



Archived by Flinders University

This is the peer reviewed, final published version of the following article:

Schuwirth, L. W. T., Durning, S. J., & King, S. M. (2020). Assessment of clinical reasoning: three evolutions of thought. In *Diagnosis* (Vol. 7, Issue 3, pp. 191–196). Walter de Gruyter GmbH

which has been published at:

<https://doi.org/10.1515/dx-2019-0096>

Copyright © 2020 Walter de Gruyter GmbH, Berlin/
Boston.

Mini Review

Lambert W.T. Schuwirth*, Steven J. Durning and Svetlana M. King

Assessment of clinical reasoning: three evolutions of thought

<https://doi.org/10.1515/dx-2019-0096>

Received December 1, 2019; accepted February 12, 2020

Abstract: Although assessing clinical reasoning is almost universally considered central to medical education it is not a straightforward issue. In the past decades, our insights into clinical reasoning as a phenomenon, and consequently the best ways to assess it, have undergone significant changes. In this article, we describe how the interplay between fundamental research, practical applications, and evaluative research has pushed the evolution of our thinking and our practices in assessing clinical reasoning.

Keywords: assessment; clinical reasoning.

Introduction

The ability to make safe and accurate clinical decisions is the bedrock of a competent health care professional. It is, therefore, fair to claim that the development of clinical reasoning is seen as central to health profession education. Yet, teaching and assessing clinical reasoning remain vexing problems. The particularly problematic domain of assessment is further complicated by the challenge of addressing poor clinical reasoning, which is often associated with diagnostic error [1, 2].

The complexity of assessment of clinical reasoning is evidenced by the lack of “magic bullet.” Yet to be developed is a “revolutionary” new method or, at least, a method, which renders all others obsolete. This is not at all surprising. Young et al. [3] recently published a comprehensive literature review in which they collated various perspectives about what clinical reasoning entails. This

review demonstrated a wide variety of perceptions about clinical reasoning, defining it as a behavior, a process, an ability, the outcome of the context, and an outcome as well as a process.

While attempts have been made to develop methods using validity evidence to assess clinical reasoning, not all have been successful. They have, however, made important contributions to the development of our current understanding of the concept of clinical reasoning and what happens when it goes awry, increasing the likelihood of medical error [4]. In this paper, we describe how this development occurred from our own perspective. In our view, assessment of clinical reasoning has moved from a linear, predictable, and quantitative measurement perspective, to a complex, dynamic situation-specific and qualitative narrative perspective. We do not claim to report a complete history, nor do we suggest that this is the result of a deliberate, systematic literature review. Instead, this paper reflects our current thinking, and documents our meaning- and sense-making about clinical reasoning.

A first major development in the 1960s adopted a pragmatic approach to designing an assessment that sought to mimic real-world clinical practice. This approach was called the patient management problem (PMP) [5] and was essentially a paper-based or computerized simulation of an actual patient case. The candidate was presented with the initial complaint of a patient case and subsequently worked toward a diagnosis and management plan. This was achieved by, for example, asking history questions, performing physical examinations, and ordering laboratory or radiographic tests as needed. Although authentic in their design, and becoming increasingly so with computerization (i.e. computer-based simulations), there were three main concerns with this approach, from an assessment point of view [4].

Challenge 1: problem-solving as idiosyncratic and individual

The first challenge arose when expert panels were asked to provide scores for all the decision points in the simulation

*Corresponding author: Lambert W.T. Schuwirth, MD, PhD, Prideaux Centre for Research in Health Professions Education, Flinders University, GPO Box, 2100 Adelaide, South Australia 5001, Australia, E-mail: lambert.schuwirth@flinders.edu.au

Steven J. Durning: Uniformed Services University, Bethesda, MD, USA, E-mail: steven.durning@usuhs.edu

Svetlana M. King: Prideaux Centre for Research in Health Professions Education, Flinders University, Adelaide, South Australia, Australia, E-mail: svetlana.king@flinders.edu.au

or paper case. These scores, which would identify whether a decision was good or poor, were used in the process of calculating a final clinical reasoning score. In so doing, it became clear that experts disagreed considerably about the value of certain decisions in the process. Indeed, and consistent with the family of theories discussed in this edition, these experts did not merely disagree on how many points to attribute to each decision, but also on the optimal pathway through the simulation [4].

From a cognitive psychological perspective, this “expert idiosyncrasy” is understandable. One of the more popular cognitive psychology theories that focuses on expertise development and problem-solving ability is script theory [6, 7]. Essentially, this theory claims that the development of clinical problem-solving expertise occurs through the creation of increasingly elaborate problem-solving (or illness) scripts. The process begins by collecting “islets” of isolated factual knowledge, which gradually connect to form semantic networks [8]. Gradually, the physician aggregates these semantic networks into illness scripts. These scripts enable the expert to quickly recognize and connect a patient problem with possible solutions, for example, “This is a typical XXX patient.” As physicians develop greater expertise, these illness scripts are enriched with contextual features, which extend beyond the content needed to arrive at a reasonable diagnostic and management plan. This allows the expert to recognize – almost effortlessly – the problem as a gestalt (or whole) [6]. From this theory, it is understandable that high-level illness scripts evolve as a result of an individual’s specific experiences in connection with his/her specific background knowledge. Because of this, it is, therefore, logical to assume that when an expert is asked to unpack this highly aggregated script, s/he will do so in a highly individual, idiosyncratic way. So, from an illness script theory perspective, understanding the development of clinical problem-solving expertise helps to explain why long, branched scenarios were an ineffective approach to the assessment of clinical reasoning.

Challenge 2: domain specificity

The second challenge—domain specificity – relates to the first. Successful clinical reasoning assumes that the physician has a good, relevant working knowledge as a prerequisite [9–11]. Herein lies the problem. One can be very knowledgeable in one domain and ignorant in another; hence, knowledge is “domain specific.” Because knowledge is both domain specific and a prerequisite for successful clinical reasoning, the process of clinical

reasoning is also domain specific, leading to the phenomenon of “content specificity” [12]. Hence, the way in which a candidate reasons through one problem is a poor predictor of the way s/he reasons through another problem, even within the same domain. Studies examining this phenomenon identified correlations between cases of 0.1 and 0.2 [4]. The implication of domain specificity is that large numbers of cases are required to produce a sufficiently reliable score. Given the elaborate nature of such simulations, lengthy testing times rendered this impracticable and unfeasible [4].

Challenge 3: expertise is associated with efficiency

When measures were put in place to mitigate idiosyncrasy and content specificity concerns, it was found that practicing physicians were generally outperformed by final year medicine students and doctors in their first and second year post-graduation [13, 14]. From a cognitive psychological perspective, this is unsurprising. Expert performance is often associated with efficiency due to the formation of illness scripts. Consequently, experts generally require less information than non-experts and even intermediates to identify a correct solution [15, 16]. This finding seriously challenged the validity of PMPs because they reward thoroughness of information collection, rather than efficiency. In this way, experts were essentially punished for their expertise.

Clinical problem-solving developments: key feature approach and extended matching items

Apart from key learnings related to idiosyncrasy, domain and content specificity, another important lesson was derived from this: authenticity or fidelity is not the same as validity [4]. Of course, this was not a new insight as construct validity theory already indicated that what looks valid by observation does not necessarily need to have construct validity [17, 18]. This distinction is important because of its application to simulation, and therefore, any modern virtual reality attempt to assess clinical reasoning. If the problems surrounding the assessment of clinical reasoning are not about lack of authenticity, virtual reality simulations are not the solution.

In the late 1980s and early 1990s, in response to the three challenges described above, two approaches were developed to assess clinical reasoning: key

feature approach and extended matching items. These approaches are similar in that they both use short cases and only require a small number of essential decisions to be made in relation to the case. In key feature assessment, short case presentations are coupled with a few questions aimed at essential decisions [19, 20]. Formats for these questions vary and may include open-ended and multiple-choice questions [21–23]. By keeping cases short and only asking a limited number of questions per case, a wide variety of problems can be explored per hour of testing time [24].

Extended matching items differ slightly in that they begin with the presentation of possible answers, together with a series of short cases (i.e. vignettes). The aim of this approach is to link one of the answers with each vignette [25]. For example, vignettes may be patient cases with a series of likely diagnoses from which to choose from. As with the key feature approach, extended matching uses short vignettes to enable the assessment of multiple problems per hour of testing time. These approaches were found to be successful in mitigating issues related to domain specificity, idiosyncrasy, and the intermediate effect [22, 25].

Although these two approaches were successful in addressing what they purported to achieve, they did not satisfy all components of the assessment of clinical reasoning. While both methods assess clinical decision making (with its focus on the final decision), they do not fully assess clinical reasoning (which focuses on the process *and* the decision). Although clinical decision making is generally seen as a pivotal component of clinical expertise, the challenge remained to develop assessment methods that would capture the thought processes behind the decision, to illuminate reasoning and error. An interesting development in this direction was a script concordance test [26, 27].

The script concordance test

In the script concordance test item, a very short vignette is presented in combination with a suggestion for a hypothesis (e.g. in a patient with symptoms A, B, and C, you are considering diagnosis X). An additional finding is presented, and the candidate is asked whether this finding makes the hypothesis more or less likely. The answer key and subsequent, weighted, scores are determined by the responses of a panel of experts [26, 27]. This test approach is accepted for its recognition of, and allowance for, candidates' own idiosyncratic patterns or "scripts," which is central to the illness script theory [28].

From an illness script theory perspective and our understanding of the idiosyncrasy of clinical reasoning, we may find it defensible to claim that for some, a certain finding makes the hypothesis more likely, and for others, less likely. This, however, is contentious. Some argue that this idiosyncrasy cannot be the case; in other words, a certain finding in a given context could result in either a hypothesis being more or less likely, but not both [29]. This paradox seems to be the result of the scoring approach used in the script concordance test. The stimulus (i.e. the short case and hypothesis) operates under the assumption that illness scripts are idiosyncratic, and there is an acceptable level of diversity. The scoring system operates under the assumption that individuals' illness scripts must be concordant with the average of a group of experts. Highlighting this paradox is essential in our explanation of further developments about the assessment of clinical reasoning.

To further explore this paradox, it is helpful to make another distinction between diagnostic decision making and clinical reasoning. The literature suggests that diagnostic decision making is not the same as other aspects of decision making and reasoning and is likely to be prone to different types of errors [1]. Klein [30] argued that in order to better understand error in naturalistic decision making, a distinction between the quality of the decision-making process and the quality of the outcomes is important. If we make this distinction, diagnostic decision making is generally treated more as a "convergent" process involving a broad collection of information that is ultimately summarized into one (or two) best solutions. This is reflected in the diagnostic process, in which all available information (e.g. history, physical examination, laboratory results) is summarized into one diagnostic classification. For assessment, this process is then, logically, relatively straightforward, as the stimulus and a correct response can be pre-determined. In contrast to diagnostic decision making, clinical reasoning can often involve the concurrent identification of multiple, "good" solutions. This process is then likely to be better understood as a non-linear (often complex), divergent, and, to a certain extent, unpredictable process.

Clinical reasoning and situational awareness

Complex processes are not necessarily difficult; for example, history taking could be considered a complex process. Prior to a history taking, there is no way to predict what will be said at a certain point in time (e.g. 4 min and 15 s into the conversation). This is because the ensuing

conversation is the result of myriads of interactions, both verbal and non-verbal, which are highlighted in the family of theories in this special edition. Yet, most clinicians typically manage this complexity without encountering any major problems. This example demonstrates that, although we might accept that there are multiple, concurrent, “good” solutions, it is also not a case of “anything goes” [31]. Rather, there are *boundaries* between acceptable and unacceptable actions and utterances [32, 33]. These boundaries may not be clear-cut or even pre-defined. Rather, they are fuzzy and can evolve or change during the course of the conversation. Consequently, it is important to utilize a range of strategies (i.e. actions and utterances) to flexibly deal with the interaction at any given point in time. Selecting and implementing an optimal strategy at every moment requires “situation awareness” – being mindful and responsive to what is effective and ineffective, and agilely moving between strategies. This (hopefully intuitive) understanding of how to navigate complexity is readily transferable to our understanding of clinical reasoning [34].

From this point of view, one could argue that clinical reasoning occurs as a narrative; either as an internal dialogue, or between the physician and others (e.g. the patient, colleagues, the examiner, additional staff). Consequently, the distinction between a good/poor clinical reasoner does not lie merely in the identification of the single best solution, but rather, the extent to which s/he can agilely utilize a repertoire of strategies and demonstrate situation awareness to convey his/her message. From a clinical reasoning assessment perspective, this means that the assessment must first establish whether the physician oversteps boundaries (e.g. utilizing incorrect or untrue pathophysiological explanations). Second, the physician’s capacity to flexibly explain the clinical situation from various viewpoints (and combine these perspectives) must be assessed (e.g. the ability to explain influenza from public health, infectious disease, immunological, and microbial perspectives, etc.). Finally, it is important to assess the physician’s ability to adapt his/her reasoning process in the moment, to connect with the other communication partner(s) present in the encounter (be it the patient, patient’s family, examiner, or colleague).

Clinical reasoning and situativity theory

The features outlined above, and particularly the situation-specific nature of clinical reasoning, has given rise to a final perspective on clinical reasoning; a social cognitive family of theories often referred to as “situativity

theory” [35]. This family of theories suggests that clinical reasoning occurs in the moment and, therefore, should only be assessed in the “here-and-now.” For example, ecological psychology emphasizes that each situation provides “affordances” (i.e. opportunities for action) to the people – the actors – in the situation who, in turn, have effectivities (i.e. skills and abilities to act) [34, 36]. Affordances indicate what the situation enables and disallows the actor(s) to do. Effectivities relate to what the actor(s) is/are able and unable to do with the given situation. For example, the information that the patient is willing to share during history taking is an affordance that the clinician can use in the clinical reasoning process. Information that the patient does not share does not contribute to the affordances. The extent to which the clinician can, for example, record and remember patient information and make sense of that information in relation to his/her prior knowledge constitutes the effectivities. From a situativity perspective, good clinical reasoning is, therefore, the continual and purposeful alignment of effectivities and affordances. From this perspective, assessing clinical reasoning still requires direct human observation. This is because it involves constant evaluation of whether the physician’s utterances and actions remain within acceptable boundaries, are sufficiently varied, and adequately agile during the interaction.

The role of human judgment in assessment

Human judgment has long been discarded in assessment as being too subjective and unreliable. We have now, however, come to realize that human judgment is integral to the assessment of clinical reasoning and error, as for many other aspects of medical competence [37]. This is not to say that we have returned to the unstructured, often biased, bedside assessment of clinical reasoning of 60 years ago. On the contrary, we have learned much about the need for broad sampling [38–40] and the factors to address and avoid in direct observation- and workplace-based assessment [41, 42].

Indeed, literature is emerging, which helps to provide us with a better understanding of the role of expert human judgment in the assessment of clinical reasoning and error [37, 43–46]. We have come to realize that clinical reasoning and error cannot be predefined and “objectively” measured but, rather, involve the creation of a shared subjectivity. Creating this shared subjectivity requires an understanding of narratives (or stories), and this is a focus of current research. Studies, which explore the nature of human decision making in assessment have

adopted script theory from clinical decision making to view assessment as a “diagnostic” process [47–49]. Other studies explore the value of different perspectives on the same reasoning process. Traditional perspectives suggest that one view may be more correct than another, adopting the assumption that there is one single truth. From the perspective that there might be concurrent, multiple, good solutions, it is important to understand how different views contribute to our understanding of clinical reasoning as a multifaceted concept. Research by Gingerich et al. [37] recognizes this. Another line of research seeks to understand how narratives contribute to judgment and decision making – the narratives that assessors use and how assessors evaluate candidates’ narratives to judge the quality of their clinical reasoning [37, 43, 49]. Finally, another line of inquiry has explored circumstances surrounding “context specificity” – a situation in which a physician arrives at two different diagnoses after seeing two patients present with identical symptomatology. This is clearly a source of unwanted variance and error in health care. This research is seeking to unveil the unique affordances and effectivities that physicians use in a clinical situation, the findings of which will assist both our teaching and assessment of clinical reasoning [50–52].

Conclusions

In summary, the literature in recent decades has helped to progress our understanding of clinical reasoning and its assessment in the context of health professions education. In our view, assessment of clinical reasoning has transitioned from an endeavor designed to capture a linear, predictable process with instruments that predominantly collect quantitative measurements, to a view of clinical reasoning as a complex, interactive process that occurs in dynamic contexts and which is situation and context specific. In this paper, we have sought to explain why we see this change as underlying the current drive to include qualitative narrative perspectives. This change is neither small nor insignificant. We believe these developments offer one of few instances wherein the term “paradigm shift” is applicable. From a research perspective, it seems as though this shift is already occurring. It is, however, likely to take time before current scientific thinking becomes enconced in practical application. Effective knowledge translation (or knowledge mobilization) activities will play an essential role in achieving this.

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

Research funding: None declared.

Employment or leadership: None declared.

Honorarium: None declared.

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.

Disclaimer: The views expressed herein are those of the authors and not necessarily those of the Department of Defence or other federal agencies.

References

1. Graber ML, Franklin N, Ruthanna G. Diagnostic error in internal medicine. *Arch Intern Med* 2005;165:1493–9.
2. Norman G. Dual processing and diagnostic errors. *Adv Health Sci Educ Theory Pract* 2009;14:37–49.
3. Young M, Thomas A, Gordon D, Gruppen L, Lubarsky S, Rencic J, et al. The terminology of clinical reasoning in health professions education: implications and considerations. *Med Teach* 2019;41:1277–84.
4. Swanson DB, Norcini JJ, Grosso LJ. Assessment of clinical competence: written and computer-based simulations. *Ass Eval High Educ* 1987;12:220–46.
5. Berner ES, Hamilton LA, Best WR. A new approach to evaluating problem-solving in medical students. *J Med Educ* 1974;49:666–72.
6. Schmidt HG, Boshuizen HP. On acquiring expertise in medicine. Special Issue: European Educational Psychology. *Educ Psychol Rev* 1993;5:205–21.
7. Custers EJ, Boshuizen H, Schmidt HG. The role of illness scripts in the development of medical diagnostic expertise; results from an interview study. *Cogn Instr* 1998;16:367–98.
8. Boshuizen HP. De ontwikkeling van medische expertise; een cognitief-psychologische benadering [On the development of medical expertise; a cognitive-psychological approach] [Dissertation]. Rijksuniversiteit Limburg [Maastricht University], 1989.
9. Chi MT, Glaser R, Rees E. Expertise in problem solving. In: Sternberg RJ, editor. *Advances in the psychology of human intelligence*. Hillsdale NJ: Lawrence Erlbaum Associates, 1982;1:7–76.
10. Glaser R, Chi MT. Overview. In: Chi MT, Glaser R, Farr MJ, editors. *The nature of expertise*. Hillsdale, NJ, USA: Lawrence Erlbaum Associates, Inc, 1988:xv–xxviii.
11. Polsen P, Jeffries R. Problem solving as search and understanding. In: Sternberg RJ, ed. *Advances in the psychology of human intelligence*. Hillsdale NJ: Lawrence Erlbaum Associates, 1982:367–411.
12. Eva K. On the generality of specificity. *Med Educ* 2003;37:587–8.
13. Schmidt HG, Boshuizen HP. On the origin of intermediate effects in clinical case recall. *Mem Cognit* 1993;21:338–51.
14. Schmidt HG, Boshuizen HP, Hobus PP. Transitory stages in the development of medical expertise: The “intermediate effect” in clinical case representation studies. Proceedings of the 10th Annual Conference of the Cognitive Science Society. Montreal, Canada: Lawrence Erlbaum Associates, 1988:139–45.

15. Boreham NC. The dangerous practice of thinking. *Med Educ* 1994;28:172–79.
16. Kahneman D. *Thinking, fast and slow*. Penguin Books. New York, Farrar: Straus and Giroux, 2011.
17. Kane M. Current concerns in validity theory. *J Educ Meas* 2001;38:319–42.
18. Cronbach LJ, Meehl PE. Construct validity in psychological tests. *Psychol Bull* 1955;52:281–302.
19. Page G, Bordage G, Allen T. Developing key-feature problems and examinations to assess clinical decision-making skills. *Acad Med* 1995;70:194–201.
20. Page G, Bordage G, Harasym P, Bowmer I, Swanson DB. A new approach to assessing clinical problem-solving skills by written examination: conceptual basis and initial pilot test results. In: Bender W, Hiemstra RJ, Scherpbier A, et al., editors. *Teaching and Assessing Clinical Competence, Proceedings of the fourth Ottawa conference*. Groningen, The Netherlands: Boekwerk Publications, 1990:403–7.
21. Schuwirth LW, Van der Vleuten CP, De Kock CA, Peperkamp AG, Donkers HH. Computerized case-based testing: a modern method to assess clinical decision making. *Med Teach* 1996;18:295–300.
22. Schuwirth LW. An approach to the assessment of medical problem solving: computerised case-based testing. PhD Thesis, Universiteit Maastricht, 1998.
23. Schuwirth LW, Blackmore DB, Mom E, Van den Wildenberg F, Stoffers H, Van der Vleuten CP. How to write short cases for assessing problem-solving skills. *Med Teach* 1999;21:144–50.
24. Norman G, Bordage G, Page G, Keane D. How specific is case specificity? *Med Educ* 2006;40:618–23.
25. Case SM, Swanson DB. Extended-matching items: a practical alternative to free response questions. *Teach Learn Med* 1993;5:107–15.
26. Charlin B, Brailovsky C, Leduc C, Blouin D. The diagnostic script questionnaire: a new tool to assess a specific dimension of clinical competence. *Adv Health Sci Educ Theory Pract* 1998;3:51–8.
27. 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;12:185–91.
28. Lubarsky S, Dorie V, Duggan P, Gagnon R, Charlin B. Script concordance testing: from theory to practice: AMEE Guide No. 75. *Med Teach* 2013;35:184–93.
29. Lineberry M, Kreiter CD, Bordage G. Threats to validity in the use and interpretation of script concordance test scores. *Med Educ* 2013;47:1175–83.
30. Klein G. Sources of error in naturalistic decision making tasks. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. Los Angeles, CA, USA: SAGE Publications, 1993.
31. Rosas SR. Systems thinking and complexity: considerations for health promoting schools. *Health Promot Int* 2015;32:301–11.
32. Checkland P. From optimizing to learning: a development of systems thinking for the 1990s. *J Op Res Soc* 1985;36:757–67.
33. Ulrich W. The quest for competence in systemic research and practice. *Sys Res Beh Sci* 2001;18:3–28.
34. Durning SJ, Artino A, Pangaro L, Van der Vleuten C, Schuwirth L. Redefining context in the clinical encounter: implications for research and training in medical education. *Acad Med* 2010;85:894–901.
35. Durning SJ, Artino AR. Situativity theory: a perspective on how participants and the environment can interact: AMEE Guide No. 52. *Med Teach* 2011;33:188–99.
36. Young M. An ecological psychology of instructional design: learning and thinking by perceiving–acting systems. In: Jonassen D, Driscoll M, editors. *Handbook of research on educational communications and technology*. New York: Routledge, 2013:180–88.
37. Gingerich A. Questioning the rater idiosyncrasy explanation for error variance by searching for multiple signals within the noise. PhD thesis, Maastricht University, 2015.
38. Norcini JJ, Swanson DB. Factors influencing testing time requirements for measurements using written simulations. *Teach Learn Med* 1989;1:85–91.
39. Swanson DB. A measurement framework for performance-based tests. In: Hart I, Harden R, editors. *Further developments in Assessing Clinical Competence*. Montreal: Can-Heal Publications, 1987:13–45.
40. Swanson DB, Norcini JJ. Factors influencing reproducibility of tests using standardized patients. *Teach Learn Med* 1989;1:158–66.
41. Nair BR, Hensley MJ, Parvathy MS, Lloyd DM, Murphy B, Ingham K, et al. A systematic approach to workplace-based assessment for international medical graduate. *Med J Aust* 2012;196:399–402.
42. Govaerts MJ, Van der Vleuten CP, Schuwirth LW, Muijtjens AM. Broadening perspectives on clinical performance assessment: rethinking the nature of in-training assessment. *Adv Health Sci Educ Theory Pract* 2007;12:239–60.
43. Valentine N, Schuwirth L. Identifying the narrative used by educators in articulating judgement of performance. *Persp Med Educ* 2019;8:1–7.
44. Cook DA, Kuper A, Hatala R, Ginsburg S. When assessment data are words: validity evidence for qualitative educational assessments. *Acad Med* 2016;91:1359–69.
45. Ginsburg S, Regehr G, Lingard L, Eva K. Reading between the lines: faculty interpretations narrative evaluation comments. *Med Educ* 2015;49:296–306.
46. Ginsburg S, Vleuten CP, Eva KW, Lingard L. Cracking the code: residents' interpretations of written assessment comment. *Med Educ* 2017;51:401–10.
47. Govaerts MJ, Schuwirth LW, Van der Vleuten CP, Muijtjens AM. Workplace-based assessment: effects of rater expertise. *Adv Health Sci Educ Theory Pract* 2011;16:151–65.
48. Govaerts MJ, Van de Wiel MW, Schuwirth LW, Van der Vleuten CP, Muijtjens AM. Workplace-based assessment: raters' performance theories and constructs. *Adv Health Sci Educ Theory Pract* 2013;18:375–96.
49. Berendonk C, Stalmeijer RE, Schuwirth LW. Expertise in performance assessment: assessors' perspectives. *Adv Health Sci Educ Theory Pract* 2013;18:559–71.
50. Durning SJ, Artino AR, Boulet JR, Dorrance K, Van der Vleuten C, Schuwirth L. The impact of selected contextual factors on experts' clinical reasoning performance (does context impact clinical reasoning performance in experts?). *Adv Health Sci Educ Theory Pract* 2012;17:65–79.
51. Durning SJ, Artino Jr AR, Pangaro L, Van der Vleuten CP, Schuwirth LW. Context and clinical reasoning: understanding the perspective of the expert's voice. *Med Educ* 2011;45:927–38.
52. Durning SJ, Trowbridge RL, Schuwirth L. Clinical reasoning and diagnostic error: a call to merge two worlds to improve patient care. *Acad Med* 2019. DOI: 10.1097/ACM.0000000000003041 [Epub ahead of print].