Chapter 5
Approaches to Assessing the Clinical Reasoning of Preclinical Students

Olle ten Cate and Steven J. Durning

If clinical reasoning is considered critical for any physician and an ability a student should acquire during undergraduate medical education, then educators should attempt to assess whether students satisfactorily meet this objective.

In earlier chapters we have establish that clinical reasoning has two components: analytic reasoning and nonanalytic reasoning (i.e., pattern recognition). Hence these two may be the focus of assessment: (1) Do students understand physiology and the pathophysiologic mechanisms and enabling conditions that lead to disease and consequently recognize signs and symptoms observable in patients? and (2) Do students build a mental repository of illness scripts that allow them to recognize patterns in the patients they encounter?

Clearly these objectives require substantial medical knowledge and substantial experience in patient care. And if clinical reasoning by definition, as some say, must include the context in which the physicians works (Woods and Mylopoulos 2015), how reasonable is it to test preclinical student on their clinical reasoning ability? According to Bowen and Ilgen, diagnostic reasoning is not a discrete, enduring, or reliably measurable skill. Accurate measurement in fact requires an observer to interpret processes that are heavily context dependent, usually not explicitly articulated, and often occur below conscious awareness of the observed clinician (Bowen and Ilgen 2014). Nevertheless, authors have attempted to infer progress in clinical reasoning ability across years using a written progress test (Williams et al. 2011).

O. ten Cate (✉)
Center for Research and Development of Education, University Medical Center Utrecht, Utrecht, The Netherlands
e-mail: T.J.tenCate@umcutrecht.nl

S.J. Durning
Uniformed Services University of the Health Sciences, Bethesda, MD, USA
e-mail: steven.durning@usuhs.edu

Case-based clinical reasoning education, or any other approach recommended for preclinical education, attempts to prepare students for clinical encounters. While assessing clinical reasoning *in context* may not be reasonable for these students, a more limited approach, using written test approaches, is possible. Analytic reasoning is practiced in basic science or integrated courses, and pattern recognition ability may already be acquired on a very basic level. The CBCR course, as described in Part II of this book, has the deliberate intention to help students build a limited illness script mental repository for a number of common medical conditions including the differential diagnosis of adjacent conditions. This can be the focus of a test.

Without mentioning the word validity, these introductory sentences pertain to validity. The validity of educational and psychological tests has been reconceptualized in the past decades by scholars such as Messick and Kane (Cook et al. 2015). The validity of a test should be argued from the perspective of the content, response process, internal structure, relationship to other variables, and consequences of the test (AERA/APA/NCME 2014; Downing 2003). For clinical reasoning in preclinical students, the consequences should be the readiness to encounter patients in the clinical setting. The content should focus on important knowledge to allow analytic reasoning as can be expected in such encounters and for the recognition of patterns they have encountered in preclinical education. Response processes, or the way questions in such tests are asked, should resemble the clinical thinking pathways that happen in such encounters and the relationship to other variables may be a hindsight evaluation whether students with a high score indeed seem to do well in clinical reasoning in practice. While we have stressed the limitations in clinical reasoning that must be faced in the preclinical period, it is important to simulate situations they will face once they assume patient-related clinical tasks. As assessment is a powerful stimulus for learning, tests should be designed in such a way that students spend their energy optimally in anticipation of clinical encounters.

Current Methods of Assessing Clinical Reasoning

Educators looking for methods to assess clinical reasoning will find most recommended approaches to be used in clinical education, such as at the bedside, and only few focusing on the testing of reasoning in the preclinical phase, e.g., in a written test format. In terms of Miller's four-level pyramid of assessment in medical education (knows – knows how – shows how – does), the highest three are all to some extent suitable for the assessment of clinical reasoning (Miller 1990). A "knows how" test would present a patient case and asks the candidate to arrive at a diagnosis and/or a therapy. During a "shows how" test, an examiner would ask the student to clinically reason in a standardized patient encounter such as during an objective structured clinical examination (OSCE), and an assessment at the "does" level would ask a student to reason related to a real patient case in the hospital. Table 5.1 summarizes some frequently used, or specifically designed, methods to assess clinical

Table 5.1 Approaches to the assessment of clinical reasoning from the literature

Miller level	Format	Specific methods	Selected references
Knows	[While knowledge is essential for clinical reasoning, factual knowledge tests per se are less suitable to assess clinical reasoning]		
Knows how	Written or electronic format	Constructed response methods	
		Short-answer open questions' test	Rademakers et al. (2005)
		Clinical reasoning problems' test	Groves et al. (2002)
		Written case summaries	Dory et al. (2016)
		Forced choice methods	
		Extended matching questions' test	Case and Swanson (1998)
		Script concordance test	Charlin et al. (2000)
		Comprehensive integrative puzzle test	Ber (2003)
		Case-based clinical reasoning test	See Chap. 7
Shows how	Standardized simulation format	Standardized patient station in an objective structured clinical examination (OSCE)	Sloane et al. (1995) and Hawkins and Boulet (2008)
		Patient assessment and management examination (PAME)	Macrae et al. (2000)
Does	Oral format	Chart-stimulated recall and case-based discussion (CSR/CBD)	Tekian and Yudkowsky (2007) and Singh and Norcini (2013)
		Standardized oral examination	Tekian and Yudkowsky (2007) and Norcini and Burch (2007)
		Mini clinical evaluation exercise	

reasoning with reference to Miller's Pyramid. In addition to this list, a specific test format has been developed for CBCR courses, which is discussed in Chap. 7.

For preclinical students, Miller's levels of *shows how* and *does* are less applicable. To assess students' clinical reasoning ability in students before they encounter patients, a written or electronic test format is more suitable for several reasons. Cohort of students can be tested at once, standards can be set, and reliable scores can be generated. One can argue that clinical reasoning should ideally measure actual performance. That would yield the best construct alignment between the goals and objectives, what is taught, and what is tested.

For CBCR courses with large numbers of students a written, or preferably an electronic, test format is recommended to establish a reliable examination. In a recent literature review on question types for clinical reasoning tests suitable for electronic tests, Van Bruggen and colleagues identified eight types (van Bruggen

et al. 2012): script concordance test questions, extended matching questions, comprehensive integrative puzzle questions, modified essay/short-answer questions, long-menu questions, multiple-choice questions, and true/false questions. The latter two were identified as least suitable, and we added two formats, all briefly discussed in Table 5.2. Features from different formats have been combined in the CBCR test format explained more extensively in Chap. 7.

Table 5.2 Questions suitable for written or electronic assessment of clinical reasoning ability

Question type	Item and tests' description	Features and comment
Script concordance test items (Charlin et al. 2000; Lubarsky et al. 2011)	A short patient vignette is given + a diagnostic hypothesis. Next, a new finding is presented. The candidate must score how this finding renders the hypothesis (much) less to (much) more likely, on a scale from −2 to +2, with score 0 being "no change"	Model answers are constructed using a panel of experts answering the questions. As they may disagree, a weighting is applied to scale values based on the number of experts choosing that value
		SCT is widely used but is also criticized for its validity and practicality (van den Broek et al. 2012; Lineberry et al. 2013)
Modified essay or short-answer questions (Rademakers et al. 2005)	Short-answer case-based questions that result in reliable tests have a short case vignette, require an answer of no more than 20 words (preferably much less), have predetermined model answers and scoring instructions to guide correction, and yield a scaled score (e.g., 0–3 points)	Experience learns that 40–50 questions should make a reliable test (ten Cate 1997)
		The major drawback of SAQs is that they require hand scoring, which may take time, specifically if there are many students
Clinical reasoning problems (Groves et al. 2002)	CRP questions contain a case vignette and ask for (a) a most likely diagnosis and (b) features from the vignette that support or oppose the hypothesis, each with a weighting (1–3), (c) an alternative diagnosis with (d) similar follow-up question as b	Groves et al. report satisfactory reliability and construct and external validity with a voluntary 10 CRP test, but without test conditions (Groves et al. 2002)
		The major drawback of CRPs is that they require hand scoring, which may take time, specifically if there are many students
Extended matching questions (Case and Swanson 1998)	EMQs have a theme (e.g., "fatigue"), a list of options (e.g., 10–20 diagnoses or lab results), a lead question ("what is the most likely diagnosis?" "which lab result do you expect?"), and then two or more case vignettes	Used by the National Board of Medical Examiners, EMQs are well known in the United States; less so outside the United States
		Number of EMQs and testing time required for a reliable test (up to 100 items and 4 hours) is quite large (Beullens et al. 2002)

Table 5.2 (continued)

Question type	Item and tests' description	Features and comment
Comprehensive integrative puzzle test CIP (Ber 2003)	One CIP is a table of 4*4–6*6 cells with in the first column a series of related (differential) diagnoses. Other columns are headed *history, physical examination, test results, X-ray, management*, or similar. Empty cells must be filled from separate option lists to construct, horizontally, logical illness scripts. The sum of correct cells yields a score	Four to five CIP cells may constitute a reliable test. Construct validity has been established (Groothoff et al. 2008)
		A potential drawback is the difficulty of item writing. A too narrow differential diagnosis column may make the construction of valid option lists hard; a too diverse differential diagnosis column may make CIP too easy
Long-menu questions (Schuwirth et al. 1996)	Long-menu questions are used in electronic testing as an alternative for open questions and have a very long list of options to eliminate guessing. Advanced formats match typed-in questions with the list to enable automatic scoring	A drawback is that more than one entry word is difficult to recognize automatically, and mistakes can be made if multiple words are required. In addition, the same drawbacks as with multiple-choice questions apply, without the cueing disadvantage
Written case summaries (Dory et al. 2016)	Candidates receive multiple case vignettes describing in lay language a patient's history of present illness, past medical history, and physical examination findings. They must summarize the case as they would present to an attending staff, in a few sentences using medical terminology (semantic qualifiers) to measure problem representation. Answers are scored using a 3-item rubric focusing on pertinent findings, semantic quality, and a global rating	This approach aligns well with Bowen's prerequisites for clinical reasoning (Chap. 4). The authors report "good evidence regarding scoring and generalizability" in a study with 8 case summary questions among 700 medical students, but acceptable reliability may require more cases. The method may be part of a battery of different items. Scoring time per rater is estimated 1 min per case, and rater training may be needed. Technology may assist the rating in the future

Almost all of the test forms in Table 5.2 use a key-feature approach. Key-feature questions focus on critical steps in the solution of a clinical problem and may pertain to aspects that learners generally find difficult or that are critical in patient management (Page et al. 1995). The development of the key-feature approach in the 1990s was a move away from the traditional assessment of clinical reasoning using a comprehensive examination of a patient management problem (Page and Bordage 1995). A recent review reconfirmed the generally favorable psychometric

properties of question types derived from the key-feature approach (Hrynchak et al. 2014).

In this chapter, we have provided a brief overview of current methods assessment of clinical reasoning, with a focus on methods suitable for preclinical students in a written fashion. We acknowledge this overview is limited. An excellent recent overview of more clinically oriented approaches was provided by Rencic and colleagues (2016). In addition, many studies have been conducted to measure clinical reasoning ability, and several of these have used experimental outcome measures that might be suitable for standard assessment at some time. Computer-based tests (Kunina-Habenicht et al. 2015), virtual reality assessment (Forsberg et al. 2016), eye-tracking (Kok and Jarodzka 2017), neuroimaging (Durning et al. 2015), and other sophisticated methods require however further evaluation before they translate to established and feasible methods, meeting Van der Vleuten's utility criteria of reliability, validity, cost-effectiveness, educational impact and acceptability, and other useful measures of quality (van der Vleuten and Schuwirth 2005).

References

AERA/APA/NCME. (2014). In B. Plake, L. Wise, et al. (Eds.), *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Ber, R. (2003). The CIP (comprehensive integrative puzzle) assessment method. *Medical Teacher, 25*(2), 171–176.

Beullens, J., et al. (2002). Are extended-matching multiple-choice items appropriate for a final test in medical education? *Medical Teacher, 24*(4), 390–395.

Bowen, J. L., & Ilgen, J. S. (2014). Now you see it, now you don't: What thinking aloud tells us about clinical reasoning. *Journal of Graduate Medical Education, 6*, 783–785.

Broek, W. E. S., et al. (2012). Effects of two different instructional formats on scores and reliability of a script concordance test. *Perspectives on Medical Education, 1*(3), 119–128.

Case, S. M., & Swanson, D. B. (1998). *Constructing written test questions for the basic and clinical sciences* (2nd ed.). Philadelphia: National Board of Medical Examiners.

Charlin, B., et al. (2000). The script concordance test: A tool to assess the reflective clinician. *Teaching and Learning in Medicine, 12*(4), 189–195.

Cook, D. A., et al. (2015). A contemporary approach to validity arguments: A practical guide to Kane's framework. *Medical Education, 49*(6), 560–575.

Dory, V., et al. (2016). In brief: Validity of case summaries in written examinations of clinical reasoning. *Teaching and Learning in Medicine, 0*(0), 1–10. Available at: http://ezproxy.usher-brooke.ca/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=mnh&AN=272 94400&site=ehost-live

Downing, S. M. (2003). Validity: On the meaningful interpretation of assessment data. *Medical Education, 37*(9), 830–837.

Durning, S. J., et al. (2015). Neural basis of nonanalytical reasoning expertise during clinical evaluation. *Brain and Behaviour, 309*, 1–10.

Forsberg, E., et al. (2016). Assessing progression of clinical reasoning through virtual patients: An exploratory study. *Nurse Education in Practice, 16*(1), 97–103.

Groothoff, J. W., et al. (2008). Growth of analytical thinking skills over time as measured with the MATCH test. *Medical Education, 42*(10), 1037–1043.

Groves, M., Scott, I., & Alexander, H. (2002). Assessing clinical reasoning: A method to monitor its development in a PBL curriculum. *Medical Teacher, 24*(5), 507–515.

Hawkins, R. E., & Boulet, J. R. (2008). Direct observation: Standardized patients. In E. S. Holmboe & R. E. Hawkins (Eds.), *Practical guide to the evaluation of clinical competence* (pp. 102–118). Philadelphia: Mosby Elsevier.

Hrynchak, P., Glover Takahashi, S., & Nayer, M. (2014). Key-feature questions for assessment of clinical reasoning: A literature review. *Medical Education, 48*(9), 870–883.

Kok, E. M., & Jarodzka, H. (2017). Before your very eyes: The value and limitations of eye tracking in medical education. *Medical Education, 51*(1), 114–122.

Kunina-Habenicht, O., et al. (2015). Assessing clinical reasoning (ASCLIRE): Instrument development and validation. *Advances in Health Sciences Education, 20*(5), 1205–1224.

Lineberry, M., Kreiter, C. D., & Bordage, G. (2013). Threats to validity in the use and interpretation of script concordance test scores. *Medical Education, 47*(12), 1175–1183.

Lubarsky, S., et al. (2011). Script concordance testing: A review of published validity evidence. *Medical Education, 45*(4), 329–338.

Macrae, H., et al. (2000). A comprehensive examination for senior surgical residents. *American Journal of Surgery, 179*, 190–193.

Miller, G. E. (1990). The assessment of clinical skills/competence/performance. *Academic Medicine, 87*(7), S63–S67.

Norcini, J., & Burch, V. (2007). Workplace-based assessment as an educational tool: AMEE guide no. 31. *Medical Teacher, 29*(9), 855–871.

Page, G., & Bordage, G. (1995). The Medical Council of Canada's key features project: A more valid written examination of clinical decision-making skills. *Academe, 70*(2), 104–110.

Page, G., Bordage, G., & Allen, T. (1995). Developing key-feature problems and examinations to assess clinical decision-making skills. *Academic Medicine, 70*, 194–201.

Rademakers, J., ten Cate, O., & Bär, P. R. (2005). Progress testing with short answer questions. *Medical Teacher, 27*(7), 578–582.

Rencic, J., et al. (2016). Understanding the assessment of clinical reasoning. In P. Wimmers & M. Mentkowski (Eds.), *Assessing competence in professional performance across disciplines and professions* (pp. 209–235). Cham: Springer International Publishing.

Schuwirth, L. W., et al. (1996). Computerized long-menu questions as an alternative to open-ended questions in computerized assessment. *Medical Education, 30*(1), 50–55.

Singh, T., & Norcini, J. (2013). Workplace-based assessment. In W. McGaghie (Ed.), *International best practices for evaluation in the health professions* (pp. 257–279). London: Radcliffe Publishing Ltd.

Sloane, D., et al. (1995). The objective structured clinical examination. The new gold standard for evaluating. *Annals of Surgery, 222*(6), 735–742.

Tekian, A., & Yudkowsky, R. (2007). Oral examinations. In S. Downing & R. Yudkowsky (Eds.), *Assessment in health professions education* (pp. 269–286). New York: Routledge.

ten Cate, O. (1997). In A. Scherpbier et al. (Eds.), *Comparing reliabilities of true/false and short-answer questions in written problem solving tests* (pp. 193–196). Dordrecht: Kluwer Academic Publishers.

van Bruggen, L., et al. (2012). Preferred question types for computer-based assessment of clinical reasoning: A literature study. *Perspectives on Medical Education, 1*(4), 162–171.

van der Vleuten, C. P. M., & Schuwirth, L. W. T. (2005). Assessing professional competence: From methods to programmes. *Medical Education, 39*(3), 309–317.

Williams, R. G., et al. (2011). Tracking development of clinical reasoning ability across five medi-
cal schools using a progress test. *Academic Medicine: Journal of the Association of American
Medical Colleges, 86*(9), 1148–1154.

Woods, N. N., & Mylopoulos, M. (2015). On clinical reasoning research and applications:
Redefining expertise. *Medical Education, 49*(5), 543–543.