

Eric S. Holmboe\* and Steven J. Durning

# Assessing clinical reasoning: moving from in vitro to in vivo

**Abstract:** The last century saw dramatic changes in clinical practice and medical education and the concomitant rise in high-stakes, psychometrically-based examinations of medical knowledge. Higher scores on these high-stakes “in-vitro” examinations are modestly associated with better performance in clinical practice and provide a meaningful degree of assurance to the public about physicians’ competency in medical knowledge. However, results on such examinations explain only a small fraction of the wide variation currently seen in clinical practice and diagnostic errors remain a serious and vexing problem for patients and the healthcare system despite decades of high-stakes examinations. In this commentary we explore some of the limitations of high-stakes examinations in assessing clinical reasoning and propose utilizing situated cognition theory to guide research and development of innovative modes of “in-vivo” assessments that can be used in longitudinally and continuously in clinical practice.

**Keywords:** clinical decision making; clinical reasoning; diagnostic error.

---

\*Corresponding author: Eric S. Holmboe, MD, ABIM, 510 Walnut Street, Philadelphia, PA 19106, USA, Phone: +(215)-446-3606, E-mail: eholmboe@abim.org

Eric S. Holmboe: American Board of Internal, Philadelphia, PA, USA  
Steven J. Durning: Uniformed Services University of the Health Sciences, Bethesda, MD, USA

## Introduction

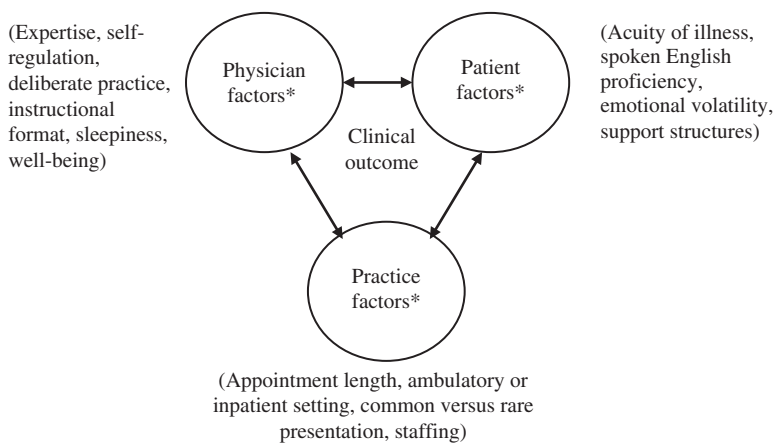
The last century saw dramatic changes in both clinical practice and medical education. Fueled in part by scientific and technological advances, many diseases and conditions became either curable or capable of being better managed and controlled. Typically, appropriate and successful therapy requires a correct diagnosis and to be sure concomitant advances in diagnostic testing and imaging have undoubtedly facilitated the diagnostic process. Yet, even in this age of profound technological change the very human process of clinical reasoning leading to accurate diagnosis remains central to high quality and safe patient

care. Diagnosis drives therapeutic decisions with physicians and patients.

Catalyzed by the Flexner report in 1910 and subsequently other calls for reforms over the last century, medical education in the US and other countries placed a premium on “science” resulting in an intense focus on acquiring and assessing medical knowledge among medical students, residents and fellows [1]. Medical knowledge is essential to clinical expertise and continues to be viewed as one of a physician’s core competencies [2]. Yet studies have shown that medical knowledge, while important to clinical reasoning, is necessary but not sufficient for diagnostic expertise. Take for example, the problem of context specificity or the phenomenon whereby a physician can see two patients with the same chief complaint and the same (or nearly so) presenting history and physical examination and yet the physician comes to two different diagnostic decisions [3, 4].

What is clinical reasoning and how does it differ from clinical decision making? For purposes of this essay, we define clinical reasoning as the cognitive operations allowing clinicians to observe, collect, and analyze information that ultimately leads to an action (i.e., diagnosis and therapy). Clinical reasoning refers to the steps up to and including establishing the diagnosis and treatment, which differs from clinical decision making where the emphasis is on the decision step (establishing the diagnosis and treatment). In other words, clinical reasoning typically entails a more inclusive view of what is occurring in the clinical encounter.

To further set the stage for this essay’s focus on assessment, we will also frame clinical reasoning in contemporary educational theory, notably situated cognition [5–8]. Situated cognition argues that thinking (cognition) emerges from individual(s) acting in concert with their environment. It shifts the emphasis from solely the physician to the physician interacting with the patient within a specific setting or encounter (e.g., a specific situation; some examples of component parts are shown in the Figure 1 illustrating the participants and the setting). Viewed through this perspective, the component parts of clinical reasoning and how these parts can and often do interact becomes apparent which can help with defining where reasoning goes awry when it does, (e.g., diagnostic



**Figure 1** Situated cognition in the clinical encounter.

\*Circles could also be added for other participants (and their individual elements) such as nurse(s), medical student(s) or resident(s) or other participants in the encounter. The number of potential interactions with including additional participants can grow substantially.

error) as well as better understand when it goes well. This in turn can lead to methods and approaches to help physicians (and trainees) in the messy and, at times, complex nature of clinical practice (e.g., clinical reasoning “in vivo”).

We should also briefly mention two other important concepts, namely distributed cognition and situational awareness, we will not address that have been noted by others studying diagnostic error [9, 10]. Distributed cognition is a form of situativity theory and thus shares many features with situated cognition. However, the emphasis is different in that distributed cognition emphasizes information is shared by all (situated cognition does not specify this) and distributed cognition assumes that the components of the system must rely on each other to get the job done (this is not an assumption of situated cognition). For this paper we argue that situated cognition is more inclusive in that it does not make these assumptions and thus may be more helpful with “diagnosing” when things go awry in clinical encounters, especially in the context of assessing the individual practitioner [8].

Situation awareness takes into account the complex interplay between elements in a given situation but the emphasis is on the individual decision maker. However, it does not provide a straightforward means, like situated cognition, for identifying and exploring the elements that may impact individual clinical performance in an encounter. Situated cognition provides different ways of viewing context in the clinical encounter [11].

Assessment is a critical activity in medical education and can help educators in many ways. First, assessment helps to codify and define a professional field of study. Second, assessment could help to assure the public a physician possessed the necessary clinical reasoning

expertise to provide high quality and safe care. Finally, assessment can enhance learning through feedback and further study [12]. However, there are many challenges to constructing assessments for clinical reasoning in the clinical practice setting. For example, studies have found that it is difficult to disentangle knowledge, experience, and reasoning [13, 14]. Indeed, it has been argued that clinical reasoning is revealed only in action; such action implicitly would incorporate the potential influence of the unique context of the clinical encounter as we have outlined in Figure 1 [15, 16].

We will now discuss some of what we as a medical profession have learned and will propose some ideas to consider pursuing in terms of assessing clinical reasoning as a strategy to reduce the impact of diagnostic errors. We will use an activity that has been described in the expertise literature (the analogy of pitching in baseball) in addition to situated cognition theory to help illustrate and illuminate our perspectives and propositions.

## Assessing medical knowledge: where we have been

The last 100 years witnessed the rise of the science of testing, grounded in psychometrics, with advances in theory and subsequent application in high-stakes testing of medical knowledge as a routine and growing component of medical education in many countries [17]. The National Board of Medical Examiners was formed to provide licensing exams in 1915 and the first specialty certification examination was given just a few years later in ophthalmology [18, 19]. Exams have served as a useful

assessment methodology because of the need for physicians to possess and apply medical knowledge which is a necessary component of clinical reasoning required for effective and efficient patient care [20, 21]. In short, physicians cannot function with a “hard drive” devoid of medical knowledge and demonstrate expert performance in clinical reasoning. Multiple studies illustrate that performance on high-stakes examinations in the form of well-written patient vignettes are associated with quality of clinical care [22–24]. In-training examinations, also using MCQ formats, have also grown in popularity as formative assessments for medical knowledge and valid predictors of performance on high-stakes certification exams [25, 26].

Yet, despite the clear value of psychometrically-based examinations, usually delivered through vignette-based or other formats of multiple choice questions that assess the “hard drive”, the rate and impact of diagnostic errors has not changed dramatically over the last few decades and remains a vexing and serious problem for healthcare systems around the globe [27, 28]. This sobering reality highlights that assessment through examinations is insufficient to address the diagnostic error problem. There are many reasons for faulty clinical reasoning and we will explore and propose some additional assessment strategies that we believe deserve attention for research.

Returning to our Figure 1, exams emphasize the knowledge acquisition model by the circle represented by the physician. It does not take into account situational factors (such as the patient interacting with the physician within a system). A useful analogy found in major league baseball might be helpful to better understand the various components of clinical reasoning in the context of a “live or in-vivo performance”. Pitchers prepare by throwing in the bullpen prior to starting a game or coming in to provide relief. A lot can be learned about pitching skill and ideas for improvement in the bull pen (“simulated practice”) but it is devoid of the actual game situations that the pitcher will experience. Any baseball fan can tell you that what happens in the bull pen may or may not predict how the pitcher performs in the game when he confronts a batter (in medicine the patient) in tense game conditions characterized by who is on base, weather conditions, inning of the game, type of turf (“system or encounter” factors in medicine) are taken into account.

There are other approaches for testing medical knowledge currently in limited use that resemble a form of an “oral exam” and can function as a form of “game tape review” for physicians. While oral exams have fallen out of favor in many high-stakes testing programs, techniques such as chart stimulated recall (CSR) and a variant known

as case-based discussion (CBD) are still used in several contexts. CSR is a validated component of the Physician Assessment Review (PAR) program in Canada (most notably the province of Alberta) and CBD has been studied as part of the United Kingdom’s Foundation programme [29–33]. Both techniques used mostly for formative assessments have been found to be reliable and perceived as useful by examiners and examinees alike. Typically, the medical record of a clinical encounter, chosen by the physician, trainee or assessor, is first reviewed by the assessor against a structured template that produces a series of questions designed to probe the “why” behind the physician’s actions and decisions. The assessor uses these questions in a one-on-one session with the physician, following and documenting the rationale and reasoning behind the physician’s choices as documented in the medical record, plus any additional pertinent information not documented.

Because both methods use the physician’s own patient encounters as the clinical material for the assessment, these techniques probably deserve greater attention for more routine and systematic use in examining clinical reasoning and diagnostic error. The challenge with such assessments, in addition to time and trained assessors, is sufficient sampling of patient encounters and associated contexts, but as we’ll discuss below health information technology may be a facilitator in the future. This is akin to major league baseball pitching – to improve their performance, a pitcher will probably need to review a lot of game films (sufficient sampling). Once one engages in authentic practice-based assessments with multiple moving parts (in vivo assessment of clinical reasoning) making sure that there is sufficient sampling is essential.

## What we do not know (well enough) – challenges of “in vivo” complexity

There has been substantial research on the clinical reasoning process from the physician’s viewpoint, usually depicted as a model using specific processes and steps (e.g., dual processing theory, illness script formation, etc.) that the decision maker (e.g., the physician) goes through [34–36]. While the decision-maker (physician) is the “final common pathway” for reasoning, things other than knowledge and the clinical facts of the patient’s presentation can impact their performance. More research and attention needs to be given to the role of context in diagnostic reasoning as we argue above.

The growing complexity of many patient presentations complicated by multiple co-morbidities and occurring in the context of chaotic and fragmented clinical environments has grown exponentially in the last 25 years [37]. In exams and controlled laboratory experiments around diagnostic reasoning, context is a controllable element, not so in actual clinical practice, let alone ambiguity which can further complicate the process of assessment. For example, there are clinical situations where more than one answer is acceptable in practice – this is often the crux of shared decision making with patients. Work is urgently needed to better understand how the clinical environment (system factors, see Figure 1) and how the different participants (and their associated factors, Figure 1) interact to affect clinical reasoning. For example, in a study of hospitalized patients experiencing adverse events, Graber and colleagues found that a combination of “system” and “cognitive” errors, on average roughly totaling seven combined factors, contributed to each adverse event [38]. Durning and colleagues found that board certified internists faced with typical presentations of common conditions in internal medicine were impacted by a variety of contextual factors (see Figure 1) [39, 40]. What we need to better understand is how the factors in Figure 1 interact to affect clinical reasoning and produce diagnostic errors and harm. Furthermore, we need to understand how physicians can be trained to deal with contextual complexity and how systems can be better designed to reduce the potential to exacerbate diagnostic errors. Returning to our baseball analogy, there are several ways to address the issue of context. Our above example of reviewing game films would represent one means to do this. Another example would entail having a pitcher practice in different game situations (e.g., different pitch counts, runners on base).

## The “hard drive” has limits

You simply cannot carry everything on the hard drive any more, as if you really ever could. To be sure, physicians could carry a relatively greater proportion of what was known in the past, but it was probably a misguided notion to believe you could always carry “enough” on the hard drive to provide high quality care without looking things up. Regrettably, medicine has stubbornly held onto the belief that the hard drive should still reign “all supreme”. The amount of new knowledge now being generated is simply staggering. In a compelling article,

Stead and colleagues foreshadowed that personalized medicine, guided by genomics and proteomics, is the next “explosion” in clinical medicine that will mandate the use of effective clinical decision support (CDS) given the potential number of combinations available for additional diagnostic work-up and therapy across patients [41]. It is now simply untenable to carry everything you need in your head. This is akin to expecting a major league baseball player knowing when to throw the “right pitch” in every situation despite the substantially lower complexity of baseball compared to medicine. Extrapolating this to medicine highlights the order of magnitude in complexity a physician encounters in “game situations”. We already know that physicians recognize multiple gaps and clinical questions each day in practice that requires use of CDS [42]. Use of clinical decision support (e.g., when the pitcher receives signs from a catcher or talks to the pitching coach) will likely be a mandatory competency for all physicians. Unfortunately, current clinical decision support tools are still relatively early in their development and often difficult to use, but on the diagnostic front several web-based tools show promise and have already generated some empiric evidence for their effectiveness [43, 44]. The main question is how best to assess physician’s ability to recognize when and how to use diagnostic decision support tools.

## Does therapy affect diagnostic accuracy?

Another area lacking in understanding is how therapeutic options affect the clinical reasoning process. Akin to the scientific method where you “start with the end in mind”, physicians’ diagnostic process is likely affected by what they believe is therapeutically available to the patient. For example, if a physician is aware of an effective therapy, could such knowledge contribute to premature closure (i.e., choosing a diagnosis without considering other viable alternatives), especially in situations requiring timely decision making such as acute chest pain or acute respiratory distress? Could physicians jump to a diagnosis because they so desperately want to be able to provide a treatment to a patient? From personal experience, we have certainly seen cases where physicians provide a treatment on the hope that a patient actually has a diagnosis with the consequence, however, being the therapy inadvertently or inappropriately becomes part of a diagnostic “test” in the belief response to therapy validates the initial diagnosis. In some cases this may make sense, but

in others it likely creates a confirmation bias error. Indeed, while our understanding of diagnostic reasoning has dramatically grown as mentioned above, our understanding of therapeutic reasoning is still in its infancy.

In all of the examples above high-stakes, psychometrically-based testing is not likely to provide the methods and approaches to better understand these vexing challenges. New methods are needed.

## Rethinking medical record audit

One promising approach potentially foreshadowed by Singh and colleagues in a recent study involved the review of medical records of patients unexpectedly hospitalized or who returned to the emergency department or outpatient clinic visit within 14 days of the index visit. In this study, the authors found diagnostic error was a relatively common cause for the unanticipated re-visits [45]. This study suggests that novel uses of medical records may help to advance research in the in vivo setting to include effective use of clinical decision support and the interaction between diagnostic and therapeutic reasoning. In our simplistic baseball analogy, this approach would entail focusing on the pitch chart showing the distribution of balls and strikes and specifically focusing on why the pitcher threw balls (a form of “error” in baseball).

## Alternative (maybe radical) methods of diagnostic assessment?

### Alternative 1: team-based diagnosis?

Diagnosis has almost exclusively been viewed as the domain of the physician and as a “solo act”. While this continues to be the case in many settings, most notably ambulatory care in the form of a medical visit, this is not as true for settings such as the hospital, rehabilitation facilities, and other forms of institutionally-based care such as ambulatory surgical centers. Many residents, for example, can recall an incident in the ICU, ward or emergency department when an astute observation by a nurse, PA or NP led to the correct diagnosis or a revision in a diagnosis. We know that no single individual holds all the information needed for optimal patient care – however, this is often viewed through the lens of treatment and not diagnosis. Ironically, problem-based learning and team-based learning deliberately use group process for learning [46]. Why

then wouldn’t similar group processes be useful in clinical diagnosis? Should we not assess physicians and other healthcare providers for their ability to engage in group diagnostic processes when appropriate? This would be akin to team meetings at the mound to discuss the optimal pitching strategy in a given situation. Indeed, a clinical example growing in popularity is the use of “huddles” in patient-centered medical homes to proactively plan care for patients being seen in the clinic that day.

### Alternative 2: audio and video-review of diagnostic reasoning

In this age of smart phones, mini-cameras, iPads and the like, video review of the clinical encounter would seem to hold substantial potential to review diagnostic and treatment decisions as well as the intermediate steps in clinical reasoning. For example, a study by Gennis and colleagues, using a simple audio recorder, some years ago found that a resident’s presentation and assessment to one faculty attending who had not directly seen the patient was discordance up to a third of the time when the patient was evaluated concomitantly the same day by an independent second faculty physician [47]. Video is routinely used for standardized patients as a debriefing tool; why not actual clinical practice using peer coaching? While CSR and CBD can effectively utilize the written record to explore clinical reasoning, use of the written record often does not provides sufficient specifics on why things did or did not go well. The baseball analogy here would be having the pitcher review game films of hitters and situations.

## Conclusions

To return to our baseball analogy of clinical reasoning we would argue that a group of expert physicians seeing the same patient in an identical situation (Figure 1) would likely have *similar trajectories* (i.e., most arrive at the most reasonable diagnostic and treatment plans), just like a group of major league baseball pitchers could throw a baseball to a very similar, but not identical, location across the plate – most major league pitchers would still throw a “strike”. We would argue, as outlined in Figure 1, that a successful “in vivo” clinical encounter is a bounded condition much like a strike zone. In other words, we want our expert physicians to consistently “throw strikes” and show our physicians in training how to most efficiently learn how to do so.



What is needed now are more robust approaches to teaching, practicing and assessing clinical reasoning (use of teams, clinical decision support, better feedback loops) that can be effectively deployed longitudinally and continually. Assessing medical knowledge periodically through high-stakes examinations, while playing an important and necessary assurance role in professional self-regulation, will not be sufficient to “move the needle” in addressing the vexing and consequential diagnostic error problem. While this commentary covered only a

small portion of issues in assessing clinical reasoning, we hope this commentary has prompted some provocative thoughts for discussion within the medical community.

**Conflict of interest statement:** The authors declare no conflict of interest.

Received September 22, 2013; accepted November 11, 2013

## References

1. Cooke M, Irby D, O'Brien B. *Educating physicians: a call for reform of medical school and residency*. San Francisco, CA: Jossey-Bass, 2010.
2. Holmboe ES, Lipner R, Greiner A. Commentary: assessing quality of care: knowledge matters. *J Am Med Assoc* 2008; 299:338–40.
3. Eva K. On the generality of specificity. *Med Educ* 2003;37: 587–8.
4. Eva KW, Neville AJ, Norman GR. Exploring the etiology of content specificity: factors influencing analogic transfer and problem solving. *Acad Med* 1998;73:S1–5.
5. Wilson BG, Myers KM. Situated cognition in theoretical and practical context. In: Jonassen D, Land S, editors. *Theoretical foundations of learning environments*. Englewood Cliffs, NJ: Erlbaum, 1999.
6. Bredo E. Reconstructing educational psychology: situated cognition and Deweyan pragmatism. *Educ Psychol* 1994; 29:23–35.
7. Robbins P, Aydede M. *The Cambridge handbook of situated cognition*. New York: Cambridge University Press, 2009.
8. Durning SJ, Artino AR. Situativity theory: a perspective on how participants and the environment can interact: AMEE Guide no. 52. 2011;33:188–99.
9. Henriksen K, Brady J. The pursuit of better diagnostic performance: a human factors perspective. *BMJ Qual Saf* 2013;22(Suppl 2):ii1–ii5.
10. Singh H, Giardina TD, Petersen LA, Smith MW, Paul LW, Dismukes K, et al. Exploring situational awareness in diagnostic errors in primary care. *Brit Med J Qual Saf* 2012; 21:30–8.
11. Durning SJ, Artino AR, Pangaro LN, van der Veluten CP. Redefining context in the clinical encounter: implications for research and training in medical education. *Acad Med* 2010;85:894–901.
12. Norcini J, Anderson B, Bollela V, Burch V, Costa MJ, Duvivier R, et al. Criteria for good assessment: consensus statement and recommendations from the Ottawa 2010 Conference. *Med Teach* 2011;33:206–14.
13. Higgs J, Jones MA, Loftus S, Christensen N. *Clinical reasoning in the health professions*, 3rd ed. Philadelphia, PA: Elsevier, 2008.
14. Ericsson KA, Charness N, Feltovich P, Hoffman RR, editors. *The Cambridge handbook of expertise and expert performance*. New York: Cambridge University Press, 2006.
15. 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.
16. Brailovsky C, Charlin B, Beausoleil S, Coté 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;35:430–6.
17. Downing SM, Yudkowsky R. Introduction to assessment in the health professions. In: Downing SM, Yudkowsky R, editors. *Assessment in the health professions education*. New York: Routledge, 2009:1–20.
18. Melnick DE. History of public protection. Available at: <http://www.nbme.org/stakeholders/historypublicprotection.html>. Accessed on Sep 22, 2013.
19. American Board of Ophthalmology. Available at <http://abop.org/about/>. Accessed on Sep 22, 2013.
20. Downing SM. Written tests: constructed-response and selected-response formats. In: Downing SM, Yudkowsky R, editors. *Assessment in the health professions education*. New York: Routledge, 2009:149–84.
21. Lipner RS, Lucey CR. Putting the secure exam to the test. *J Am Med Assoc* 2010;304:1379–80.
22. Holmboe ES, Wang Y, Meehan TP, Tate JP, Ho S-Y, Starkey KS, et al. Association between maintenance of certification M examination scores and quality of care for medicare beneficiaries. *Arch Intern Med* 2008;168:1396–403.
23. Norcini JJ, Kimball HR, Lipner RS. Certification and specialization: do they matter in the outcome of acute myocardial infarction? *Acad Med* 2000;75:1193–8.
24. Norcini JJ, Lipner RS, Kimball HR. Certifying examination performance and patient outcomes following acute myocardial infarction. *Med Educ* 2002;36:853–9.
25. Babbott SF, Beasley BW, Hinchey KT, Blotzer JW, Holmboe ES. The predictive validity of the internal medicine in-training examination. *Am J Med* 2007;120:735–40.
26. de Virgilio C, Yaghoubian A, Kaji A, Collins JC, Deveney K, Dolich M, et al. Predicting performance on the American Board of Surgery qualifying and certifying examinations: a multi-institutional study. *Arch Surg* 2010;145:852–6.
27. Newman-Toker DE, Pronovost PJ. Diagnostic errors, the next frontier of patient safety. *J Am Med Assoc* 2009;301:1060–2.

28. Graber ML. The incidence of diagnostic error in medicine. *BMJ Qual Saf* 2013;22(Suppl 2):ii21–ii27.
29. Jennett P, Affleck L. Chart audit and chart stimulated recall as methods of needs assessment in continuing professional health education. *J Cont Educ Health Prof* 1998;18:163–71.
30. Hall W, Violato C, Lewkonja R, Lockyer J, Fidler H, Toews J, et al. Assessment of physician performance in Alberta: the physician achievement review. *Can Med Assoc J* 1999;161:52–7.
31. Schipper S, Ross S. Structured teaching and assessment: a new chart-stimulated recall worksheet for family medicine residents. *Can Fam Physician* 2010;56:958–9.
32. Davies H, Archer J, Southgate L, Norcini J. Initial evaluation of the first year of the Foundation Assessment Program. *Med Educ* 2009;43:74–81.
33. Norcini J, Burch V. Workplace-based assessment as an educational tool: AMEE Guide No. 31. 2007;29:855–71.
34. Norman G. Research in clinical reasoning: past history and current trends. *Med Educ* 2005;39:418–27.
35. Croskerry P. A universal model of diagnostic reasoning. *Acad Med* 2009;84:1022–8.
36. Schuwirth LW, van der Vleuten CP. General overview of the theories used in assessment. AMEE Guide No. 57. *Med Teach* 2011;33:783–97.
37. Boyd CM, Darer J, Boulton C, Fried LP, Boulton L, Wu AW. Clinical practice guidelines and quality of care for older patients with multiple comorbid diseases: implications for pay for performance. *J Am Med Assoc* 2005;294:716–24.
38. Graber ML, Franklin N, Gordon R. Diagnostic error internal medicine. *Arch Intern Med* 2005;165:1493–9.
39. Durning SJ, Artino AR, Boulet JR, Dorrance K, van der Veluten CP, Schuwirth L. The impact of selected contextual factors on experts' clinical reasoning performance (does context impact clinical reasoning performance experts?). *Adv Health Sciences Educ* 2012;17:65–79.
40. 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.
41. Stead WW, Searle JR, Fessler HE, Smith JW, Shortliffe EH. Biomedical informatics: changing what physicians need to know and how they learn. *Acad Med* 2011;86:429–34.
42. Green ML, Reddy SG, Holmboe ES. Teaching and evaluating point of care learning with an internet-based clinical question portfolio. *J Cont Educ Health Professions* 2009;29:209–19.
43. Graber ML, Ashlei M. Performance of a Web-based clinical diagnosis support system for internists. *J Gen Intern Med* 2008;23:37–40.
44. Reed DA, West CP, Holmboe ES, Halvorsen AJ, Lipner RS, Jacobs C, et al. Relationship of electronic medical knowledge resource use and practice characteristics with Internal Medicine Maintenance of Certification examination scores. *J Gen Intern Med* 2012;27:917–23.
45. Singh H, Giardina TD, Meyer AN, Forjuoh SN, Reis MD, Thomas EJ. Types and origins of diagnostic errors in primary care settings. *J Am Med Assoc Intern Med* 2013;173:418–25.
46. Parmelee DX. Team-based learning in health professions education: why is it a good fit? In: Michaelsen LK, Parmelee DX, McMahon KK, Levine RE, editors. *Team-based learning for health professions education*. Stylus: Sterling Virginia, 2008:3–8.
47. Gennis VM, Gennis MA. Supervision in the outpatient clinic: effects on teaching and patient care. *J Gen Intern Med* 1993;9:116.