Clinical reasoning is a core component of clinical competency that is used in all patient encounters from simple to complex presentations. It involves synthesis of myriad clinical and investigative data, to generate and prioritize an appropriate differential diagnosis and inform safe and targeted management plans. The literature is rich with proposed methods to teach this critical skill to trainees of all levels. Yet, ensuring that reasoning ability is appropriately assessed across the spectrum of knowledge acquisition to workplace-based clinical performance can be challenging. In this perspective, we first introduce the concepts of illness scripts and dual-process theory that describe the roles of analytic system 1 and non-analytic system 2 reasoning in clinical decision making. Thereafter, we draw upon existing evidence and expert opinion to review a range of methods that allow for effective assessment of clinical reasoning, contextualized within Miller’s pyramid of learner assessment. Key assessment strategies that allow teachers to evaluate their learners’ clinical reasoning ability are described from the level of knowledge acquisition, through to real-world demonstration in the clinical workplace.

KEY WORDS: medical education-assessment method; medical education-assessment/evaluation; medical education-cognition/problem solving; clinical reasoning.

J Gen Intern Med 34(8):1631–6
DOI: 10.1007/s11606-019-04953-4
© The Author(s) 2019

INTRODUCTION

Clinical reasoning is the “thinking and decision making processes associated with clinical practice.” It involves pattern recognition, knowledge application, intuition, and probabilities. It is integral to clinical competency and is gaining increasing attention within medical education. Much has been written about the cognitive processes underpinning reasoning strategies and the myriad ways educators can enhance learners’ clinical reasoning abilities. Literature on assessing clinical reasoning, however, is more limited with focus on written assessments targeting the lower levels of “knows” and “knows how” of Miller’s pyramid (Fig. 1). This article offers a more holistic perspective on assessing clinical reasoning by exploring current thinking and strategies at all levels.

CLINICAL REASONING MODELS

Although many clinical reasoning models have been proposed, script theory and dual-process theory have attracted particular attention among medical educators. Script theory suggests clinicians generate and store mental representations of symptoms and findings of a particular condition (“illness scripts”) with networks created between existing and newly learnt scripts. Linked to this is dual-process theory which suggests that clinical decision making operates within two systems of thinking. System 1 thinking utilizes pattern recognition, intuition, and experience to effortlessly activate illness scripts to quickly arrive at a diagnosis. Conversely, clinicians utilizing system 2 thinking analytically and systematically compare and contrast illness scripts in light of emerging data elicited from history and examination while factoring in demographic characteristics, comorbidity, and epidemiologic data. In this approach, clinicians test probable hypotheses, using additional information to confirm or refute differential diagnoses. Although system 2 thinking requires more cognitive effort, it is less prone to the biases inherent within system 1 thinking. Several cognitive biases have been characterized with key examples outlined in Table 1. With increasing experience, clinicians skillfully gather and apply relevant data by continually shifting between non-analytic and analytic thinking.

Attempting to assess complex internal cognitive processes that are not directly observable poses obvious challenges. Furthermore, it cannot be assumed that achieving correct final outcomes reflects sound underpinning reasoning. Potential strategies however have been suggested to address these difficulties and are described below at each level of Miller’s pyramid.

ASSESSING CLINICAL REASONING AT THE “KNOWS” AND “KNOWS HOW” LEVELS

In the 1970s, patient management problems (PMPs) were popular and utilized a patient vignette from which
candidates selected management decisions. Originally designed to assess problem-solving strategies, later work suggested PMPs were likely only testing knowledge acquisition. Often, there was disagreement among experts on the possible correct answer along with poor case specificity (performance on one case poorly predicting performance on another). Furthermore, experienced clinicians did not always score higher than juniors. As a result, the use of PMPs has declined.

Subsequently, script concordance tests (SCTs) were developed based on the previously mentioned concept of “illness script.” An example SCT is shown in Text box 1. Examinees are faced with a series of patient scenarios and decide, using a Likert-type scale, whether a particular item (such as a symptom, test, or result) would make a diagnosis more or less likely. Examinees’ answers are compared with those from experts, with weighted scoring applied to responses chosen by more expert clinicians. SCTs offer reliable assessments (achieving alpha of 0.77–0.82) with agreement from both examiners and candidates that real-world diagnostic thinking is being assessed. They predict performance on other assessments (such as Short Answer Management Problems and Simulated Office Orals), allow for discrimination across the spectrum of candidate ability, and show improved performance with increasing experience (construct validity).

**Text Box 1. Example SCT**

A 55-year-old man presents to your clinic with a persistent cough of 6 weeks.

| If you were thinking of: | And then you find: | This diagnosis becomes |
|--------------------------|--------------------|------------------------|
| Q1: Lung cancer          | Patient has smoked 20 cigarettes a day for 30 years | – 2 + 0 + 1 + 2 |
| Q2: Drug side effect     | Patient started ace inhibitor 6 weeks ago          | – 2 + 0 + 1 + 2 |
| Q3: COPD                 | Patient has never smoked                            | – 2 + 0 + 1 + 2 |

– 2 Ruled out or almost ruled out; – 1 Less likely; 0 Neither more nor less likely; + 1 More likely; + 2 Certain or almost certain

**Key feature questions** (KFQs) require candidates to identify essential elements within a clinical vignette in relation to possible diagnoses, investigations, or management options. In keeping with SCTs, KFQs similarly demonstrate good face validity, construct validity, and predictive validity of future performance. In addition, KFQs are thought to have an advantage over SCTs as they can minimize the cueing effect within the latter’s response format.

The clinical integrative puzzle (CIP) bases itself on the extended matching question concept but utilizes a grid-like appearance which requires learners to compare and contrast a group of related diagnoses across domains such as history, physical examination, pathology, investigations, and management. CIPs encourage integration of learning and consolidation of illness scripts and demonstrate good reliability (up to 0.82) but only modest validity.
The ASCLIRE method uses computer-delivered patient scenarios that allow learners to seek additional data from a range of diagnostic measures in order to select a final diagnosis from a differential list. Diagnostic accuracy, decision time, and choice of additional diagnostic data are used to differentiate reasoning abilities. It is estimated that 15 scenarios, over 180 min, would achieve a reliability of 0.7. ASCLIRE is well received by candidates and demonstrates appropriate construct validity, with experts outscoring novices.

More recently, virtual patients have been developed with software enabling students to categorize diagnoses as unlikely or not to be missed through illness script-based concept maps. Although currently proposed for use as a learning tool, future development could offer assessment possibilities.

### Assessing Clinical Reasoning at the “Shows How” Level

Objective structured clinical examinations (OSCEs) are widely accepted as robust assessments of learners’ clinical competencies. From the first papers describing OSCEs in the 1970s to the multitude of publications since, their ability to assess a range of clinical skills including problem-solving abilities has been emphasized. Despite this stated aim, the literature however is limited to components of clinical competency such as history taking, physical examination, or explanation of diagnoses, with less attention paid to understanding how OSCEs can be used to assess clinical reasoning ability.

Given the paucity of published work in this area, assessment and teaching academics from the lead authors’ institution have worked collaboratively to transform historically used OSCE stations, which often operated on simple pattern recognition, into stations that require analytical system 2 thinking.

![Table 2 Suggested OSCE design strategies](image)

**Table 2 Suggested OSCE design strategies**

| Station type              | Design strategies                                                                 |
|---------------------------|----------------------------------------------------------------------------------|
| History taking stations   | Create simulated patient scripts such that not all possible symptoms are present and/or add in symptoms that may suggest more than one plausible differential diagnosis. |
| Physical examination stations | Design hypothesis-driven or presentation-based examinations (requiring the candidate to conduct an appropriate examination from a stated differential list or short clinical vignette) rather than full system-based examinations. |
| Data interpretation stations | Utilize clinical data (either at once or sequentially) along with a clinical vignette and examiner questions to assess not just the candidate’s data reporting ability, but interpretation ability in light of the clinical situation described. |
| Explanation stations      | Provide clinical results/data requiring candidate interpretation and offering real-world clinical context to base explanations and justifications. |

Lastly, how OSCE stations are scored may impact candidates’ demonstration of reasoning ability. Checklist-based rubrics often trivialize the complexity of patient encounters and thus may discourage the use of analytical system 2 approaches. Conversely, rating scales that assess component parts of performance (analytic) or overall performance (global) offer improved reliability and validity in capturing a more holistic perspective on candidates’ overall performance. However, whether analytic or global rating scales differ in their assessment of clinical reasoning remains unclear.

Scale issues aside, the challenge for OSCE examiners remains trying to score, through candidate observation, the internal cognitive process of clinical reasoning. Recent work however has provided guidance, suggesting that there are certain observable behaviors demonstrated by candidates which reflect their reasoning processes as shown in Table 3.

### Assessing Clinical Reasoning at the “Does” Level

Since clinical reasoning proficiency in knowledge tests or simulated settings does not automatically transfer to real-life clinical settings, it is critical to continue assessment in the workplace, thereby targeting the top of Miller’s pyramid at the “does” level. Clinical teachers should assess how learners tackle uncertainty and detect when the reasoning process is derailed by limitations in knowledge or experience, cognitive biases, or inappropriate application of analytic and non-analytic thinking. For example, if novice learners demonstrate non-analytic thinking, clinical teachers should question...
their reasons for prioritizing certain diagnoses over others. However, advanced learners can apply non-analytic thinking to simpler clinical scenarios. Experts demonstrate higher diagnostic accuracy rates and lower decision making time than novices and skillfully utilize both non-analytic and analytic thinking. Therefore, learners will benefit when expert clinical teachers think out loud as they develop diagnostic hypotheses.

Clinical teachers routinely observe learners to assess their clinical skills; however, such assessment is often informal, in the moment and impression-based rather than systematic. Moreover, it is often the end point that is assessed rather than the process of reasoning. While summative assessment can determine whether learners have achieved expected competencies, formative assessment fosters a climate of assessment for learning. Frameworks that allow for formative systematic assessment of clinical reasoning are therefore valuable and exemplars are described below.

Bowen’s framework lists sequential steps that can be demonstrated and assessed by clinical teachers as shown in Figure 2. These include data gathering (history, examination

Table 3 Observable Behaviors of Clinical Reasoning During Patient Interactions

| Level | Student acts                                                                 | Patient acts                  | Course of the conversation | Data gathered and efficiency |
|-------|-------------------------------------------------------------------------------|-------------------------------|----------------------------|-----------------------------|
| 1     | Taking the lead in the conversation  |
|       | Recognizing and responding to relevant information  |
|       | Specifying symptoms  |
|       | Asking specific questions pointing to pathophysiological thinking  |
|       | Putting questions in a logical order  |
|       | Checking with patients  |
|       | Summarizing  |
|       | Body language  |
| 2     | Patient body language, expressions of understanding or confusion  |
| 3     | Students and patients talking at cross purposes, repetition  |
| 4     | Quantity and quality of data gathered  |
|       | Speed of data gathering  |

Based on Haring et al.31

Figure 2 Steps and strategies for clinical reasoning (supplied also as a .tif file). Adapted from Bowen JL. Educational strategies to promote clinical diagnostic reasoning. N Engl J Med. 2006;355(21):2217–25.
findings, results of investigations), summarizing key features of the case (problem representation), generating differential diagnoses (diagnostic hypothesis), applying prior knowledge (illness scripts), and final diagnosis. The assessment of reasoning tool similarly describes a five-component assessment process (hypothosis-directed data collection, problem representation, prioritized differential diagnosis, high-value testing, and metacognition) with a simple scoring matrix that rates the learner’s degree of mastery on each component. The IDEA framework moves beyond observed assessments and instead evaluates the clinician’s written documentation for evidence of reasoning across four elements: interpretive summary, differential diagnosis, explanation of reasoning, and alternative diagnoses considered.

Finally, the one-minute preceptor model offers a simple and time-efficient framework for formative assessment of clinical reasoning during short case presentations. Learners should develop the skills to synthesize all clinical clues from the history and physical examination, generate an appropriate differential diagnosis, and commit to the most likely diagnosis. Teachers can then pose questions exploring their learners’ skills in diagnostic hypothesis generation, investigative intent, and management planning, seeking their justifications for each. “Getting learners to make a commitment” requires “what” questions and “probing for supportive evidence” requires “how” or “why” questions. “Teaching general rules” assesses how well learners can compare and contrast similar presentations for different patients. Assessment is only meaningful when learners receive ongoing feedback on accuracy of their diagnostic reasoning processes and errors resulting from inappropriate use of non-analytic reasoning and this is achieved through “tell them what they did right” and “correct errors gently”. These steps are depicted in Table 4 in relation to corresponding steps of Bowen’s model with potential methods of assessment for each stage.

### CONCLUSIONS

This article describes a range of clinical reasoning assessment methods that clinical teachers can use across all four levels of Miller’s pyramid. Although this article has not focused on strategies to help address identified deficiencies in reasoning ability, other authors have developed helpful guidelines in this regard. As with all areas of assessment, no one level or assessment tool should take precedence and clinical teachers should be prepared and trained to assess from knowledge through to performance using multiple methods to gain a more accurate picture of their learners’ skills. The challenge of case specificity also requires teachers to repeatedly sample and test reasoning ability across different clinical contexts.

Clinical reasoning is a core skill that learners must master to develop accurate diagnostic hypotheses and provide high-quality patient care. It requires a strong knowledge base to allow learners to build illness scripts which can help expedite diagnostic hypothesis generation. As it is a critical step that synthesizes disparate data from history, physical examination, and investigations into a coherent and cogent clinical story, teachers cannot assume that their learners are applying sound
Corresponding Author: Harish Thampy, Division of Medical Education, School of Medical Sciences, Faculty of Biology, Medicine & Health University of Manchester, Manchester, UK (e-mail: harish.thampy@manchester.ac.uk).

Compliance with Ethical Standards:
Conflict of Interest: The authors declare that they do not have a conflict of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

REFERENCES

1. Higgs J, Jones M, Loftus S, Christensen N. Clinical Reasoning in the Health Professions. 3rd Edition. The Journal of Chiropractic Education. 2008;2(2):161–2.
2. Miller GE. The assessment of clinical skills/competence/performance. Acad Med. 1990;65(9 Suppl):S63–7.
3. Charlin B, Tardif J, Boshuizen HP. Scripts and medical diagnostic knowledge: theory and applications for clinical reasoning instruction and research. Acad Med. 2000;75(2):182–90.
4. Crossley P. A universal model of diagnostic reasoning. Acad Med. 2009;84(8):1022–8.
5. Crossley P. From mindless to mindful practice-cognitive bias and clinical decision making. N Engl J Med. 2013;368(26):2445–8.
6. Crosskey P. The importance of cognitive errors in diagnosis and strategies to minimize them. Acad Med. 2003;78(8):775–80.
7. Eva KW. What every teacher needs to know about clinical reasoning. Med Educ. 2005;39(1):98–106.
8. Harden RM. Preparation and presentation of patient-management problems (PMPs). Med Educ. 1983;17(4):256–76.
9. Norcini JJ, Swanson DB, Grosso LJ, Weber GD. Reliability, validity and efficiency of multiple choice question and patient management problem item formats in assessment of clinical competence. Med Educ. 1985;19(3):238–47.
10. Newble DI, Hoare J, Baxter A. Patient management problems. Issues of validity. Med Educ. 1982;16(3):137–42.
11. Charlin B, van der Vleuten C. Standardized assessment of reasoning in contexts of uncertainty: the script concordance approach. Eval Health Prof. 2004;27(3):304–19.
12. Carrière B, Gagnon R, Charlin B, Downing S, Bordage G. Assessing clinical reasoning in pediatric emergency medicine: validity evidence for a Script Concordance Test. Ann Emerg Med. 2009;53(9):647–52.
13. Pievezan RD, Custodio O, Cendoroglio MS, Batista NA, Labarsky S, Charlin B. Assessment of undergraduate clinical reasoning in geriatric medicine: application of a script concordance test. J Am Geriatr Soc. 2012;60(10):1946–50.
14. Charlin B, Roy L, Brailovsky C, Goulet F, van der Vleuten C. The Script Concordance test: a tool to assess the reflective clinician. Teach Learn Med. 2000;12(4):189–95.
15. Brailovsky C, Charlin B, Beausoleil S, Cote S, van der Vleuten C. Measurement of clinical reflective capacity early in training as a predictor of clinical reasoning performance at the end of residency: an experimental study on the script concordance test. Med Educ. 2001;35(5):430–6.
16. Hrychnak P, Takahashi SG, Nayer M. Key-feature questions for assessment of clinical reasoning: a literature review. Med Educ. 2014;48(9):870–83.
17. Wass V, Van der Vleuten C, Shatzer J, Jones R. Assessment of clinical competence. Lancet. 2001;357(9269):945–9.
18. Ber R. The CIP (comprehensive integrative puzzle) assessment method. Med Teach. 2003;25(2):171–6.
19. Capaldi VF, Durning SJ, Pangaro LN, Ber R. The clinical integrative puzzle for teaching and assessing clinical reasoning: preliminary feasibility, reliability, and validity evidence. MJ Med. 2015;180(4 Suppl):S56–60.
20. Kumina-Habenicht O, Hautz WE, Knigge M, Spies C, Ahlers O. Assessing clinical reasoning (ASCLIRE): Instrument development and validation. Advances in health sciences education : theory and practice. 2015;20(5):1205–24.
21. Hege I, Kononowicz AA, Adler M. A Clinical Reasoning Tool for Virtual Patients: Design-Based Research Study. JMIR medical education. 2017;3(2):e21.
22. Harden RM, Stevenson M, Downie WW, Wilson GM. Assessment of clinical competence using objective structured examination. Br Med J. 1975;1(5955):447–51.
23. Yudkowsky R, Otaki J, Lowenstein T, Riddle J, Nishigori H, Bordage G. A hypothesis-driven physical examination learning and assessment procedure for medical students: initial validity evidence. Med Educ. 2009;43(8):729–40.
24. Durning SJ, Artino A, Boulet J, La Rochelle J, Van der Vleuten C, Arze B, et al. The feasibility, reliability, and validity of a post-encounter form for evaluating clinical reasoning. Med Teach. 2012;34(11):30–7.
25. Berger AJ, Gillespie CC, Tewkbury LR, Overstreet IM, Tsai MC, Kalet A. Validation of a tool to assess medical student clinical reasoning by “lay” vs physician raters: inter-rater reliability using a scoring guide in a multidisciplinary objective structured clinical examination. Am J Surg. 2012;203(1):81–6.
26. Myung SJ, Kang SH, Phylo SR, Shin JS, Park WB. Effect of enhanced analytic reasoning on diagnostic accuracy: A randomized controlled study. Med Teach. 2013;35(3):248–50.
27. Park WB, Kang SH, Lee YS, Myung SJ. Does Objective Structured Clinical Examinations Score Reflect the Clinical Reasoning Ability of Medical Students? The American Journal of the Medical Sciences. 2015;350(1):64–7.
28. Yudkowsky R, Park YS, Hyderi A, Bordage G. Characteristics and Implications of Diagnostic Justification Scores Based on the New Patient Note Format of the USMLE Step 2 CS Exam. Acad Med. 2015;90(11 Suppl):S556–62.
29. Hodges B, McIroy JJ. Analytic global OSCE ratings are sensitive to level of training. Med Educ. 2003;37(11):1012–6.
30. Norman GR, Brooks LR, Allen SW. Recall by expert medical practitioners and novices as a record of processing attention. J Exp Psychol Learn Mem Cogn. 1989;15(6):1166–74.
31. Haring CM, Cools BM, van Gurp PJM, van der Meer JWM, Postma CT. Observable phenomena that reveal medical students’ clinical reasoning ability during expert assessment of their history taking: a qualitative study. BMC Med Educ. 2017;17(1):147.
32. Ericsson KA. The Cambridge handbook of expertise and expert performance. 2nd edition. Cambridge: Cambridge University Press; 2018. 969.
33. Bowen JL. Educational strategies to promote clinical diagnostic reasoning. N Engl J Med. 2006;355(21):2217–25.
34. Society to Improve Diagnosis in Medicine. Assessment of Reasoning Tool [Available from: https://www.improvediagnosis.org/page/art. Accessed 1 Nov 2018.
35. 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(2):163–73.
36. Neher JO, Stevens NG. The one-minute preceptor: shaping the teaching conversation. Fam Med. 2002;34(8):589–3.
37. Neher JO, Gordon KC, Meyer B, Stevens N. A five-step “microskills” model of clinical teaching. J Am Board Fam Pract. 1992;5(4):419–24.
38. Audétat M-C, Laurin S, Sanche G, Bélique C, Fon NC, Blais J-G, et al. Clinical reasoning difficulties: A taxonomy for clinical teachers. Med Teach. 2013;35(3):e884–e9.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.