Satid Thammasitboon*, Joseph J. Rencic, Robert L. Trowbridge, Andrew P.J. Olson, Moushumi Sur and Gurpreet Dhaliwal

The Assessment of Reasoning Tool (ART): structuring the conversation between teachers and learners

https://doi.org/10.1515/dx-2018-0052
Received July 16, 2018; accepted October 19, 2018; previously published online November 8, 2018

Abstract

Background: Excellence in clinical reasoning is one of the most important outcomes of medical education programs, but assessing learners' reasoning to inform corrective feedback is challenging and unstandardized.
Methods: The Society to Improve Diagnosis in Medicine formed a multi-specialty team of medical educators to develop the Assessment of Reasoning Tool (ART). This paper describes the tool development process. The tool was designed to facilitate clinical teachers' assessment of learners' oral presentation for competence in clinical reasoning and facilitate formative feedback. Reasoning frameworks (e.g. script theory), contemporary practice goals (e.g. high-value care [HVC]) and proposed error reduction strategies (e.g. metacognition) were used to guide the development of the tool.
Results: The ART is a behaviorally anchored, three-point scale assessing five domains of reasoning: (1) hypothesis-directed data gathering, (2) articulation of a problem representation, (3) formulation of a prioritized differential diagnosis, (4) diagnostic testing aligned with HVC principles and (5) metacognition. Instructional videos were created for faculty development for each domain, guided by principles of multimedia learning.
Conclusions: The ART is a theory-informed assessment tool that allows teachers to assess clinical reasoning and structure feedback conversations.

Keywords: clinical reasoning; cognitive bias; diagnostic process; feedback; instrument validation.

***Corresponding author: Satid Thammasitboon,** MD, MHPE, Center for Research, Innovation and Scholarship in Medical Education, Texas Children's Hospital, Baylor College of Medicine, 6651 Main St., Houston, TX 770303, USA, Phone: +832-826-6238, E-mail: satidt@bcm.edu
Joseph J. Rencic: Department of Internal Medicine, Tufts University School of Medicine, Boston, MA, USA
Robert L. Trowbridge: Tufts University School of Medicine, Maine Medical Center, Portland, ME, USA
Andrew P.J. Olson: Department of Internal Medicine, University of Minnesota, Minneapolis, MN, USA
Moushumi Sur: Department of Pediatrics, Texas Children's Hospital, Baylor College of Medicine, Houston, TX, USA
Gurpreet Dhaliwal: University of California, San Francisco and San Francisco VA Medical Center, San Francisco, CA, USA

Introduction

Developing and ensuring competence in clinical reasoning is one of the most important goals of medical education programs. The 2015 National Academies of Science, Engineering and Medicine report *Improving Diagnosis in Health Care* [1] highlighted the prevalence of diagnostic errors and recommended that training programs improve learners' performance in the diagnostic process to decrease diagnostic errors.

A productive feedback conversation requires that clinical teachers and learners share a common understanding of the processes used to make diagnostic decisions. This shared mental model is an essential but often elusive starting point because clinical reasoning is idiosyncratic [2]. Clinicians rely upon their own unique clinical experiences and knowledge to address clinical problems; hence, there may be as many paths to a diagnosis as there are diagnosticians. Further, expert diagnosticians often use non-analytic, experience-based approaches to make decisions and may be unaware of the underlying cognitive processes. The context-specific nature of clinical reasoning necessitates that teachers converse with learners about situational factors (e.g. student's clinical experience with the chief complaint) that can impact clinical reasoning performance [3, 4]. Lastly, teachers' own knowledge, experience, ability and even personal cognitive biases may influence his or her assessment of learners [5, 6].

Despite these limitations and challenges, assessment of learners' clinical reasoning in medical education

programs is essential to identify opportunities for improvement. Such identification is useful when teachers can follow with questions that help learners calibrate their reasoning processes and provide learners with strategies for improvement in future encounters. Questions asked during a clinical reasoning assessment allows learners to reactivate existing knowledge, incorporate newly acquired knowledge and then assimilate the two in increasingly organized and rich illness scripts [7]. In summary, a tool that facilitates assessment for learning will be most useful in guiding learners in the development toward mastery of clinical reasoning skills. Though numerous approaches can be used to assess components of clinical reasoning ability, very few assessment tools to guide feedback are comprised entirely of elements specific to the reasoning process [4, 6, 8–14].

To address this gap, we developed the Assessment of Reasoning Tool (ART). The goal of the tool is to facilitate teachers' formative assessment of learners' oral presentations for competence in clinical reasoning and provide a structure for conversations between teachers and learners during a feedback session. Recognizing that an informed faculty member is essential to successful ART use, we developed instructional videos for asynchronous, just-in-time faculty development for the five ART domains. This paper describes the tool development process and provides preliminary validity evidence.

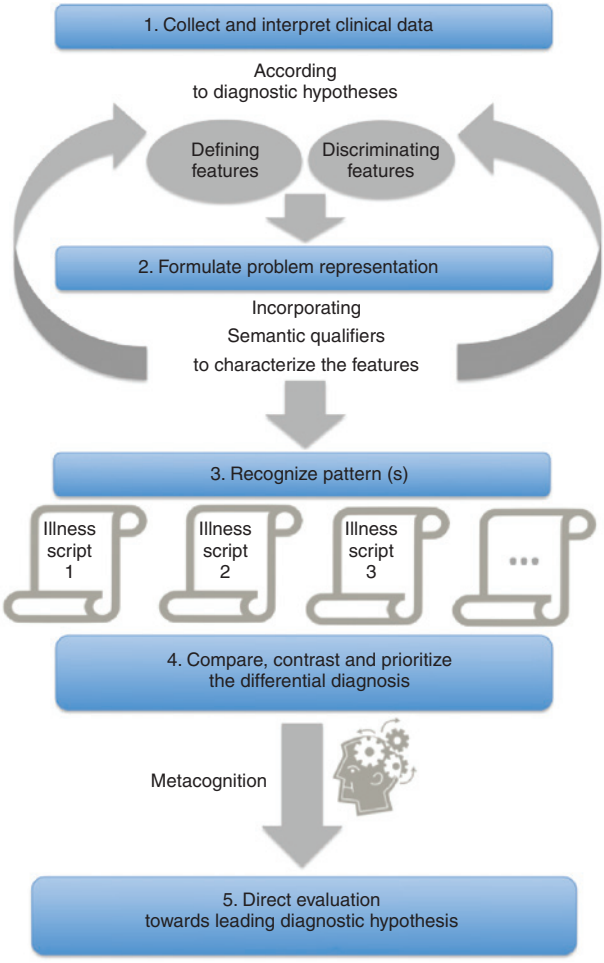

Figure 1: A proposed conceptual framework for the clinical reasoning process derived from literature and expert consensus. Defining features are clinical elements that are characteristic of a disorder; discriminating features are clinical elements that are useful for distinguishing one disorder from another. A problem representation is a concise synthesis of the clinical problem that needs to be solved. Semantic qualifiers are specific terms (e.g. epidemiological features, severity description, temporal association) used to frame the patient presentations into clear and recognizable logical paths or patterns. Embedded in these terms are some opposing qualities (e.g. acute vs. chronic) used to facilitate differentiation between potential diagnoses. Pattern recognition is a form of reasoning that is driven by recognition of familiar combination of data. Metacognition is an awareness of one's own thinking process that enables self-assessment, monitoring and reflection during the reasoning process. Cognitive debiasing is a bias-reduction strategy within judgment or decision-making. Cognitive biases are systematic patterns of deviation from normative judgment. Metacognition, not debiasing strategy, was used as a construct of the final Assessment of Reasoning Tool.

Methods

A subcommittee of the Society to Improve Diagnosis in Medicine Education Committee, composed of nine medical educators with expertise in teaching clinical reasoning and an interest in decreasing diagnostic errors, was formed. The development process for the ART followed the scale development guidelines recommended by DeVellis [15]. The study protocol was approved by the Institutional Review Board at the Baylor College of Medicine.

Determining the assessment target

We determined the ART should be used to assess learners' performance in clinical reasoning during an oral presentation of a patient encounter. The ART may be used in the context of a direct observation of a clinical encounter although that was not our primary intent, as the opportunities for direct observation in many medical education programs are limited. Through an iterative process, we formulated a conceptual framework for a diagnostic reasoning process based on prevalent reasoning concepts (e.g. structural semantics, script theory) [16–27], and proposed error-reduction strategies (e.g. cognitive debiasing) [28, 29]. We used this framework to define domains to be assessed (Figure 1).

Determining the measurement format

We developed a behaviorally anchored rating scale with descriptors to serve as standards for assessment and provide a common language

for effective feedback on performance. We created a three-point scale which calls upon teachers to determine the presence or absence of specific behaviors within a domain, rather than the extent or quality of those behaviors.

Generating descriptor pool

Based on the conceptual framework, we established six preliminary domains as a construct for generating behavioral descriptors. The preliminary domains were: (1) collect history and physical examination in a hypothesis-directed manner, (2) formulate a problem representation, (3) develop a prioritized differential diagnosis, (4) select appropriate illness scripts, (5) direct evaluation and treatment towards high-priority diagnoses, and (6) recognize potential cognitive and affective biases as sources of diagnostic errors.

We generated behavioral descriptors for each domain and categorized them into different levels (i.e. minimal, partial, complete). Through multiple iterations, we revised the descriptors to focus on specific behaviors reflected in a learner's oral presentation. During these revisions, we minimized clinical reasoning jargon and semantic ambiguity to make the ART easier to use by all teachers.

The preliminary pool of behavioral descriptors was reviewed by non-committee members with expertise in diagnostic and clinical reasoning, survey design, or measurement and psychometrics. The key tasks for this expert validation process included determining (1) how well the descriptors represented the corresponding domains, (2) how relevant the descriptors were to specific aspects of the domains, and (3) how difficult it would be for teachers to distinguish the three levels of performance. These consultants also gave feedback on clarity of the preliminary domains, ambiguity of terms, feasibility of use and additional concepts to be considered. For instance, we determined that consideration of a "do not miss" diagnosis is a promising strategy for reducing diagnostic errors and should be taught explicitly. Thus, it was incorporated as a specific behavioral descriptor in domain "develop a prioritized differential diagnosis." Realizing the importance of high-value care (HVC) [30, 31], we used this concept to specify the desired behaviors in domain "direct evaluation/treatment towards high-priority diagnoses." Whereas the third domain emphasizes that "do not miss" diagnoses must be on the learner's differential diagnosis (i.e. worst-case scenario medicine), the fourth domain stresses the logic of testing for these "do not miss" diagnoses (i.e. HVC). The synergy of these two domains provides a structure for sharing the thinking processes in how one determined which diagnoses are high priority. Performing additional tests for a "do not miss" diagnosis when the probability of disease is exceedingly low neither avoids potential errors nor characterizes HVC. We additionally recognized that cognitive and affective biases are difficult to discern during an oral presentation [32], so we replaced the goal of identifying biases with the goal of promoting metacognition (i.e. thinking about one's own thinking). Based on further discussion, we dichotomized this metacognitive domain into a two-point scale (instead of three) and added a prompting question to help teachers inquire about this issue. We eliminated the preliminary domain of "selection of appropriate illness scripts" as we realized that teachers cannot identify this step explicitly when listening to learners' oral presentation. This brought the final number of assessment domains from six to five. However, we recognize that illness scripts are an important clinical reasoning concept for feedback conversations and featured them in the faculty development videos.

Piloting the tool

To gather the validity of the content and response process [33, 34], we asked a convenience sample of committee members to administer the ART with their learners and provide feedback. The tool was further revised to reduce confusing terms and eliminate double-barreled questions. A global assessment scale was added when multiple teachers suggested that their gestalt would capture something that the sum of the individual domains might not. The final version of the ART was piloted with an additional convenience sample of 10 clinician-educators. They were asked to administer the ART with a learner, and then completed a five-point Likert scale questionnaire about six characteristics of the ART (see supplementary document). The investigator also asked them for verbal feedback about the tool.

Developing faculty training videos

We determined that faculty who would be using the ART to assess learners and provide feedback would benefit from enhanced knowledge about the clinical reasoning process and from instructions on using ART. The authors explored various forms of web-based learning and used principles of multimedia learning [35] as a framework to design the training modules. A whiteboard animation platform was chosen as a simple, cost-effective approach to convey the relevant content. We created scripts for the audiovisual whiteboard animation pertaining to each of the final five domains of ART. In each script, we specified the behavioral descriptors of performance and explained key terms and concepts related to clinical reasoning that faculty would be called upon to assess using the ART. Video scripts and video prototypes went through multiple reviews by all the authors. We asked committee members to view and evaluate the videos regarding their quality, educational value and utility.

Results

The ART is an assessment tool that delineates five domains of the clinical reasoning process and allows for both specific and global assessment of learners' diagnostic reasoning performance (Figure 2). The tool provides teachers with benchmarks and language that can be used during formative feedback discussions with learners.

Table 1 demonstrates the results of the pilot testing by 10 clinical faculty who evaluated the initial six characteristics of the ART. The faculty favorably rated the ART's relevance to their own assessment practices and provision of feedback to learners. The potential for learning about clinical reasoning through the tool was rated lower than the other metrics (median rating of 4 on a five-point scale). Faculty commented that some terms used in the ART such as problem representation and high-priority diagnoses were either unfamiliar or ambiguous and that they would benefit from training to properly guide their learners. A

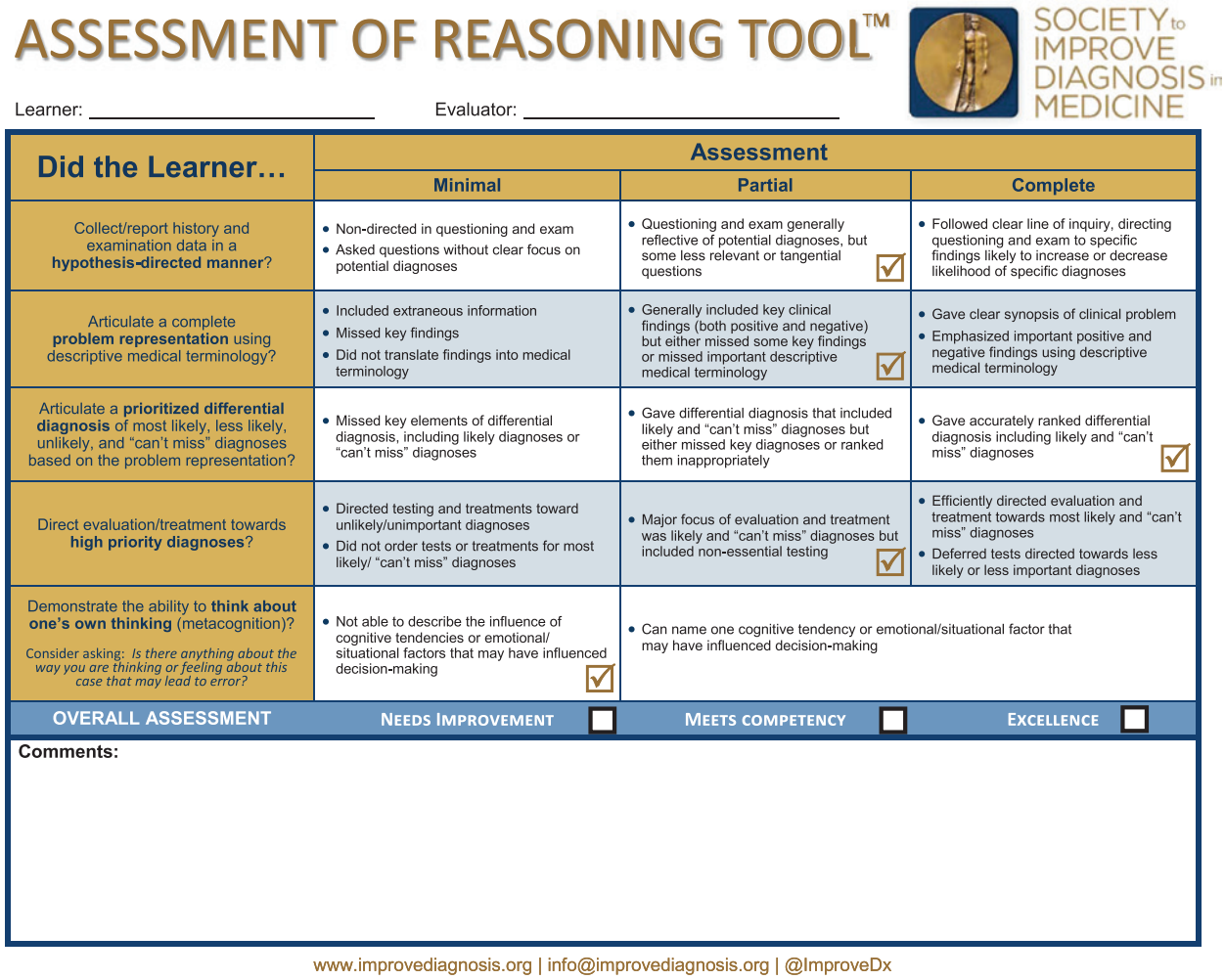

Figure 2: The Assessment of Reasoning Tool includes five domains of the clinical reasoning process and an overall assessment of the learner's performance.

Table 1: Faculty ratings of six ART characteristics using five-point Likert scale.

Characteristic	Median rating
1. Relevant to my practice	5
2. Covers important domains of clinical reasoning	4.5
3. Provides structure for assessment	4.5
4. Provides structure for performance feedback	5
5. Easy to use	4.5
6. Promote learning about clinical reasoning	4.0

1, Poor; 5, very good. n = 10.

common suggestion was to establish a process for faculty training using practice cases so that faculty could be better prepared to teach the components of clinical reasoning to their learners and accurately assess their presentations. Surveyed faculty also mentioned that clarity regarding which level of performance reflected competence would be helpful for summative assessments.

The faculty training module comprises five short (3–4 min) whiteboard animation videos (available at www.improvediagnosis.org/art). Each video highlights one domain of the ART, describes the rationale for the domain, defines clinical reasoning terms and gives examples of high- and low-performing learners. Ten clinical faculty rated the faculty training module favorably. They rated the all five videos as useful, educational and enjoyable (medians between 4 and 5 of five-point scale) and anticipated that the videos would provide effective guidance to faculty using the ART (median 4).

Discussion

Clinical reasoning has been called the "Holy Grail" of assessment in the health professions due to its great importance juxtaposed against the lack of a gold-standard

method of measurement [2]. Through a collaborative effort of multi-specialty experts, we created a theory-informed tool for the assessment of clinical reasoning along with faculty development videos. The ART provides an explicit structure for assessment of clinical reasoning and shared terminology to facilitate formative feedback.

Multiple tools have been developed to assess clinical reasoning. The methods often capture particular components or sub-tasks of the clinical reasoning process [4]. Few of the published general assessment tools that include clinical reasoning specify multiple domains of competence within the clinical reasoning process [12]. For example, the mini-Clinical Evaluation Exercise (mini-CEX) asks clinicians to rate "clinical judgment" on a nine-point scale [13]. The Integrated Direct Observation Clinical Encounter Examination (IDOCEE) tool includes three domains (data gathering, reasoning and analysis, decision-making) which address clinical reasoning but does not deconstruct it. The Problem Representation, Background Evidence, Analysis, Recommendation (PBEAR) tool [14] includes a comprehensive list of concrete tasks or actions within history taking and physical examination, which provides useful guidance for communication during an oral case presentation. Though the PBEAR construct implies several theoretical principles in clinical reasoning, few terms or concepts from the clinical reasoning literature are used explicitly.

The Interpretive summary, Differential diagnosis, Explanation of reasoning, Alternatives (IDEA) tool [8] developed for assessing medical students' clinical documentation includes important elements of clinical reasoning along with concise descriptors of those elements, but omits several elements of the clinical reasoning process featured in contemporary literature (e.g. hypothesis-directed data gathering). The Reading Hospital mini-CEX rating instrument assesses some domains of clinical reasoning (e.g. data collection, interpretation and synthesis of data, diagnostics and therapeutic reasoning). However, the tool also covers other competencies (e.g. interpersonal/communication skills, professional conduct and system-based practice) [10]. Furthermore, each domain combines multiple behaviors encompassing knowledge, skills and attitudes, so teachers must synthesize several aspects into one assessment rating. Most aligned with our proposed conceptual framework is a diagnostic skill evaluation form developed by Haight and DePriest [11] that uses milestone language to create clinical reasoning benchmarks within domains of competence.

The ART is distinguished from these tools by providing detailed behavioral anchors using contemporary clinical reasoning terminology to guide assessor observations or judgments. The ART is grounded in a theoretical framework integrated with a contemporary practice goal (i.e. HVC) and proposed strategies for reduction of errors (e.g. metacognition).

The ART is also distinguished from other assessment tools by the accompanying faculty development modules that facilitate implementation. Teachers can use these videos to familiarize themselves with core concepts of clinical reasoning and prepare for an upcoming ART assessment with a learner. Clinical reasoning terms represent a shared language for teachers and learners to dissect how they think about and learn from clinical problems [36]. Providing specific feedback to learners can be difficult if the teacher is not versed in vocabulary that can be used to describe specific behaviors within the clinical reasoning process. For example, some learners may be unable to give a clear synopsis of the clinical problem (e.g. merely restating facts and findings). Asking the learners to synthesize a case without providing an explicit guidance may leave some learners wondering how they can improve. A teacher can use the ART to assess the oral case presentation and then guide learners in applying the concepts and terms pertaining to a problem representation to structure the case synthesis. Teachers can also refer learners to the accompanied training module to expand their conceptual understanding about the concepts. Teachers can use these videos for either asynchronous or synchronous learning in a faculty development program to enhance teaching and learning about the diagnostic reasoning process. Teachers can also share these videos with learners in advance of or following an ART-guided conversation.

The ART has limitations. The conceptual framework derived from our interpretation and synthesis of literature to guide the construct of the ART reflects one model of clinical reasoning. By its nature, a conceptual framework emphasizes certain aspects and inherently disregards or dismisses other aspects of the subject being studied [37]. During the development process, our expert consultants advised us to capture important elements but not all aspects of the clinical reasoning process. Thus the ART captures only parts of the complex and broad construct of diagnostic reasoning process. Summative assessment of diagnostic reasoning skills requires a combination of non-workplace-based and workplace-based longitudinal assessments across multiple chief complaints and diseases in diverse clinical settings with numerous raters. The ART was not developed to displace well-established summative diagnostic reasoning assessment methods (e.g. multiple choice question, objective structured clinical examination). Rather, it complements them in the formative domain because it promotes explicit teaching

of the diagnostic reasoning process. Another invaluable suggestion from experts was to develop a simple tool for a broad range of users and then spend an equal amount of time developing the faculty training module. The training videos introduce clinical reasoning concepts but are not a comprehensive overview of the field. Among the specific behavioral descriptors in the ART, some are difficult to assess objectively. Hypothesis-directed data gathering in the first domain, for instance, is a partially subconscious process that can only be inferred but not directly assessed, even with real-time observation at the bedside. This can be improved by direct questioning of the learner. The fifth domain addresses metacognition with a simplified approach. Metacognitive abilities cannot be easily assessed or taught by simply asking "What else could this be?" or "Have you reflected about this case?". Teachers can use the ART to prompt a conversation with learners around cognitive tendencies and emotional/affective factors and to raise awareness of how these factors may have influenced their thinking.

The collective evidence of validity of the ART was embedded in the development process. We ascertained evidence of validity pertaining to construct and content through a systematic approach and integration of theories, expert opinion, consensus and comments from the end-users. By piloting the tool with clinician-educators we demonstrated some response process validity. A validation study is underway to gather further evidence about the response process, internal structure and relationships to other variables.

Conclusions

Through a collaborative effort of multi-specialty experts, we created a theory-informed assessment tool of clinical reasoning along with faculty training videos. The ART provides behavioral descriptors which serve as standards for assessment and benchmark practices that can be used by evaluators and learners for effective performance feedback.

Acknowledgments: The authors thank the following members of the SIDM Education Committee Assessment subcommittee for their many contributions to this project: Ethan Fried, MD, William Follansbee, MD, Frank Papa, DO, PhD, and Brent Smith, MD.

Author contributions: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.

Research funding: The creation of the faculty development videos was supported by a grant from the American Board of Internal Medicine Foundation.

Employment or leadership: None declared.

Honorarium: None declared.

Disclosure statement: Dr. Dhaliwal reports receiving lecture fees from ISMIE Mutual Insurance Company, Physicians' Reciprocal Insurers, & GE Healthcare.

Competing interests: The funding organization(s) played no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the report for publication.

References

1. Committee on Diagnostic Error in Health Care; Board on Health Care Services; Institute of Medicine; The National Academies of Sciences, Engineering, and Medicine; Balogh EP, Miller BT, Ball JR, editors. Improving diagnosis in health care. Washington, DC: National Academies Press (US), 2015 Dec 29:9, The Path to Improve Diagnosis and Reduce Diagnostic Error. Available from: https://www.ncbi.nlm.nih.gov/books/NBK338589/.
2. Schuwirth L. Is assessment of clinical reasoning still the holy grail? Med Educ 2009;43:298–300.
3. Durning S, Artino AR, Pangaro L, van der Vleuten CP, Schuwirth L. Context and clinical reasoning: understanding the perspective of the expert's voice. Med Educ 2011;45:927–38.
4. Gruppen L. Clinical reasoning: defining it, teaching it, assessing it, studying it. West J Emerg Med 2017;18:4–7.
5. Kogan JR, Hess BJ, Conforti LN, Holmboe ES. What drives faculty ratings of residents' clinical skills? The impact of faculty's own clinical skills. Acad Med 2010;85(10 Suppl):S25–8.
6. Ilgen JS, Humbert AJ, Kuhn G, Hansen ML, Norman GR, Eva KW, et al. Assessing diagnostic reasoning: a consensus statement summarizing theory, practice, and future needs. Acad Emerg Med 2012;19:1454–61.
7. Chamberland M, Mamede S, St-Onge C, Setrakian J, Bergeron L, Schmidt H. Self-explanation in learning clinical reasoning: the added value of examples and prompts. Med Educ 2015;49:193–202.
8. van der Vleten CP, Norman GR, Schuwirth L. Assessing clinical reasoning. In: Higgs J, Jones M, Loftus S, Christensen N, editors. Clinical reasoning in the health professions. Edinburgh: Elsevier, Churchill Livingstone, 2008:21–55. P413–422.
9. Baker EA, Ledford CH, Fogg L, Way DP, Park YS. The IDEA assessment tool: assessing the reporting, diagnostic reasoning, and decision-making skills demonstrated in medical students' hospital admission notes. Teach Learn Med 2015;27:163–73.
10. Donato AA, Pangaro L, Smith C, Rencic J, Diaz Y, Mensinger J, et al. Evaluation of a novel assessment form for observing medical residents: a randomised, controlled trial. Med Educ 2008;42:1234–42.
11. Haight M, DePriest JL. Using the internal medicine milestones to teach and assess resident clinical diagnostic reasoning skills. Academic Internal Medicine Insight 2013;11:14–5. Available from: https://www.im.org.

12. Kogan JR, Holmboe ES, Hauer, KE. Tools for direct observation and assessment of clinical skills of medical trainees: a systematic review. J Am Med Assoc 2009;302:1316–26.
13. Norcini JJ, Blank LL, Duffy FD, Fortna GS. The mini-CEX: a method for assessing clinical skills. Ann Intern Med 2003;138:476–81.
14. Carter C, Akar-Ghibril N, Sestokas J, Dixon G, Bradford W, Ottolini M. Problem representation, background evidence, analysis, recommendation: an oral case presentation tool to promote diagnostic reasoning. Acad Pediatr 2018;18:228–30.
15. DeVellis RF. Scale development: theory and applications, 2nd ed. Thousand Oaks, CA: Sage, 2003.
16. Nendaz MR, Bordage G. Promoting diagnostic problem representation. Med Educ 2002;36:760–6.
17. Chang R, Bordage G, Connell K. The importance of early problem representation during case presentations. Acad Med 1998;73(Suppl):109–11.
18. Bordage G, Lemieux M. Semantic structures and diagnostic thinking of experts and novices. Acad Med 1991;66(9 Suppl):S70–2.
19. Bordage G. Elaborated knowledge: a key to successful diagnostic thinking. Acad Med 1994;69:883–5.
20. Bordage, G. Prototypes and semantic qualifiers: from past to present. Med Educ 2007;41:1117–21.
21. Bordage G, Connell KJ, Chang RW, Gecht MR, Sinacore JM. Assessing the semantic content of clinical case presentations: studies of reliability and concurrent validity. Acad Med 1997;72(10 suppl):S37–9.
22. Bowen JL. Educational strategies to promote clinical diagnostic reasoning. N Engl J Med 2006;355:2217–25.
23. Schmidt HG, Norman GR, Boshuizen HP. A cognitive perspective on medical expertise: theory and implications. Acad Med 1990;65:611–21.
24. Audétat MC, Laurin S, Dory V, Charlin B, Nendaz MR. Diagnosis and management of clinical reasoning difficulties: part I. Clinical reasoning supervision and educational diagnosis. Med Teach 2017;39:792–6.
25. Charlin B, Tardif J, Boshuizen HP. Scripts and medical diagnostic knowledge: theory and applications for clinical reasoning instruction and research. Acad Med 2000;75:182–90.
26. Charlin B, Boshuizen HP, Custers EJ, Feltovich PJ. Scripts and clinical reasoning. Med Educ 2007;41:1178–84.
27. Schmidt HG, Rikers RM. How expertise develops in medicine: knowledge encapsulation and illness script formation. Med Educ 2007;41:1133–9.
28. Croskerry P. Diagnostic failure: a cognitive and affective approach. In Advances in patient safety: from research to implementation. Rockville, MD: Agency for Healthcare Research and Quality (Publication No. 050021), 2005:2:241–54.
29. Croskerry P. A universal model of diagnostic reasoning. Acad Med 2009;84:1022–8.
30. Stammen LA, Stalmeijer RE, Paternotte E, Oudkerk Pool A, Driessen EW, Scheele F, et al. Training physicians to provide high-value, cost-conscious care: a systematic review. J Am Med Assoc 2015;314:2384–400.
31. Korenstein D. Charting the route to high value care: the role of medical education. J Am Mes Assoc 2015;314:2359–61.
32. Zwaan L, Monteiro S, Sherbino J, Ilgen J, Howey B, Norman G. Is bias in the eye of the beholder? A vignette study to assess recognition of cognitive biases in clinical case workups. BMJ Qual Saf 2017;26:104–10.
33. Downing S, Haladyna T. Validity and its threats. In: Downing S, Yudkowsky R, editors. Assessment in health professions education. New York, NY: Routledge, 2009:21–55.
34. Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. Am J Med 2006;119:166.e7–16.
35. Mayer RE. Applying the science of learning: evidence-based principles for the design of multimedia instruction. Am Psychol 2008;63:760–9.
36. Dhaliwal G, Ilgen JS. Clinical reasoning: talk the talk or just walk the walk? J Grad Med Educ 2016;8:274–6.
37. Bordage G. Conceptual frameworks to illuminate and magnify. Med Educ 2009;43:312–9.

Supplementary Material: The online version of this article offers supplementary material (https://doi.org/10.1515/dx-2018-0052).