Review

Joseph Rencic*, Lambert W.T. Schuwirth, Larry D. Gruppen and Steven J. Durning

A situated cognition model for clinical reasoning performance assessment: a narrative review

Abstract

Background: Clinical reasoning performance assessment is challenging because it is a complex, multi-dimensional construct. In addition, clinical reasoning performance can be impacted by contextual factors, leading to significant variation in performance. This phenomenon called context specificity has been described by social cognitive theories. Situated cognition theory, one of the social cognitive theories, posits that cognition emerges from the complex interplay of human beings with each other and the environment. It has been used as a valuable conceptual framework to explore context specificity in clinical reasoning and its assessment. We developed a conceptual model of clinical reasoning performance assessment based on situated cognition theory. In this paper, we use situated cognition theory and the conceptual model to explore how this lens alters the interpretation of articles or provides additional insights into the interactions between the assessee, patient, rater, environment, assessment method, and task.

Methods: We culled 17 articles from a systematic literature search of clinical reasoning performance assessment that explicitly or implicitly demonstrated a situated cognition perspective to provide an "enriched" sample with which to explore how contextual factors impact clinical reasoning performance assessment.

Results: We found evidence for dyadic, triadic, and quadratic interactions between different contextual factors, some of which led to dramatic changes in the assessment of clinical reasoning performance, even when knowledge requirements were not significantly different.

Conclusions: The analysis of the selected articles highlighted the value of a situated cognition perspective in understanding variations in clinical reasoning performance assessment. Prospective studies that evaluate the impact of modifying various contextual factors, while holding others constant, can provide deeper insights into the mechanisms by which context impacts clinical reasoning performance assessment.

Keywords: assessment; clinical reasoning; situated cognition.

Introduction

Assessment of clinical reasoning performance is challenging because clinical reasoning is a complex, multi-dimensional construct and is context-specific [1–3]. We define it as the cognitive and physical processes that emerge as a healthcare professional consciously and unconsciously adapts to interactions with the patient and environment to solve problems and make decisions by collecting and interpreting patient data, predicting potential outcomes, weighing the benefits and risks of actions, and accounting for patient preferences to determine a working diagnostic and therapeutic management plan to improve a patient's well-being [4]. Studies have shown that significant variation in clinical reasoning performance exists, even when doctors are tested on *identical* cases on different occasions, suggesting that subtle changes in context may prevent physicians from transferring their cognitive approach to the subsequent occasion [2]. This finding, named "context specificity", undoubtedly plays a role in the current crisis in medical error in the US [5, 6]. The mechanisms which lead to context specificity remain uncertain [7–9]. Traditionally, information processing theories [10], which emphasize knowledge acquisition and organization, have provided valuable insights into understanding clinical reasoning, but these do not naturally lend themselves to

*Corresponding author: Joseph Rencic, MD, Department of Medicine, Boston University School of Medicine, 72 East Concord Street, Boston, MA 02118, USA, Phone: +1 (617) 358-7468, Fax: +1 (617) 358-7469, E-mail: jrencic@bu.edu. https://orcid.org/0000-0002-2598-3299
Lambert W.T. Schuwirth: Prideaux Centre for Research in Health Professions Education, Flinders University, Flinders, Australia
Larry D. Gruppen: Department of Medical Education, University of Michigan, Ann Arbor, MI, USA
Steven J. Durning: Department of Medicine, Uniformed Services University of the Health Sciences, Bethesda, MD, USA

the exploration of context. Thus, some researchers have recently sought alternative theories that could be valuable in understanding the impact of context on clinical reasoning performance and its assessment.

Situated cognition theory, which posits that cognition emerges from the complex interplay of human beings with each other and the environment, has been used as a valuable conceptual framework for exploring context specificity [7]. Durning et al. [11] developed a situated cognition model of clinical reasoning performance focused on physician, patient, and environmental factors. They demonstrated that altering patient factors (e.g. patient does not speak English) and environmental factors (e.g. time constraints) negatively impacted clinical reasoning performance as assessed by a post-encounter write-up score [12]. Furthermore, Kogan et al. [13] performed semi-structured interviews of faculty after they observed and assessed videotaped clinical encounters of standardized residents with standardized patients. They qualitatively analyzed the transcriptions of these interviews which led to the development of a situated cognition model of rater cognition in direct observation. The model highlighted the impact of clinical competence, personality factors, emotions, the clinical system, and institutional culture on the assessment of learners. Thus, evidence is emerging that situated cognition theory can aid in the understanding of context specificity.

On the basis of these studies [12, 13], we developed a simplified conceptual framework for clinical reasoning performance assessment which expands upon this work. It consists of six clinical reasoning performance assessment elements, including the clinician (termed assessee), patient, rater, assessment method, task, and environment (Figure 1). We hypothesized that applying the lens of situated cognition theory and this conceptual framework to analyze the results of articles that we culled from a systematic search of the clinical reasoning performance assessment literature would provide additional insights into the outcomes of these studies. Our aim was to demonstrate that situated cognition theory and our model specifically could serve as an explanatory framework for understanding context specificity, which could in turn influence research design and educational practice.

Methods

Our research group conducted a systematic search of the literature from 1946 to 2016 in a broad range of databases, including Ovid MEDLINE, CINAHL, ERIC, PsycINFO, Scopus, Google Scholar, and the New York Academy of Medicine Grey Literature Report, to determine the scope of the clinical reasoning assessment literature. Based on this search, a scoping review was published which provides additional details of our methodology [14]. The extraction form used for that review included a question about whether an article explicitly or implicitly referenced specific theories relevant to clinical reasoning. For this paper, we reviewed 44 of the articles from that review, which at least one full-text article reviewer believed explicitly or implicitly used a situated cognition framework. We used
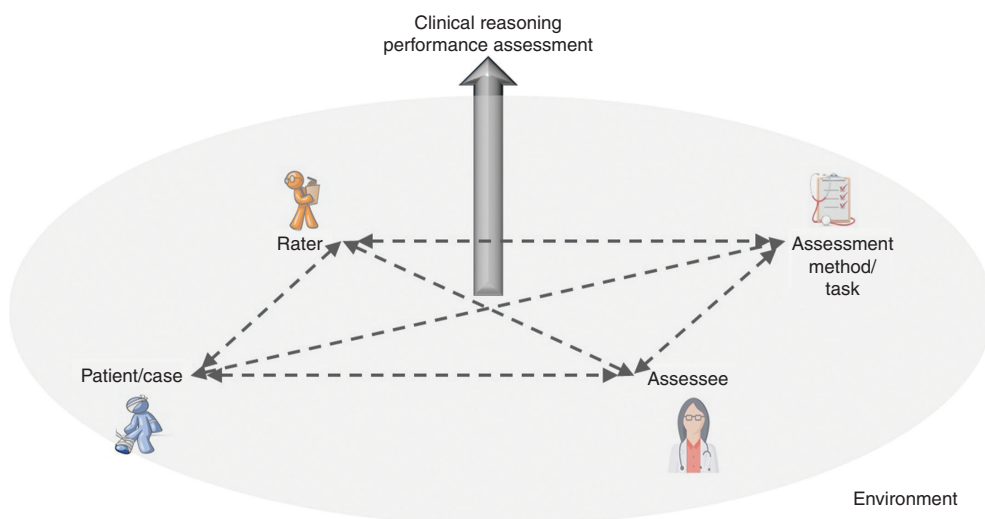

Figure 1: Assessee, patient, rater, environmental, assessment method, and task interactions lead to the emergence of clinical reasoning performance assessment.

this approach to maximize the potential for finding manuscripts that demonstrated interactions between different elements of clinical reasoning performance assessment. Of these 44 articles, two authors (JR and SJD) selected 17 that furnished the most valuable insights into the utility of situated cognition theory in understanding some of the contextual factors that impact the assessment of clinical reasoning performance.

Results

Using our conceptual model (Figure 1), we categorize articles based on the contextual factors that interact with assessees to influence their clinical reasoning performance and its assessment. We present the following dyadic interactions: assessee-case/patient, assessee-environment, assessee-assessment method, assessee-task, and assessee-rater. Definitions of these and other key terms may be found in the Table 1 glossary. We will then discuss a few articles that demonstrate triadic or tetradic contextual factor interactions.

Assessee-case/patient interaction

The "case" element in a clinical reasoning performance assessment method typically refers to the stimulus format, the way a problem is presented to an assessee. Stimulus formats include written or computer-based vignettes, low or high-fidelity simulation, and standardized or real patients. A dramatic study by Nendaz et al. [15] demonstrated the significant impact that the stimulus format can have on clinical reasoning performance assessment. Assessees (n = 91) that were randomized to receive only a chief complaint and had to ask questions to obtain the history and physical exam findings to determine the diagnosis had significantly lower diagnostic accuracy at all levels of experience (students, 10% versus 36%; residents, 47% versus 81%; and general internists, 59% versus 100%) as compared with assesses (n = 64) who received a complete clinical vignette. All the assessees who only received the chief complaint obtained more data than that provided in the vignette but many failed to obtain the key information required to make the correct diagnosis. The authors speculated that assessees receiving the vignette more readily generated the correct hypothesis (i.e. the vignette triggered more relevant hypothesis generation).

Similarly, a systematic review of clinical reasoning performance showed that certain stimulus formats can alter physician behaviors by causing increased data collection [16]. When physician clinical reasoning performance on similar medical problems was assessed on paper cases versus real patient cases through chart audits, some researchers found a significant increase in data collection (e.g. lab testing) in paper case formats [16]. Several study design elements appeared important in leading to differences in clinical reasoning performance between the two stimulus formats. These included: (1) limited instructions about the nature of the task, (2) how closely the task on the paper cases mirrored real-life clinical reasoning tasks, (3) differences in the response formats (see Table 1 for definition) between the paper cases and the actual clinical documentation, and (4) differences in data analysis of the clinical reasoning performance of interest [16].

From a traditional information-processing perspective, the clinical reasoning performance assessment variation seen in these studies does not make sense. The performance should not be impacted by the stimulus format, given that the clinical content and knowledge required was the same in both formats. On the other hand, situated cognition theory provides a clear framework for understanding these performance and assessment differences because it accounts for contextual differences, such as changes in stimulus format. By comparing two different stimulus formats, both studies showed that assessee-case format interactions have the potential to significantly alter clinical reasoning performance assessment. The situated cognition perspective provided by these results raises the possibility that the traditional approach of vignette-based stimulus formats, which do not require whole task clinical reasoning performance, may overestimate assessee performance.

Assessee-environment interactions

Environmental factors can have a significant impact on assessee performance. For example, in emergency room settings, the multitude of patients, task-shifting requirements, and high stress levels can increase the chances of poor clinical reasoning performance and diagnostic error. In their study of assesse-environment interactions, Hughes and Young [17] demonstrated that contextual factors, such as hospital setting, influenced nurses' clinical reasoning processes, highlighting the impact of environmental factors on clinical reasoning performance assessments. From an information processing view, these outcomes may seem surprising because health professionals aspire to "objective", evidence-based clinical reasoning. The situated cognition view recognizes that nurses'

Table 1: Glossary of key terms.

Category	Term	Definition
Clinical reasoning	Clinical reasoning	The cognitive and physical processes that emerge as a healthcare professional consciously and unconsciously adapts to interactions with the patient and environment to solve problems and make decisions by collecting and interpreting patient data, predicting potential outcomes, weighing the benefits and risks of actions, and accounting for patient preferences to determine a working diagnostic and therapeutic management plan to improve a patient's well-being
	Clinical reasoning performance	The observable activities of clinical reasoning, including all those mentioned in the previous definition
	Part-task clinical reasoning performance	The observable activities of part- or sub-tasks of clinical reasoning performance, such as data gathering, problem representation, diagnosis, or treatment
	Whole task clinical reasoning performance	The observable activities of the entire task of clinical reasoning (i.e. all the steps up to and including diagnosis and treatment)
Assessment	Clinical reasoning performance assessment	The assessment of whole task or part-task clinical reasoning performance through one or more methods
	Response format	The type of "answer" required of the assessee for a given assessment method. Examples include selected responses (e.g. multiple-choice options) and constructed responses (e.g. post-encounter form requiring differential diagnosis and diagnostic justification).
	Scoring activity	The process by which an assessment method produces a judgment of clinical reasoning performance
	Stimulus format	The manner in which a problem is presented to an assessee (e.g. the "case") that initiates clinical reasoning performance. Examples include actual patients, computer- or paper-based simulated patients, or brief clinical questions (e.g. "What radiology study would be most likely to diagnose cholecystitis?"
Expertise	Deliberate Practice	Systematic, purposeful, intense, focused effort to improve specific sub- and whole-task performance of a specific ability using timely and accurate feedback to calibrate performance
	Mastery learning	A type of learning that requires learners to acquire essential knowledge and skill, measured rigorously against fixed achievement standards, without regard to the time needed to reach the outcome
Information processing	Information-processing theory	A theory that uses the analogy of a computer for the brain, focusing on the way in which the brain processes information to effectively act within its environment
Situated cognition	Situated cognition theory	A theory that posits cognition emerges from the complex interplay of human beings with each other and the environment
	Context specificity	The phenomenon that an individual's performance on a particular problem in a particular situation is only weakly predictive of the same individual's performance on an identical problem in a slightly different situation
	Contextual factors	Elements of a situation that can impact clinical reasoning performance and its assessment
	Assessee factors	The factors that relate to or emerge from a learner being evaluated in a given situation that can impact clinical reasoning performance and its assessment (e.g. fatigue, physical appearance)
	Assessment method factors	The factors that relate to or emerge from the selected assessment method in a given situation that can impact clinical reasoning performance and its assessment (e.g. time constraints, selected multiple-choice responses)
	Environmental factors	The factors that relate to or emerge from the environment in the broadest sense in a given situation that can impact clinical reasoning performance and its assessment (e.g. noise, institutional culture)
	Patient factors	The factors that relate to or emerge from a patient in a given situation that can impact clinical reasoning performance and its assessment (e.g. fatigue, stress)
	Rater factors	The factors that relate to or emerge from a rater in a given situation that can impact clinical reasoning performance and its assessment (e.g. bias, expertise)
	Task factors	The factors that relate to or emerge from a task in a given situation that can impact clinical reasoning performance and its assessment (e.g. authentic documentation, physical examination)

interactions with their hospital environment shape their clinical reasoning performance and its subsequent assessment. These types of cultural environmental interactions, which can occur in any inpatient or outpatient setting, are rarely considered in the assessment of clinical reasoning performance because these assessments often occur within single institutions.

Assessee-method interactions

Our systematic search [14] revealed a broad range of clinical reasoning assessment methods (Table 2). Assessment methods or tools are usually characterized by one specific response format (e.g. multiple-choice questions), but some allow for the use of a variety of response formats (e.g. a written note to assess diagnostic justification or learner performance assessed by a standardized patient checklist). The target goal of assessment and feasibility issues typically determine the method and response format selection. From a situated cognition view, increasing the quantity of contextual factors leads to more interactions with the assessee's and rater's thinking processes which may potentially lead to greater variation in performance assessment. We will highlight the impact of specific aspects of assessment methods on clinical reasoning performance assessment in subsequent sections.

Table 2: Major clinical reasoning assessment methods reviewed.

– Biologic (e.g. functional MRI)
– Chart stimulated recall
– Comprehensive integrative puzzle
– Concept maps
– Direct observation (e.g. mini-CEX, clinical examination exercise)
– Extended matching questions
– Free-text responses/short essay
– Global assessment
– Simulation with technology (simulation)
– Key features testing
– Multiple-choice questions
– Objective structured clinical examination (OSCE)
– Oral case presentation
– Oral examination
– Patient management problem
– Script concordance testing
– Self-regulated learning/microanalytic techniques (SRLMAT)
– Think aloud
– Written notes (charted documents, admission notes)

MRI, magnetic resonance imaging.

Assessee-response format interactions: cueing effects

In medical education, cueing effects refer to the impact that the presentation of lists of answer choices has on assessee performance. McCarthy demonstrated that such lists increase both the number of selected answers (when allowed) and the correctness of the answer [18]. Three studies [18–20] demonstrated a significant increase in data collection in the patient management problem format due to the lists of options visible to examinees (i.e. "cueing"), as compared with chart audit, oral examination, or simulated patient exercises. In a study comparing data collection in standardized patient encounters versus patient management problems, differences in response format accounted for 36% of performance variation, while performance differences between students only accounted for 24% [19].

Schuwirth et al. [21] noted that cueing effects have their greatest impact on diagnostic questions and deconstructed the cueing effect on diagnostic accuracy using constructed response items (i.e. open-ended questions) followed by selected response items (i.e. four- to eight-option multiple-choice questions) for the same case. Positive cueing (i.e. improved diagnostic accuracy) occurred in about 14% and negative cueing (i.e. worsened diagnostic accuracy) in about 6% of multiple-choice questions. Positive cueing occurred with harder items while negative cueing occurred with easier items. More advanced practitioners (e.g. physicians) were less impacted by cueing effects of multiple-choice questions than students.

These studies demonstrate the dramatic impact that assessment method can have on clinical reasoning performance assessment. The situated cognition perspective provides deeper insights into the nature of knowledge and cueing effects because it explains that knowledge is not a static entity that exists in an assessee's mind. Rather, knowledge emerges from the interaction of the assessee with contextual factors. Certain contextual factors are more or less likely to lead to the emergence of knowledge that results in acceptable clinical reasoning performance (e.g. multiple-choice questions cue the mind to recognize the correct option in harder questions). Positive cueing in difficult questions may relate to diagnostic hypothesis generation, a clinical reasoning subtask that multiple-choice questions allow assessees to bypass. Experts who have more knowledge appear to be less easily swayed by distractors than novices, although even they are prone to negative cueing effects on easier questions. This insight has important implications for

how educators choose assessment methods to prove assessee competence.

Assessee-scoring activity: sequence of data collection

Scoring activity refers to the process by which an assessment method produces a judgment of performance. Typical scoring activities include marking multiple-choice questions correct or incorrect, directly observing clinical performance by global (i.e. gestalt) assessment, and assessing the quality of diagnostic justification provided in a written note. Clearly, such activities differ significantly from one another in terms of the quantity and quality of interactions occurring between assessees and contextual factors. LaRochelle et al. [22] created a unique measure of data collection focused on the coherence or logical sequencing of videotaped student history-taking in a standardized patient encounter. Two study authors determined an ideal approach for the sequence of critical historical questions and the sequence of history questions for a given global score was then processed into an average ordering matrix, and the probabilities for any key data factor following another were calculated. The probabilities were then added together to create a coherence score which was compared to faculty assessors' global ratings for these encounters. For higher average ordering matrix numbers, higher coherence scores demonstrated moderate associations with global ratings and aided in discriminating among students in the highest global rating score category. For low scores on the average ordering matrix, a weak correlation between coherence and checklist scores existed (approximate $r = 0.1–0.3$). For high scores on the average ordering matrix, moderate correlations were seen between coherence and checklist scores (approximate $r = 0.5$).

The correlation of history-taking sequence with global assessment again provides support for the value of a situated cognition perspective for clinical reasoning performance assessment. LaRochelle et al. [22] went beyond standard knowledge/process checklists to study the dynamic interactions of assessees with patients in collecting history. This approach demonstrates the value of situated cognition in clinical reasoning performance assessment because it highlights how the study of interactions between contextual factors (i.e. the assessee and standardized patient) can increase the capability to discriminate between student performances. The use of a situated cognition perspective in this study led to a novel scoring activity which measured the sequence and

relationship of questions asked during history-taking. Although not every institution can replicate the statistics of this approach, most could incorporate assessment items focused on the logical sequencing and organization of students' histories into their objective structured clinical examinations.

Assessee-task effects: the intermediate effect

The tasks of clinical reasoning can consist of outcomes, specifically a diagnosis or a treatment, and/or processes that generically consist of data collection, problem representation, and hypothesis generation. For a more comprehensive list of potential clinical reasoning tasks, we recommend reviewing Table 3 [23, 24]. These tasks occur within a spectrum of authenticity with standardized written examinations representing the less authentic pole (i.e. in vitro assessment) and direct observation representing the more authentic pole (i.e. in vivo assessment). Clinical reasoning activities range from part tasks, such as a multiple-choice question that asks what historical question is most likely to aid in a patient's diagnosis, to whole tasks, such as a directly observed admission of a complicated intensive care unit patient.

Inaccurate conclusions about clinical reasoning performance can be caused by assessee interactions with certain tasks related to clinical reasoning response formats. For example, advanced students and/or residents score higher than clinical practitioners in process-focused assessments [e.g. objective structured clinical examination (OSCE) standardized patient checklist] that reward examinees for comprehensive recall of patient's clinical findings [25]. This "intermediate effect" likely relates to the fact that experienced clinicians make diagnoses through System 1 (e.g. pattern recognition) rather than a System 2 (e.g. analytic) approach. On the other hand, assessment tasks that measure more authentic clinical tasks, such as diagnostic accuracy, will lead to higher clinical reasoning performance scores for expert clinicians. The intermediate effect again demonstrates how contextual factor interactions, in this case the assessee with the task, artifactually lead to worsened clinical reasoning performance scores.

Assessee-rater effects

In clinical reasoning assessment methods, such as OSCEs, direct observations, and global rating scores, raters play

Table 3: Tasks of clinical reasoning.

Framing the encounter
1. Identify active issues
2. Assess priorities (based on issues identified, urgency, stability, patient preference, referral question, etc.)
3. Reprioritize based on assessment (patient perspective, unexpected findings, etc.)
   a. *Consider the impact of prior therapies*

Diagnosis
4. Consider alternative diagnoses and underlying cause(s)
   a. *Restructure and reprioritize the differential diagnosis*
5. Identify precipitants or triggers to the current problem(s)
6. Select diagnostic investigations
7. Determine most likely diagnosis with underlying cause(s)
8. Identify modifiable risk factors
   a. *Identify non-modifiable risk factors*
9. Identify complications associated with the diagnosis, diagnostic investigations, or treatment
10. Assess rate of progression and estimate prognosis
11. Explore physical and psychosocial consequences of the current medical conditions, or treatment

Management
12. Establish goals of care (treating symptoms, improving function, altering prognosis or cure; taking into account patient preferences, perspectives, and understanding)
13. Explore the interplay between psychosocial context and management
14. Consider the impact of comorbid illnesses on management
15. Consider the consequences of management on comorbid illnesses
16. Weigh alternative treatment options (taking into account patient preferences)
17. Consider the implications of available resources (office, hospital, community, and inter- and intra-professionals) on diagnostic or management choices
18. Establish management plans (taking into account goals of care, clinical guidelines/evidence, symptoms, underlying cause, complications, and community spread)
19. Select education and counseling approach for patient and family (taking into account patients' and their families' levels of understanding)
20. Explore collaborative roles for patient and family
21. Determine follow-up and consultation strategies (taking into account urgency, how pending investigations/results will be handled)
22. Determine what to document and who should receive the documentation

Self-reflection
23. Identify knowledge gaps and establish personal learning plan
24. Consider cognitive and personal biases that may influence reasoning

From: Goldszmidt M, Minda JP, Bordage G. Developing a unified list of physicians' reasoning tasks during clinical encounters. Acad Med 2013;88:390–4. Italicized tasks from: McBee E, Ratcliffe T, Goldszmidt M, et al. Clinical reasoning tasks and resident physicians: what do they reason about? Acad Med 2016;91:1022–102.

a critical role. The exact extent of their impact on holistic clinical reasoning performance assessment, though uncertain, is undeniable. Nevertheless, significant evidence exists regarding the impact of rater variation on clinical reasoning sub-task performance, such as data collection. Rater effects may account for up to 34% of performance variance on mini-clinical examination exercises (CEXs) with an additional 7% and 4% of the variance associated with rater-case effects and rater-trainee interactions, respectively [26]. Both construct-relevant (i.e. content knowledge) and construct-irrelevant factors (i.e. verbal style and dress) can impact rater stringency [27]. In addition, as compared with independent raters, teachers scored their students higher on a script concordance test (72 vs. 76.3, $p < 0.01$), the primary outcome, although relative ranking remained the same and the small absolute difference seems clinically insignificant [28].

Clearly, information-processing theory is critical to understanding a significant portion of rater-based variation seen in clinical reasoning performance assessment, because raters' clinical and assessment knowledge and experience is the basis for their judgments. However, the situated cognition perspective provides insights into construct-irrelevant contextual factors, such as gender and dress, which do not relate to rater knowledge and experience. Furthermore, situated cognition theory illuminates potential modifications to assessment methods. For example, a separate rater assesses a transcript of an OSCE encounter blinded to assessee contextual factors, such as gender, to reduce

the risk for rater bias. Kogan et al. [13] have created a conceptual model of rater cognition highlighting the influence of clinical competence, emotions, personality traits, frame of reference, clinical systems, and institutional culture, but more research is required to understand the impact of these on clinical reasoning performance assessment.

Multi-component interactions

Assessee-case-environment interactions

Two articles highlight how case and environment interact with the assessee to impact clinical reasoning performance and its assessment. A study [29] of nurses' ability to use vital sign cues to categorize patients as high or low risk for deterioration compared their accuracy using identical clinical scenarios presented either as written vignettes or high-fidelity simulation. The authors used judgment analysis, an approach which applies the Lens Model theory [30]. Briefly, the authors used logistic regression analysis to retrospectively determine the value of various clinical variables to predict the acute deterioration of real patients. The results of this regression analysis then became the gold standard by which study nurses' predictions about the risk of the simulated patients deteriorating were judged. This type of representation of clinical reasoning is called "paramorphic" because it predicts the results of judgments even though clinicians do not actually perform such a detailed mental mathematical calculation (i.e. Bayesian reasoning). For both experienced and student nurses, nursing judgment accuracy, r_a, was statistically significantly lower for the high-fidelity simulation [mean r_a 0.502, standard deviation (SD) 0.15] as compared to paper cases (mean r_a 0.553, SD 0.14, p = 0.007) [29].

This difference in judgment accuracy stemmed from less-accurate utilization of cues in the high-fidelity simulation cases as compared with paper cases, despite the presentation of identical information. The Lens Model or other knowledge-based theories such as information processing do not provide obvious insights into what caused less-accurate utilization of the high-fidelity simulation cues. On the other hand, situated cognition theory emphasizes how multiple factors may have impacted cognition negatively. The patient/case and tasks were clearly different given that the nurses in the high-fidelity simulation had to consciously recognize the signal (i.e. the key cues/data) among the significant noise of a simulated environment as opposed to reading the cues from a paper case. In addition, the motivation, emotions, and stress of high-fidelity simulation differ significantly from a low-fidelity environment. A situated cognition view could have helped the researchers consider additional hypotheses and develop a study design better equipped to explore the causes of the differences in the assessments of the nurses' clinical reasoning performance. For example, a third arm of the study using standardized patients could have been developed which might have demonstrated improved performance by experienced nurses who might more readily detect key signals and clinical findings in an actor than a mannequin.

Similarly, a web-based simulated patient study demonstrated that second-year medical students from the class of 2010 (n = 155) with unlimited time in a pass-fail examination requested half the number of history and physical examination inquiries (i.e. 100–150) and achieved 91% diagnostic accuracy (average time to exam completion 4 h, 13 m) as compared with students from the class of 2012 (n = 175) taking a time-constrained (3-h time limit), stratified examination (i.e. honors, high pass, pass, fail grade) (i.e. average time to exam completion 2 h, 1 m) who requested 250–300 inquiries in a stratified grades examination but only achieved 31% diagnostic accuracy [31]. Despite being informed that efficiency (# of required inquiry items requested/total # of inquiry items requested) impacted scoring, students in the time-constrained situation asked far more questions than the non-time-constrained students.

This example again highlights the value of a situated cognition perspective in understanding the powerful impact of contextual factors on clinical reasoning performance assessment. Test format modifications from untimed to timed and pass/fail to stratified reduced the diagnostic accuracy rate to 31%. It is improbable that the entire class of students in the time-constrained test condition had 61% poorer diagnostic ability on identical knowledge content than their colleagues from a class 2 years earlier. Because of the stress produced by time-constraint and the stratified examination and its interaction with case format [32], students asked nearly twice as many inquiries of the simulated patient than their unstressed colleagues, even when they were explicitly told that such behavior would negatively impact their grade. This quasi-experimental study highlights the importance of using a situated cognition framework in developing clinical reasoning performance assessments because contextual factors can have a profound impact on assessment outcomes.

Assessee-case-environment-method interactions

Durning et al. [12] studied the impact of contextual factors on diagnostic performance of general internists on video vignettes that portrayed different chief complaints. Each vignette varied based on patient factors [i.e. diagnostic suggestion, an atypical presentation, English as a second language (ESL), and/or emotional volatility] or environmental factors (e.g. time constraints) on a constructed written response to a validated clinical reasoning post-encounter form [33] as well as a subsequent verbal think-aloud protocol. Analysis of covariance provided support for the mechanisms of cognitive load and time constraints negatively impacting post-encounter form scores. The authors concluded that cognitive load, wherein the interactions of the clinician with multiple elements within a task can increase demands on working memory and hence worsen performance, provides an explanatory model for their findings [12]. Likewise, a study of oral examinations, where performance scores emerging from the interactions of assessees, raters, method, and task, demonstrated that scoring leniency increased over the course of the day and on succeeding examination days [34].

In both these studies, researchers varied or analyzed non-traditional contextual factors to see how they impacted clinical reasoning performance assessment. A situated cognition framework makes such research questions obvious because clinical reasoning performance and its assessment performance are recognized as context dependent rather than independent.

Assessee-environment-method-task interactions

The time allowed to perform a clinical reasoning task clearly impacts the assessment of clinical reasoning performance. Time limitation to complete a post-encounter form accounted for 4–16% of the variation in assessments of clinical reasoning performance, with the largest variance seen in the task of providing supporting evidence for the diagnosis [12]. Although an information-processing view certainly recognizes time as a critical contextual factor in assessment, a situated cognition perspective generalizes this perspective to non-traditional contextual factors (e.g. institutional culture, patient's poor health literacy) and encourages educators to consider assessment tasks as "contextual factor(s)"-sensitive or -insensitive and modify them as needed. For example, given that most clinicians practice in an "open-book" environment, a situated

cognition view might lead to a greater focus on open-book assessments to determine assessee competence.

Assessee-method-rater interactions

Durning et al. [35] compared a post-encounter form (PEF) to standardized patient checklist, oral presentation scores, and scores on standardized written multiple-choice question examinations. In univariate analyses, they found low correlations of the PEF across different methods: total oral presentation r 0.17, total SP checklist r 0.20, and National Board of Medical Examiners (NBME) shelf exam 0.23. The causes of these low correlations are uncertain but potentially include the following: (1) rater effects given that a different rater scored each method, (2) attenuation effects due to measurement error, or (3) measurement of a different construct. The third seems likely to be a significant contributor given the different tasks required of the assessee in each method.

The situated cognition perspective absorbs these results readily. Different assessment methods are contextual factors that interact with learners and raters in unique ways, leading to variations in clinical reasoning performance assessment and decreased correlations across methods.

Discussion

Understanding clinical reasoning assessment through contextual factor interactions

Applying the lens of situated cognition theory to the selected studies reveals how contextual factors can interact with each other in complex ways to alter clinical reasoning performance and its assessment. Medical educators or researchers who do not account for such interactions may misinterpret assessment outcomes (e.g. most students cannot perform diagnostic reasoning) [31]. Our situated cognition-based conceptual model (Figure 1) provides a simplified set of contextual factors for researchers to consider as they design research studies to assess the impact of contextual factors and for medical educators as they develop programs of clinical reasoning performance assessment. It complements other valuable theoretical frameworks relevant to clinical reasoning performance and its assessment (e.g. information processing, self-regulated learning, and motivation theories).

Potential applications of situated cognition theory and the model in clinical reasoning assessment

The conceptual model encourages educators to consider how contextual factors impact clinical reasoning assessment (Table 4). These factors can be modified to achieve various assessment goals. For example, efficiency in a diagnostic task is an important aspect of competence in

residents, so patient complexity could vary with the duration of a clinical reasoning OSCE station to assess competence in a time-constrained environment. For example, residents could be provided an evaluation time of 15 min for diagnosis of a complex case with an atypical disease presentation versus 8 min for an uncomplicated, typical disease presentation.

An additional value of the model is conceptualizing rater variability as "context specificity for raters".

Table 4: Examples of impact of conceptual model on assessment strategies.

Interaction type	Contextual factor	Contextual factor impact	Modifications
Assessee-Assessment method	Checklist-based OSCE assessment	Intermediate effect	Outcome-based as well as process-based assessment
	Closed-book assessments	Negatively impacts students who have difficulty with memorization	Open-book testing allowing time to look up needed information prior to answering, mimicking actual physician practice
	Multiple-choice questions	Cueing effect leading to changes in assessee responses	Constructed response items (i.e. free text answers)
	Scoring activity	Increased or decreased discrimination of student performance	Assess clinical reasoning performance tasks such as diagnostic efficiency and justification, history and physical examination sequence/ organization of data collection, in addition to prioritized differential diagnosis
Assessee-case/patient	Clinical vignette case format	Framing effect	Assessee obtains patient data
	Frustrated patient	Stress	Monitor learner stress, modify stress based on assessee developmental level
Assessee-environment	Cultural norms	Culture-concordant decision-making	Evaluate differences in assessment across different institutions
	High-fidelity simulation (e.g. mannequin)	Increased motivation, stress	
	Multi-patient environment (e.g. ED)	Frequent task-shifting, increased cognitive load	Increase or decrease task-shifting based on cognitive load
	Time constraints	Decreased diagnostic accuracy	Increase time allowance
Assessee-task	Data collection via simulated paper or computer-based case	Cueing effect leading to increased data collection	Provide clear instructions regarding the goals of data collection (e.g. thoroughness, efficiency)
Rater-assessee	Assessee appearance	Construct-irrelevant bias	Blinded post-encounter form or OSCE transcript assessment
	Rater clinical competence	Greater or lesser leniency	Rater faculty development, direct observation "experts"
	Rater teaching relationship with assessee	More lenient grading	Blinded assessments or rater selection based on no or minimal relationship with student
Rater-environment	Duration of rating time	Increased leniency over the course of a day	No rater observes for more than 4 h

OSCE, objective structured clinical examination.

Assessment by raters is analogous to the diagnostic process in clinical reasoning; it is a categorization task (e.g. is the learner in the "excellent", "good", or "poor" performance category). Thus, it is not surprising that clinical reasoning performance assessment would demonstrate similar context specificity [36]. This model may be helpful in providing a framework for further explaining the black box of context specificity in clinical performance assessment [13]. We believe that medical educators familiar with the model will more easily recognize opportunities for extending clinical reasoning performance assessment beyond assessee and case factors only.

Practical implications

The phenomenon of context specificity and the complexity of clinical reasoning performance assessment have implications for medical educators who are responsible for passing/failing standards for assessees. Clinical reasoning performance variation due to context specificity requires that multiple assessments across various problems in diverse settings are essential to obtaining valid information. The conceptual model also supports the push for increased quantity and quality of workplace-based assessments that has been trumpeted by competency-based medical education proponents [37], because these assessments measure situated clinical reasoning performance in context-rich environments that provide insights into how assessee's perform in real clinical encounters. High-volume workplace-based observations by multiple raters have the potential to overcome the lack of standardization and rater variability that plague workplace-based assessment because large sample sizes will likely lead to accurate overall assessments of clinical reasoning performance. Such an approach has been employed successfully by at least one institution [38].

Although the previously described psychometric approach has been used with some success in dealing with the challenges of workplace-based summative assessment, our conceptual model depicts the significant challenges facing raters given the contextual complexity and task-shifting that is required of them. An alternative approach to increasing the validity of clinical reasoning assessments is to emphasize deliberate practice [39] and mastery learning [40] for raters (Table 1).

Society entrusts health professions institutions to ensure the quality of their graduates. When framed this way, the deliberate development and maintenance of a professional direct observation "team" for every health care institution to improve the validity of workplace-based

assessments seems reasonable and appropriate. One key criterion for selecting raters is that they have excellent clinical skills and reasoning ability as research has demonstrated that rater clinical skills play an important role in assessment variation [41]. Unfortunately, because many schools and training programs have difficulty obtaining adequate direct observations, they often accept any rater that is willing or is able to perform the task. This situation may change as coaching programs become more common in medication education [42].

Limitations

Although we systematically searched the literature, some articles may have been missed, particularly non-English language articles. We did collect additional articles through hand searching references from selected seminal papers to maximize the broadness of our search. Our last literature search occurred in 2016, so this review may be missing some newer relevant articles that shed additional light on the value of a situated cognition perspective in clinical reasoning performance assessment. Although we defined situated cognition theory for the reviewers in our group, their understanding of the term likely varied significantly and may have led to the misidentification or lack of recognition of articles that implicitly referenced situated cognition. Nevertheless, we found an adequate number of articles that provided valuable insights into the impact of contextual factors on clinical reasoning performance assessment. We chose a narrative review format because it allowed for a selection of articles that best exemplify the insights that situation cognition theory provides into clinical reasoning assessment performance. We did not feel that the narrative review format was a significant limitation given the purpose of trying to highlight the potential value of a situated cognition perspective for understanding variation in the assessment of clinical reasoning performance (i.e. a proof of conceptual framework). An article focused on teaching clinical reasoning recently used a similar approach [43].

Future directions

Using situated cognition theory and our model, further research should explore potential interactions varying one or more of the contextual factors to determine their impact on clinical reasoning performance and its assessment. Clearly, the most context-rich research settings are real clinical encounters but the lack of standardization

and opportunities for manipulating specific contextual factor "variables" make unannounced standardized patient encounters, OSCEs, and technology-enhanced simulations more useful methods for studying clinical reasoning performance assessment at this time. Complex rater interactions with multiple contextual factors seem a particularly fertile ground for research, particularly given sophisticated statistical approaches such as network analysis with or without the use of artificial intelligence [44]. Although the interactions between contextual factors may be linear, it is more likely that they will be nonlinear (e.g. they emerge and the results do not approximate a straight line such as in a correlation or regression analysis). Thus, nonlinear statistical methods may be necessary to reveal these types of interactions [45].

In terms of assessment for learning, another critical area for consideration is learning transfer. Situated cognition theory posits that thinking and, by extension, learning emerges within contexts and little evidence exists to suggest that thinking or learning readily transfer to different contexts [3, 46]. Because students typically best learn the content covered on examinations (e.g. how to ace a multiple-choice question), clinical reasoning performance assessment for learning is a critical research domain. Further study is needed to compare situational, workplace-based assessments, especially in interprofessional teams, versus non-situational assessments (e.g. multiple-choice questions), as well as high-fidelity and low-fidelity versions of these assessments, to see if transfer of learning increases. Logistical challenges have typically been the rate-limiting step in developing these "instructional" assessment strategies; however, the accessibility of affordable virtual reality in the next 5–10 years provides the opportunity for scalable high-fidelity clinical reasoning assessments [47]. Research should explore rater performance, along with assessee performance, given the complexity of the task of rating. Delving more deeply into contextual factors and context specificity in these settings may provide further insights into the mechanisms of clinical reasoning and its assessment.

Finally, we believe that trans-theoretical models combining situated cognition as a broad theory to enhance situational understanding with other, more task-focused theories may provide further insights into context specificity and clinical reasoning performance and assessment variation. Two such theories, cognitive load [48] and self-regulated learning theory [49], seem particularly promising.

Conclusions

Our conceptual model based on situated cognition theory characterizes clinical reasoning and its assessment as complex phenomena that emerge from the interactions of contextual factors. Our review of the clinical reasoning assessment literature provides support for the value of this model in enhancing knowledge of causes of variation in clinical reasoning performance assessment and thereby, potentially improving the validity of such assessments. Clinical reasoning performance cannot be isolated from contextual factors because it is interconnected with them; context specificity lends statistical credence to this claim. A good analogy is the mind-body duality problem of philosophy. Although the concepts can be understood as distinct entities, philosophers now realize that such a division was artificial and led to significant misperceptions. A holistic view of clinical reasoning performance assessment does not prevent the manipulation of contextual factors (i.e. "variables") and such manipulation with one or more contextual factors can provide valuable insights into clinical reasoning performance [12]. A greater understanding of these interactions can improve educators' ability to provide both formative and summative clinical reasoning performance assessment. Future research using a situated cognition perspective combined with modern technologies, such as virtual reality, and novel statistical methods, such as nonlinear modeling, may enable us to inch closer to reaching the "Holy Grail" of valid situational clinical reasoning performance assessment [1].

References

1. Schuwirth L. Is assessment of clinical reasoning still the Holy Grail? Med Educ 2009;43:298–300.
2. Norman GR, Tugwell P, Feightner HW, Muzzin L. Knowledge and clinical problem-solving ability. Med Educ 1985;19:344–56.
3. Eva K. On the generality of specificity. Med Educ 2003;37:587–8.
4. Trowbridge RL, Rencic JJ, Durning SJ. Teaching clinical reasoning. Philadelphia, PA: Am Coll Physicians, 2015.
5. Medicine I of, editor. To err is human: building a safer health system [Internet]. Washington, DC: National Academies Press, 2000. Available from: http://informahealthcare.com/doi/abs/10.1080/1356182021000008364.
6. National Academies of Sciences, Engineering and M. Improving diagnosis in health care. Washington, DC: National Academies Press, 2016.
7. Durning SJ, Artino A. Situativity theory: a perspective on how participants and the environment can interact: AMEE guide no. 52. Med Teach 2011;33:188–99.
8. Kreiter CD, Bergus G. The validity of performance-based measures of clinical reasoning and alternative approaches. Med Educ [Internet] 2009;43:320–5.
9. Schauber SK, Hecht M, Nouns ZM. Why assessment in medical education needs a solid foundation in modern test theory. Adv Heal Sci Educ 2018;23:217–32.
10. Newell A, Simon HA. Human problem solving. [Internet]. Vol. 2, Contemporary sociology. Englewood Cliffs, NJ: Prentice-Hall, 1972. Available from: http://www.jstor.org/stable/2063712?origin=crossref.
11. 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.
12. 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.
13. Kogan JR, Conforti L, Bernabeo E, Iobst W, Holmboe E. Opening the black box of clinical skills assessment via observation: a conceptual model. Vol. 45, Medical education. England: Oxford, 2011:1048–60.
14. Daniel M, Rencic J, Durning SJ, Holmboe E, Santen SA, Lang V, et al. Clinical reasoning assessment methods: a scoping review and practical guidance. Acad Med 2019;94:902–12.
15. Nendaz MR, Raetzo MA, Junod AF, Vu NV. Teaching diagnostic skills: clinical vignettes or chief complaints? Adv Heal Sci Educ 2000;5:3–10.
16. Jones TV, Gerrity MS, Earp J. Written case simulations: do they predict physicians' behavior? J Clin Epidemiol 1990;43:805–15.
17. Hughes KK, Young WB. The relationship between task complexity and decision-making consistency. Res Nurs Health 1990;13:189–97.
18. McCarthy WH. An assessment of the influence of cueing items in objective examinations. Acad Med 1966;41:263–6.
19. Norman GR, Feightner JW. A comparison of behaviour on simulated patients and patient management problems. Med Educ 1981;15:26–32.
20. Goran MJ, Williamson JW, Gonnella JS. The validity of patient management problems. Acad Med 1973;48:171–7.
21. Schuwirth LW, van der Vleuten CP, Donkers H. A closer look at cueing effects in multiple-choice questions. Med Educ 1996;30:44–9.
22. LaRochelle J, Durning SJ, Boulet JR, van der Vleuten C, van Merrienboer J, Donkers J. Beyond standard checklist assessment: question sequence may impact student performance. Perspect Med Educ 2016;5:95–102.
23. Goldszmidt M, Minda JP, Bordage G. Developing a unified list of physicians' reasoning tasks during clinical encounters. Acad Med 2013;88:390–4.
24. McBee E, Ratcliffe T, Goldszmidt M, Schuwirth L, Picho K, Artino AR, et al. Clinical reasoning tasks and resident physicians: what do they reason about? Acad Med 2016;91:1022–8.
25. Schmidt HG, Boshuizen HP, Hobus PP. Transitory stages in the development of medical expertise: the "intermediate effect" in clinical case representation studies. In: Proceedings of the 10th annual conference of the cognitive science society. Hillsdale, NJ: Lawrence Erlbaum Associates, 1988:139–45.
26. de Lima AA, Conde D, Costabel J, Corso J, Van der Vleuten C. A laboratory study on the reliability estimations of the mini-CEX. Adv Heal Sci Educ 2013;18:5–13.
27. Burchard KW, Rowland-Morin PA, Coe NP, Garb JL. A surgery oral examination: interrater agreement and the influence of rater characteristics. Acad Med J Assoc Am Med Coll 1995;70:1044–6.
28. Charlin B, Gagnon R, Sauvé E, Coletti M. Composition of the panel of reference for concordance tests: do teaching functions have an impact on examinees' ranks and absolute scores? Med Teach 2007;29:49–53.
29. Yang H, Thompson C, Hamm RM, Bland M, Foster A. The effect of improving task representativeness on capturing nurses' risk assessment judgements: a comparison of written case simulations and physical simulations. BMC Med Inform Decis Mak 2013;13:62.
30. Brunswik E. The conceptual framework of psychology. Psychol Bull 1952;49:654–6.
31. Gunning WT, Fors UG. Virtual patients for assessment of medical student ability to integrate clinical and laboratory data to develop differential diagnoses: comparison of results of exams with/without time constraints. Med Teach 2012;34:e222–8.
32. Lee Y-H, Chen H. A review of recent response-time analyses in educational testing. Psychol Test Assess Model 2011;53:359–79.
33. Durning S, Artino Jr 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.
34. Colton T, Peterson OL. An essay of medical students' abilities by oral examination. J Med Educ 1967;42:1005–14.
35. 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;34:30–7.
36. Park B, DeKay ML, Kraus S. Aggregating social behavior into person models: perceiver-induced consistency. J Pers Soc Psychol 1994;66:437–59.
37. Holmboe ES. Realizing the promise of competency-based medical education. Acad Med 2015;90:411–3.
38. Warm EJ, Held JD, Hellmann M, Kelleher M, Kinnear B, Lee C, et al. Entrusting observable practice activities and milestones over the 36 months of an internal medicine residency. Acad Med 2016;91:1398–405.
39. Ericsson KA. An expert-performance perspective of research on medical expertise: the study of clinical performance. Med Educ 2007;41:1124–30.

40. McGaghie WC, Issenberg SB, Cohen ER, Barsuk JH, Wayne DB. Medical education featuring mastery learning with deliberate practice can lead to better health for individuals and populations. Acad Med 2011;86:e8–9.

41. Kogan JR, Hess BJ, Conforti LN, Holmboe ES. What drives faculty ratings of residents' clinical skills? The impact of faculty's own clinical skills. Acad Med 2010;85:S25–8.

42. Gifford KA, Fall LH. Doctor coach: a deliberate practice approach to teaching and learning clinical skills. Acad Med 2014;89:272–6.

43. Schmidt HG, Mamede S. How to improve the teaching of clinical reasoning: a narrative review and a proposal. Med Educ 2015;49:961–73.

44. Dias RD, Gupta A, Yule SJ. Using machine learning to assess physician competence: a systematic review. Acad Med 2019;94:427–39.

45. Durning SJ, Lubarsky S, Torre D, Dory V, Holmboe E. Considering "nonlinearity" across the continuum in medical education assessment: supporting theory, practice, and future research directions. J Contin Educ Health Prof 2015;35:232–43.

46. Lave J. Cognition in practice: mind, mathematics and culture in everyday life. New York: Cambridge University Press, 1988.

47. Moro C, Štromberga Z, Raikos A, Stirling A. The effectiveness of virtual and augmented reality in health sciences and medical anatomy. Anat Sci Educ 2017;10:549–59.

48. van Merrienboer J, Sweller J. Cognitive load theory in health professional education: design principles and strategies. Med Educ 2010;44:85–93.

49. Artino AR, Brydges R, Gruppen LD. Self-regulated learning in healthcare profession education: theoretical perspectives and research methods. Res Med Educ 2015;155–66.