologic principles. For instructional design and analysis, selected contextual dimensions (e.g., triage settings, team dynamics, communication noise, resource availability) can be escalated while the underlying clinical domain remains recognizable. It also has limits: repeated exposure can produce case-specific familiarity that is mistaken for transfer. The sepsis illustration should therefore be paired with other acute and non-acute case families and with cases that share reasoning demands but differ in surface features. Educators can then examine whether learners recognize instability, revise priorities, and coordinate management because they have organized transferable relations or because they remember a rehearsed script. The example illustrates developmental contrasts within one case family as part of a curriculum that includes multiple diagnoses and reasoning demands.

7 Testable theoretical propositions

Table 5 presents five testable propositions. Each is falsifiable by comparing CRAFT-aligned design features with plausible alternatives. The propositions concern the incremental effects of the specified coordination decisions.

The propositions specify hypothesized causal sequences. Each begins with a manipulable design feature, identifies a mechanism that should change during learning, and predicts a proximal performance pattern before a more distal reasoning or transfer outcome. This ordering helps identify which predicted relationships are supported. For example, if adaptive fading does not reduce prompt dependence, there is little basis for attributing a later transfer difference to that mechanism. If prompt dependence falls but transfer does not improve, the hypothesized bridge from reduced support to generalization requires revision. Mechanism measures complement outcome measures by identifying where hypothesized relationships fail.

Initial studies can evaluate individual model components. Component studies can compare dynamic modeling with conclusion-only demonstration, performance-contingent with scheduled fading, or feedback-plus-reattempt with feedback alone. Factorial or sequential designs can examine interactions among components, while longitudinal comparisons can test whether progressively varied cycles add value beyond repeated same-level practice. Moderation analyses should examine prior knowledge, task domain, case complexity, and resource setting because average effects could conceal expertise reversal or implementation burden. Evaluation of the integrated model requires comparison with well-implemented existing instruction.

8 Curriculum implementation and feasibility

Curriculum-level implementation requires coordinated governance across courses. A cross-course group should define common outcomes and terminology, inventory existing cases and assessments, identify where each function is already strong or absent, and coordinate the developmental parameters that change across cycles. Faculty development should include demonstration of think-aloud modeling, calibration of coaching and feedback, recognition of readiness evidence, and use of common rubrics; longitudinal, practice-based formats are more likely than single workshops to change teaching behavior (Steinert et al., 2016). The need is practical: lack of time and faculty expertise are repeatedly reported barriers, and longitudinal implementations describe faculty development and coordination as enabling conditions (Rencic et al., 2017; Rowat and Suneja, 2022; Singh et al., 2022).

A feasible implementation can proceed in three stages. Mapping identifies where current activities already perform each function, where learners commit before receiving answers, and where progression decisions lack evidence. A bounded pilot then revises one case family or course sequence, documents faculty time and learner workload, and tests whether instructors can apply the terminology and transition criteria consistently. Scaling occurs only after the pilot identifies a sustainable case-production process, feedback workflow, faculty-development plan, and data review cycle. This staged approach allows institutions to retain effective local practices and avoids treating wholesale curricular redesign as a condition of fidelity.

Five commitments are essential to the proposed model: whole-case progression, visible modeling, learner commitment before definitive feedback, evidence-informed fading, and transfer sampling with reduced support. Delivery format, pacing, course location, technology, team size, and the mix of faculty, peer, and digital feedback are adaptable. A lower-resource implementation might use a shared bank of written cases, reusable clinician think-aloud recordings, structured peer comparison, a brief common rubric, and periodic faculty calibration. A higher-resource implementation may add standardized patients, simulation, virtual patients, analytics, or locally validated AI. Institutions should not adopt resource-intensive components if they displace case sampling, feedback quality, or faculty oversight.

Implementation changes the distribution of faculty workload. Preparing contrastive cases and calibrated think-alouds requires substantial initial effort, whereas reusable cases, shared rubrics, peer comparison, and faculty calibration may reduce recurring demands. Protected coordination time is likely to be more important than a specific technology. Adaptability should also be evaluated for equity: learners should have comparable access to cases, feedback, remediation, and trusted resources; accommodations should preserve the intended reasoning construct; and additional practice assigned by an adaptive system should not become an invisible penalty. These conditions

TABLE 4 Illustrative sepsis case family across developmental CRAFT cycles. The sepsis case family is an instructional design demonstration informed by Sepsis-3 and the Surviving Sepsis Campaign (Evans et al., 2021; Singer et al., 2016). It does not report learner outcomes or establish effectiveness.

Developmental point	Case and readiness	Learner responsibility	Support	Progression evidence
Prerequisite preparation	Foundational biomedical concepts, vital signs, basic history/exam language	Interpret short patient narratives and organize concepts in preparation for sepsis reasoning	Direct teaching and worked examples	Can explain infection, perfusion, and organ dysfunction in a simplified case
Early CRAFT cycle	Common infection with clear deterioration and limited competing diagnoses	Create a concise representation; identify acuity, likely source, immediate information needs	High support: static schema, staged model, worksheet, immediate debrief	Stable performance across several prototypical cases; can explain why cues alter urgency
Intermediate CRAFT cycle	Atypical presentation, comorbidity, incomplete data, sepsis mimics	Prioritize competing diagnoses and management while revising as data change	Moderate support: brief prompts, partial worksheet, delayed feedback	Accurate prioritization without critical omission across varied cases; prompts rarely required
Advanced CRAFT cycle	Dynamic shock state, interruptions, resource constraints, handoffs, multiple problems	Lead assessment and management, communicate uncertainty, escalate appropriately	Minimal support with supervision for safety; post-event feedback	Consistent multi-case performance, contextual adaptation, appropriate help-seeking, and delayed transfer

TABLE 5 CRAFT propositions, mechanisms, outcomes, and boundary conditions.

Proposition	CRAFT component	Hypothesized mechanism	Proximal outcome	Clinical-reasoning and transfer outcome	Boundary conditions
P1 Readiness matching	Entry criteria align case complexity and support with prior knowledge.	Working-memory demands remain interpretable and task-relevant.	More coherent problem representations and lower extraneous demand.	Better cross-case accuracy and fewer critical omissions.	Benefits diminish if readiness criteria are too restrictive or insensitive to domain-specific knowledge.
P2 Visible-to-active transition	Dynamic modeling and contrastive examples precede learner commitment.	Learners discriminate consequential cues and externalize evolving representations.	Improved explanation of probability and priority shifts.	Stronger performance than description-only instruction on related cases.	Passive observation without explanation or commitment may nullify the effect.
P3 Adaptive fading	Consistent performance across varied cases determines when scaffolds are reduced.	Support remains available until routines stabilize, then avoids redundancy.	Reduced prompt dependence without increased safety-critical error.	Greater delayed transfer and independence than scheduled fading.	Requires feasible, trustworthy performance evidence; excessive delay may create dependence.
P4 Feedback-to-action loop	Criterion-referenced feedback is paired with a specific plan and a subsequent case.	Calibration and error correction are linked to another performance opportunity.	The targeted reasoning behavior changes on the next case.	More durable improvement than feedback or reflection alone.	Ritualized reflection, weak criteria, or absent reattempts may eliminate benefit.
P5 Progressive recursion	Successive cycles increase selected dimensions of complexity and responsibility while reducing support.	Variability and retrieval reorganize knowledge for context-sensitive use.	Improved consistency across increasingly dissimilar cases.	Greater transfer and appropriate autonomy than repeating same-level cases or completing one linear sequence.	Rapid increases may overload learners; weak coordination may disrupt developmental continuity.

should be evaluated as dimensions of implementation quality from the outset.

Technology, including artificial intelligence, is an optional implementation support rather than a theoretical component of CRAFT. Tools may release information, elicit learner commitment, provide practice, or display reference paths, but they should not disclose a definitive answer before the learner commits to a representation and plan. Cases, reference materials, and outputs require faculty curation, local validation, periodic audit, explicit data governance, and a path to human

review. Evidence that a model can influence diagnostic reasoning does not establish readiness for autonomous clinical deployment (Goh et al., 2024; Hager et al., 2024; Thirunavukarasu et al., 2023; World Health Organization, 2025).

Technology selection should follow the instructional function. A low-technology design that samples more cases and produces actionable feedback may align more closely with CRAFT than a sophisticated platform that reveals answers early or generates unreviewed judgments.

9 Learner assessment and proposed evaluation strategy

Learner assessment should be construct-aligned and sampled broadly enough to support the intended inference. Problem representation, differential prioritization, use of supporting and opposing evidence, management reasoning, response to uncertainty, critical omissions, and metacognitive calibration are related but separable targets. Key-feature questions, post-encounter notes, structured case discussions, chart-stimulated recall, simulations, OSCE stations, and workplace observations each sample different parts of the construct (Daniel et al., 2019; Page et al., 1995; Sutherland et al., 2019; Thammasitboon et al., 2018). No single diagnosis, case, self-report, or global impression is sufficient.

The assessment plan should mirror the intended developmental inference without turning the five functions into five separate scores. Early evidence may emphasize knowledge organization and cue interpretation, but later decisions should require integrated performance on whole cases. Process evidence can be useful when it is linked to a defined construct, yet verbose explanations are not necessarily better reasoning, and agreement with one expert path is not the only defensible solution. Assessment should credit clinically coherent alternatives, appropriate uncertainty, and help-seeking while preserving explicit thresholds for safety-critical omissions.

Validity concerns whether the available evidence supports the proposed interpretation and use of assessment results. Case specificity and rater effects require multiple cases and, for observational judgments, multiple occasions or assessors. Programs should use a blueprint, explicit criteria, rater calibration, and psychometric or generalizability analyses appropriate to the stakes. Written case summaries and other mapping exercises demonstrate that meaningful variance may arise from learner-by-case interactions, reinforcing the need for sampling (Dory et al., 2016). Programmatic assessment can integrate heterogeneous evidence (Schuwirth and van der Vleuten, 2011), but recent multisite work shows that predictive relations vary by institution and should not be assumed to transport automatically (Guth et al., 2024).

Reliability concerns scores obtained under specified sampling conditions and must be evaluated for the intended use of a rubric. Increasing the number and diversity of cases may contribute more to dependable decisions than adding fine-grained items within one case. Rater training can improve shared interpretation, but over-standardization may suppress legitimate disciplinary judgment. Programs should examine score distributions, sources of variance, decision consistency, and consequences of false progression or false remediation. The amount of evidence required should rise with the stakes, and local monitoring should continue after implementation because task and rater conditions change.

For grading or competency decisions, the decision rule should be specified before implementation. Low-stakes data can guide feedback and case assignment; higher-stakes progression should require convergent evidence across methods, cases, contexts, and time, with documented procedures for review and remediation. Entrustment or workplace decisions should emphasize the level of supervision required for a defined activity and include appropriate help-seeking within the relevant clinical context (Ten Cate, 2013). CRAFT's proposed readiness gates can inform curriculum decisions when local outcomes, evidence, and progression rules are aligned and validated for the intended stakes; they are not established standards for learner advancement.

Consequences are part of the validity argument. A readiness gate that triggers targeted support may be defensible with relatively imprecise evidence when the cost of an incorrect decision is low and the learner can be reassessed quickly. The same evidence would be inadequate for delayed progression or an entrustment decision. Programs should examine whether gates create bottlenecks, whether remediation addresses the inferred need, and whether subgroup differences arise from access, opportunity, case familiarity, or the construct itself. Learners should know what evidence is used, how it is reviewed, and how they can demonstrate improvement. Transparent procedures reduce the risk that adaptive support becomes opaque tracking.

Implementation evaluation asks different questions. Fidelity indicators include whether whole cases were sequenced as intended, modeling occurred, learners committed before feedback, scaffolds were faded contingently, and transfer was sampled. Feasibility and acceptability indicators include faculty time, learner workload, case availability, simulation or technology demands, coordination burden, and access to feedback; program-level internal evaluation offers established methods for collecting them (Kodagoda Gamage et al., 2024). Equity indicators include differential participation, feedback quality, accommodations, and subgroup outcomes. These indicators characterize CRAFT implementation and should be analyzed separately from measures of clinical-reasoning competence.

Fidelity should include both adherence and responsiveness. Adherence asks whether the planned functions, case progression, commitment, feedback, and fading occurred. Responsiveness asks whether educators adjusted support when evidence indicated that the plan no longer fit the learner. A program could deliver every scheduled worksheet and still violate the model if prompts were withdrawn despite unsafe performance or retained after they became redundant. Dose counts are therefore insufficient on their own. Brief observations, case artifacts, learner work, faculty logs, and interviews can be combined to determine whether the hypothesized decision logic was actually used.

The following strategy is proposed for an initial empirical evaluation. A feasibility pilot could compare baseline, immediate, and delayed performance across multiple cases while documenting fidelity and resources. A contemporaneous comparator, stepped implementation, interrupted time series, or

cluster design would strengthen inference. Without a comparator, improvement after implementation could reflect maturation, concurrent curriculum changes, repeated testing, or increased clinical exposure and would not establish superiority. Mixed methods can examine how mechanisms operate and why effects differ across learners or settings; iterative educational design research can refine the model before a larger comparative study (Cook, 2010; McKenney and Reeves, 2021).

The comparator should match the question. A component trial may use the same cases with different fading or feedback rules; a course evaluation may compare CRAFT-aligned sequencing with an established local sequence; and a program-level study may use staggered adoption while preserving common assessments. Outcomes should include an immediate related-case measure, delayed performance, and dissimilar-case transfer, alongside mechanism and implementation measures. Analyses should prespecify the level of inference, account for clustering and repeated observations where applicable, and report null or adverse findings. Evidence of benefit in one domain should not be generalized automatically to other specialties or workplace conditions.

10 Discussion

CRAFT's proposed contribution is the explicit coordination of established approaches through five instructional functions, proposed readiness gates, developmental recursion, and transition criteria. Table 2 documents substantial overlap with existing approaches. Within CRAFT, progressive case families operationalize situativity by treating diagnostic and management performance as context-dependent and systematically varying the setting, workplace constraints, and interprofessional interactions (Merkebu et al., 2020). The empirical question is whether the coordinated architecture improves design decisions, implementation coherence, and learner transfer beyond established alternatives.

A practical use of the model before effectiveness is established is as an audit tool. Faculty can ask whether learners encounter an interpretable whole task, whether expert reasoning is visible, whether learners must commit to a response, whether feedback changes a later attempt, and whether reduced-support performance is sampled. The answers may reveal a missing relation without requiring adoption of every component. Such an audit can clarify design decisions; causal claims about subsequent changes require additional evidence.

The propositions create a chain from component to mechanism to outcome. They also expose potential failure. A readiness gate may delay useful early exposure if interpreted too rigidly; whole-task epitomes may still overload learners; modeling may invite passive imitation; scaffolds may produce persistent dependence; reflection may become ritualized; and programmatic evidence may be too burdensome to collect. An instructional model that requires repeated cases, expert modeling, feedback, and cross-course coordination may be less feasible where faculty time, case access, assessment infrastructure, or protected curricular space is limited. These constraints may moderate both implementation and educational outcomes.

Several competing explanations should be retained in future studies. Increased time on task, additional cases, greater expert contact, assessment practice, and clearer expectations could each account for improved outcomes independently of CRAFT's coordination decisions.

Excessive complexity, inconsistent coaching, delayed feedback, or burdensome documentation could also contribute to poorer outcomes and complicate interpretation of the underlying mechanisms. Measuring these alternatives is necessary because a flexible model can otherwise become unfalsifiable: every success could be credited to adaptation and every failure attributed to poor implementation. The propositions and fidelity criteria are intended to prevent that circularity.

CRAFT may be inappropriate when the immediate goal is isolated factual acquisition, a brief procedural orientation, or urgent just-in-time performance that cannot safely support guided experimentation. It should not be used to justify unsupervised practice by learners who lack prerequisite knowledge, or to impose identical functions on settings whose professional roles and reasoning constructs differ. Some courses already use coherent whole-case progression, coaching, feedback, and adaptive fading; in those settings, relabeling instruction as CRAFT may add no value. The model also requires adaptation before use outside undergraduate medical education or across cultural, resource, and regulatory contexts.

Transfer claims should specify the domain and conditions in which learning is expected to generalize. CRAFT proposes that organized knowledge and adaptable routines can support performance on cases that differ along specified dimensions within a defensible domain, and that repeated contextual variation may broaden that range. Studies should therefore describe the relationship between learning and transfer cases, the supports available, and the clinical contexts sampled. Far transfer claims should require correspondingly stronger evidence.

Several limitations follow from the model's development. CRAFT has not been empirically validated as an integrated curriculum, and its incremental benefit is unknown. The synthesis was informed by theory and author judgment without a systematic model-development method, and the selected traditions may omit competing perspectives. The literature selection emphasized independent consensus statements, systematic reviews, empirical studies, and implemented curricula; the authors' prior work was retained where directly relevant to situativity or the model's development. Because the review was not systematic, selection bias remains possible. Finally, the sepsis example demonstrates design translation only; it cannot establish feasibility, effectiveness, validity evidence, or transfer.

The model's parsimony is also a limitation. Five functions may conceal important distinctions within diagnostic, management, collaborative, and ethical reasoning, and the same function may require different instructional forms across specialties. The readiness logic assumes that useful evidence can be collected frequently enough to guide design. In distributed clinical placements, readiness gates may need to guide support rather than serve as mandatory progression thresholds. Fixed schedules may also limit recursive support, requiring planned reassessment

opportunities or fewer gates. These tensions are boundary conditions and may require specialty-specific variants or a different model.

The research agenda should proceed in stages. First, content validity evidence and usability studies should test whether educators can distinguish the five functions and apply the transition criteria consistently. Second, feasibility studies should estimate faculty time, case and assessment requirements, fidelity, and acceptability across contrasting institutions. Third, mechanism-focused studies should test the propositions, including readiness matching, performance-contingent fading against scheduled fading, feedback-plus-action planning, and successive developmental cycles. Only comparative designs with credible alternatives can address whether CRAFT adds value beyond existing clinical-reasoning curricula or generic instructional design models.

Progression through this agenda should include stop or revision rules. If educators cannot distinguish the functions reliably, the terminology and operational table require revision before outcome testing. If a pilot consistently exceeds available faculty time or disadvantages particular learner groups, the implementation design should be revised. If component mechanisms are not observed, a large integrated trial may be premature. Conversely, convergent evidence across usability, feasibility, mechanisms, and transfer would justify broader comparative testing. This staged approach supports revision of CRAFT in response to empirical evidence.

11 Conclusion

CRAFT is a proposed instructional design model for coordinating established educational principles in clinical-reasoning curricula. Its five functions recur in developmental cycles guided by proposed readiness criteria. The design operationalizes situativity through planned variation in clinical context, learner responsibility, and instructional support. Progression follows locally validated evidence and does not map to fixed curricular years. The model distinguishes clinical reasoning from its instruction and assessment and specifies transition criteria, testable theoretical propositions, implementation requirements, and boundary conditions. Existing approaches may already accomplish many of the same goals; CRAFT's incremental value and underlying mechanisms remain to be established empirically. Staged studies should evaluate usability, mechanisms, feasibility, learner outcomes, and transfer against credible comparators before wider adoption.

Data availability statement

No new data were created or analyzed in this study. All conceptual material is contained in the article.

Author contributions

EJ: Writing – original draft, Writing – review & editing, Conceptualization, Formal analysis, Investigation, Methodology, Resources, Visualization. CR: Writing – original draft, Writing – review & editing. BZ: Writing – original draft, Writing – review & editing. SD: Writing – original draft, Writing – review & editing. JM: Writing – original draft, Writing – review & editing.

Funding

The author(s) declared that financial support was not received for this work and/or its publication.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that generative AI was used in the creation of this manuscript. Generative AI was used to assist with language editing and organizational drafting. ChatGPT (OpenAI, GPT-5.6) was used to improve clarity, organize the manuscript, and expand explanatory text.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Author disclaimer

The opinions and assertions expressed herein are those of the authors and do not necessarily reflect the official policy or position of the Uniformed Services University or the Department of War.

References

Bowen, J. L. (2006). Educational strategies to promote clinical diagnostic reasoning. *N. Engl. J. Med.* 355 (21), 2217–2225. doi: 10.1056/NEJMra054782

Braun, L. T., Borrmann, K. F., Lottspeich, C., Heinrich, D. A., Kiesewetter, J., Fischer, M. R., et al. (2019). Scaffolding clinical reasoning of medical students with virtual patients: effects on diagnostic accuracy, efficiency, and errors. *Diagnosis* 6 (2), 137–149. doi: 10.1515/dx-2018-0090

Braun, L. T., Zottmann, J. M., Adolf, C., Lottspeich, C., Then, C., Wirth, S., et al. (2017). Representation scaffolds improve diagnostic efficiency in medical students. *Med. Educ.* 51 (11), 1118–1126. doi: 10.1111/medu.13355

Chamberland, M., Mamede, S., St-Onge, C., Setrakian, J., Bergeron, L., and Schmidt, H. (2015). Self-explanation in learning clinical reasoning: the added value of examples and prompts. *Med. Educ.* 49 (2), 193–202. doi: 10.1111/medu.12623

Charlin, B., Tardif, J., and Boshuizen, H. P. A. (2000). Scripts and medical diagnostic knowledge: theory and applications for clinical reasoning instruction and research. *Acad. Med.* 75 (2), 182–190. doi: 10.1097/00001888-200002000-00020

Choi, S., Oh, S., Lee, D. H., and Yoon, H. S. (2020). Effects of reflection and immediate feedback to improve clinical reasoning of medical students in the assessment of dermatologic conditions: a randomised controlled trial. *BMC Med. Educ.* 20, 146. doi: 10.1186/s12909-020-02063-y

Collins, A., Brown, J. S., and Holum, A. (1991). Cognitive apprenticeship: making thinking visible. *Am. Educ.* 15 (3), 6–11.

Connor, D. M., Narayana, S., and Dhaliwal, G. (2022). A clinical reasoning curriculum for medical students: an interim analysis. *Diagnosis* 9 (2), 265–273. doi: 10.1515/dx-2021-0112

Cook, D. A. (2010). Twelve tips for evaluating educational programs. *Med. Teach.* 32 (4), 296–301. doi: 10.3109/01421590903480121

Cooper, N., Bartlett, M., Gay, S., Hammond, A., Lillicrap, M., Matthan, J., et al. (2021). Consensus statement on the content of clinical reasoning curricula in undergraduate medical education. *Med. Teach.* 43 (2), 152–159. doi: 10.1080/0142159X.2020.1842343

Daniel, M., Rencic, J., Durning, S. J., Holmboe, E., Santen, S. A., Lang, V., et al. (2019). Clinical reasoning assessment methods: a scoping review and practical guidance. *Acad. Med.* 94 (6), 902–912. doi: 10.1097/ACM.0000000000002618

Dory, V., Gagnon, R., Charlin, B., Vanpee, D., Leconte, S., Duyver, C., et al. (2016). In brief: validity of case summaries in written examinations of clinical reasoning. *Teach. Learn. Med.* 28 (4), 375–384. doi: 10.1080/10401334.2016.1190730

Duca, N. S., and Glod, S. (2019). Bridging the gap between the classroom and the clerkship: a clinical reasoning curriculum for third-year medical students. *MedEdPORTAL* 15, 10800. doi: 10.15766/mep_2374-8265.10800

Elvén, M., Welin, E., Wiegleb Edström, D., Petreski, T., Szopa, M., Durning, S. J., et al. (2023). Clinical reasoning curricula in health professions education: a scoping review. *J. Med. Educ. Curric. Dev.* 10, 23821205231209093. doi: 10.1177/23821205231209093

Ericsson, K. A. (2004). Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Acad. Med.* 79 (Suppl 2), S70–S81. doi: 10.1097/00001888-200410001-00022

Eva, K. W., and Norman, G. R. (2005). Heuristics and biases: a biased perspective on clinical reasoning. *Med. Educ.* 39 (9), 870–872. doi: 10.1111/j.1365-2929.2005.02258.x

Evans, L., Rhodes, A., Alhazzani, W., Antonelli, M., Coopersmith, C. M., French, C., et al. (2021). Surviving sepsis campaign: international guidelines for management of sepsis and septic shock. *Intensive Care Med.* 47 (11), 1181–1247. doi: 10.1007/s00134-021-06506-y

Gobet, F. (2005). Chunking models of expertise: implications for education. *Appl. Cogn. Psychol.* 19 (2), 183–204. doi: 10.1002/acp.1110

Goh, E., Gallo, R., Hom, J., Strong, E., Weng, Y., Kerman, H., et al. (2024). Large language model influence on diagnostic reasoning: a randomized clinical trial. *JAMA Netw. Open* 7 (10), e2440969. doi: 10.1001/jamanetworkopen.2024.40969

Guth, T. A., Wolfe, R. M., Martinez, O., Subhiyah, R. G., Henderek, J. J., McAllister, C., et al. (2024). Assessment of clinical reasoning in undergraduate medical education: a pragmatic approach to programmatic assessment. *Acad. Med.* 99 (8), 912–921. doi: 10.1097/ACM.0000000000005665

Hager, P., Jungmann, F., Holland, R., Bhagat, K., Hubrecht, I., Knauer, M., et al. (2024). Evaluation and mitigation of the limitations of large language models in clinical decision making. *Nat. Med.* 30 (9), 2613–2622. doi: 10.1038/s41591-024-03097-1

Harden, R. M. (1999). What is a spiral curriculum? *Med. Teach.* 21 (2), 141–143. doi: 10.1080/01421599979752

Jagannath, A. D., Dreicer, J. J., Penner, J. C., and Dhaliwal, G. (2023). The cognitive apprenticeship: advancing reasoning education by thinking aloud. *Diagnosis* 10 (1), 9–12. doi: 10.1515/dx-2022-0043

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educ. Psychol. Rev.* 19 (4), 509–539. doi: 10.1007/s10648-007-9054-3

Kassirer, J. P. (2010). Teaching clinical reasoning: case-based and coached. *Acad. Med.* 85 (7), 1118–1124. doi: 10.1097/ACM.0b013e3181d5dd0d

Kiesewetter, J., Sailer, M., Jung, V. M., Schönberger, R., Bauer, E., Zottmann, J. M., et al. (2020). Learning clinical reasoning: how virtual patient case format and prior knowledge interact. *BMC Med. Educ.* 20, 73. doi: 10.1186/s12909-020-1987-y

Kodagoda Gamage, S., Jowsey, T., Bishop, J., Forbes, M., Grant, L.-J., Green, P., et al. (2024). Internal evaluation of medical programs is more than housework: a scoping review. *PLoS One* 19 (10), e0305996. doi: 10.1371/journal.pone.0305996

Konopasky, A., Artino, A. R., Battista, A., Ohmer, M., Hemmer, P. A., Torre, D., et al. (2020). Understanding context specificity: the effect of contextual factors on clinical reasoning. *Diagnosis* 7 (3), 257–264. doi: 10.1515/dx-2020-0016

Lefroy, J., Watling, C., Teunissen, P. W., and Brand, P. (2015). Guidelines: the do's, don'ts and don't knows of feedback for clinical education. *Perspect. Med. Educ.* 4 (6), 284–299. doi: 10.1007/s40037-015-0231-7

Levin, M., Cennimo, D., Chen, S., and Lamba, S. (2016). Teaching clinical reasoning to medical students: a case-based illness script worksheet approach. *MedEdPORTAL* 12, 10445. doi: 10.15766/mep_2374-8265.10445

Mamede, S., and Schmidt, H. G. (2023). Deliberate reflection and clinical reasoning: founding ideas and empirical findings. *Med. Educ.* 57 (1), 76–85. doi: 10.1111/medu.14863

Mann, K., Gordon, J., and MacLeod, A. (2009). Reflection and reflective practice in health professions education: a systematic review. *Adv. Health Sci. Educ.* 14 (4), 595–621. doi: 10.1007/s10459-007-9090-2

McKenney, S., and Reeves, T. C. (2021). Educational design research: portraying, conducting, and enhancing productive scholarship. *Med. Educ.* 55 (1), 82–92. doi: 10.1111/medu.14280

Merkebu, J., Battistone, M., McMains, K., McOwen, K., Witkop, C., Konopasky, A., et al. (2020). Situativity: a family of social cognitive theories for understanding clinical reasoning and diagnostic error. *Diagnosis* 7 (3), 169–176. doi: 10.1515/dx-2019-0100

Moulton, C. A., Regehr, G., Mylopoulos, M., and MacRae, H. M. (2007). Slowing down when you should: a new model of expert judgment. *Acad. Med.* 82 (10 Suppl), S109–S116. doi: 10.1097/ACM.0b013e3181405a76

Nelson, N. R., and Rhoney, D. H. (2024). Use of a clinical reasoning scaffolding document improves student performance. *Curr. Pharm. Teach. Learn.* 16 (7), 102102. doi: 10.1016/j.cptl.2024.04.018

Page, G., Bordage, G., and Allen, T. (1995). Developing key feature problems and examinations to assess clinical decision making skills. *Acad. Med.* 70 (3), 194–201. doi: 10.1097/00001888-199503000-00009

Reigeluth, C. M. (1999). "The elaboration theory: guidance for scope and sequence decisions," in *Instructional Design Theories and Models: A New Paradigm of Instructional Theory, Vol. II*, ed. C. M. Reigeluth (Mahwah: Lawrence Erlbaum Associates), 425–453.

Rencic, J., Trowbridge, R. L. Jr., Fagan, M., Szauter, K., and Durning, S. J. (2017). Clinical reasoning education at US medical schools: results from a national survey of internal medicine clerkship directors. *J. Gen. Intern. Med.* 32 (11), 1242–1246. doi: 10.1007/s11606-017-4159-y

Rowat, J., and Suneja, M. (2022). Longitudinal clinical reasoning theme embedded across four years of a medical school curriculum. *Diagnosis* 9 (4), 468–475. doi: 10.1515/dx-2022-0046

Ruczynski, L. I. A., Schouwenberg, B. J. J. W., Fluit, C. R. M. G., and Van De Pol, M. H. J. (2025). The power of peers: design-based research on stimulating peer-assisted learning for enhancing the clinical-reasoning learning process in the workplace. *Med. Educ.* 59 (7), 739–749. doi: 10.1111/medu.15613

Schmidt, H. G., and Mamede, S. (2015). How to improve the teaching of clinical reasoning: a narrative review and a proposal. *Med. Educ.* 49 (10), 961–973. doi: 10.1111/medu.12775

Schuwirth, L. W. T., and Van Der Vleuten, C. P. M. (2011). Programmatic assessment: from assessment of learning to assessment for learning. *Med. Teach.* 33 (6), 478–485. doi: 10.3109/0142159X.2011.565828

Singer, M., Deutschman, C. S., Seymour, C. W., Shankar-Hari, M., Annane, D., Bauer, M., et al. (2016). The third international consensus definitions for sepsis and septic shock. *JAMA* 315 (8), 801–810. doi: 10.1001/jama.2016.0287

Singh, M., Collins, L., Farrington, R., Jones, M., Thampy, H., Watson, P., et al. (2022). From principles to practice: embedding clinical reasoning as a longitudinal curriculum theme in a medical school programme. *Diagnosis* 9 (2), 184–194. doi: 10.1515/dx-2021-0031

Stark, R., Kopp, V., and Fischer, M. R. (2011). Case-based learning with worked examples in complex domains: two experimental studies in undergraduate medical education. *Learn. Instr.* 21 (1), 22–33. doi: 10.1016/j.learninstruc.2009.10.001

Steinert, Y., Mann, K., Anderson, B., Barnett, B. M., Centeno, A., Naismith, L., et al. (2016). A systematic review of faculty development initiatives designed to enhance teaching effectiveness: a 10-year update: BEME Guide No. 40. *Med. Teach.* 38 (8), 769–786. doi: 10.1080/0142159X.2016.1181851

Sutherland, R. M., Reid, K. J., Chiavaroli, N. G., Smallwood, D., and McColl, G. J. (2019). Assessing diagnostic reasoning using a standardized case-based discussion. *J. Med. Educ. Curric. Dev.* 6, 2382120519849411. doi: 10.1177/2382120519849411

Sweller, J. (1988). Cognitive load during problem solving: effects on learning. *Cogn. Sci.* 12 (2), 257–285. doi: 10.1207/s15516709cog1202_4

Sweller, J., Van Merriënboer, J. J. G., and Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educ. Psychol. Rev.* 31, 261–292. doi: 10.1007/s10648-019-09465-5

Ten Cate, O. (2013). Nuts and bolts of entrustable professional activities. *J. Grad. Med. Educ.* 5 (1), 157–158. doi: 10.4300/JGME-D-12-00380.1

Thammasitboon, S., Rencic, J., Trowbridge, R., Olson, A. P. J., Sur, M., and Dhaliwal, G. (2018). The assessment of reasoning tool (ART): structuring the conversation between teachers and learners. *Diagnosis* 5 (4), 197–203. doi: 10.1515/dx-2018-0052

Theobald, K. A., and Ramsbotham, J. (2019). Inquiry-based learning and clinical reasoning scaffolds: an action research project to support undergraduate students' learning to think like a nurse. *Nurse Educ. Pract.* 38, 59–65. doi: 10.1016/j.nepr.2019.05.018

Thirunavukarasu, A. J., Ting, D. S. J., Elangovan, K., Gutierrez, L., Tan, T. F., and Ting, D. S. W. (2023). Large language models in medicine. *Nat. Med.* 29 (8), 1930–1940. doi: 10.1038/s41591-023-02448-8

Van de Pol, J., Volman, M., and Beishuizen, J. (2010). Scaffolding in teacher-student interaction: a decade of research. *Educ. Psychol. Rev.* 22 (3), 271–296. doi: 10.1007/s10648-010-9127-6

Van Gog, T., Kester, L., and Paas, F. (2011). Effects of worked examples, example-problem, and problem-example pairs on novices' learning. *Contemp. Educ. Psychol.* 36 (3), 212–218. doi: 10.1016/j.cedpsych.2010.10.004

Van Merriënboer, J. J. G., Kirschner, P. A., and Kester, L. (2003). Taking the load off a learner's mind: instructional design for complex learning. *Educ. Psychol.* 38 (1), 5–13. doi: 10.1207/S15326985EP3801_2

Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Cambridge: Harvard University Press.

Wark, S., Drovandi, A., Alele, F. O., McGee, R. G., Mwangi, F., Malau-Aduli, B., et al. (2025). How and when should clinical reasoning be taught in undergraduate medicine: a systematic review and meta-analyses. *Perspect. Med. Educ.* 14 (1), 1021–1042. doi: 10.5334/pme.1986

Woods, N. N. (2007). Science is fundamental: the role of biomedical knowledge in clinical reasoning. *Med. Educ.* 41 (12), 1173–1177. doi: 10.1111/j.1365-2923.2007.02911.x

Woods, N. N., Brooks, L. R., and Norman, G. R. (2007). The role of biomedical knowledge in diagnosis of difficult clinical cases. *Adv. Health Sci. Educ.* 12 (4), 417–426. doi: 10.1007/s10459-006-9054-y

World Health Organization (2025). *Ethics and Governance of Artificial Intelligence for Health: Guidance on Large Multimodal Models*. Geneva: World Health Organization.

Zimmerman, B. J. (2002). Becoming a self-regulated learner: an overview. *Theory Pract.* 41 (2), 64–70. doi: 10.1207/s15430421tip4102_2