Received: 2 December 2023 | Accepted: 7 March 2024

DOI: 10.1002/jhm.13335

ORIGINAL RESEARCH

Journal of Hospital Medicine
shm Society of Hospital Medicine

Characteristics differentiating problem representation synthesis between novices and experts

Casey N. McQuade MD, MS[1] | **Michael G. Simonson MD, MS**[1] | **Julia Lister MD**[2,3] | **Andrew P. J. Olson MD**[2,3] | **Laura Zwaan PhD**[4] | **Scott D. Rothenberger PhD**[1] | **Eliana Bonifacino MD, MS**[1]

[1]Division of General Internal Medicine, Department of Medicine, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, USA

[2]Division of Hospital Medicine, Department of Internal Medicine, University of Minnesota School of Medicine, Minneapolis, Minnesota, USA

[3]Division of Hospital Medicine, Department of Pediatrics, University of Minnesota School of Medicine, Minneapolis, Minnesota, USA

[4]Erasmus Medical Center, Institute of Medical Education Research Rotterdam, Rotterdam, the Netherlands

Correspondence
Casey N. McQuade, MD, MS, Department of Medicine, University of Pittsburgh School of Medicine, 200 Lothrop St, Suite G100, Pittsburgh, PA 15224, USA.
Email: mcquadec@upmc.edu;
Twitter: @CaseyMcQuadeMD

Funding information
Society to Improve Diagnosis in Medicine's Fellowship in Diagnostic Excellence; Gordon and Betty Moore Foundation

Abstract

Background: Formulating a thoughtful problem representation (PR) is fundamental to sound clinical reasoning and an essential component of medical education. Aside from basic structural recommendations, little consensus exists on what characterizes high-quality PRs.

Objectives: To elucidate characteristics that distinguish PRs created by experts and novices.

Methods: Early internal medicine residents (novices) and inpatient teaching faculty (experts) from two academic medical centers were given two written clinical vignettes and were instructed to write a PR and three-item differential diagnosis for each. Deductive content analysis described the characteristics comprising PRs. An initial codebook of characteristics was refined iteratively. The primary outcome was differences in characteristic frequencies between groups. The secondary outcome was characteristics correlating with diagnostic accuracy. Mixed-effects regression with random effects modeling compared case-level outcomes by group.

Results: Overall, 167 PRs were analyzed from 30 novices and 54 experts. Experts included 0.8 fewer comorbidities ($p < .01$) and 0.6 more examination findings ($p = .01$) than novices on average. Experts were less likely to include irrelevant comorbidities (odds ratio [OR] = 0.4, 95% confidence interval [CI] = 0.2–0.8) or a diagnosis (OR = 0.3, 95% CI = 0.1–0.8) compared with novices. Experts encapsulated clinical data into higher-order terms (e.g., sepsis) than novices ($p < .01$) while including similar numbers of semantic qualifiers (SQs). Regardless of expertise level, PRs following a three-part structure (e.g., demographics, temporal course, and clinical syndrome) and including temporal SQs were associated with diagnostic accuracy ($p < .01$).

Conclusions: Compared with novices, expert PRs include less irrelevant data and synthesize information into higher-order concepts. Future studies should determine whether targeted educational interventions for PRs improve diagnostic accuracy.

INTRODUCTION

Formulating a problem representation (PR) is regarded as a fundamental step in solving clinical problems and is a cornerstone of clinical reasoning education.[1–4] A PR transforms clinical findings into abstract concepts to guide hypothesis generation and the search for a candidate diagnosis.[1,5,6] While PRs are conceptualized as forming unspoken in an experienced clinician's mind, trainees practice how to represent clinical problems by writing their thoughts down or presenting them aloud on rounds.[1,4,7–9] A prototypical PR, sometimes called a summary statement or "one-liner," is an expression of the patient's demographics and pertinent comorbidities, the temporal pattern of illness, and a description of the observed clinical syndrome.[1,2,4] This three-part structure, based on extrapolations from earlier studies and expert consensus of what elements are needed to solve a case, provides a starting point from which trainees are taught to craft a differential diagnosis (Ddx) to guide the patient's diagnostic evaluation.[2–5]

Apart from these general structural recommendations, medical educators share little consensus on what details characterize "gold standard" PR.[8–10] Early studies suggested that the number of semantic qualifiers (SQs)—terms like "acute," "generalized," or "severe," which transform data into meaningful abstractions—used in a summary statement correlates with participants' level of expertise and with diagnostic accuracy.[5,6] However, an intervention training novices SQ use did not improve accuracy, suggesting that SQ use alone does not capture PR quality.[11] Smith et al. developed a rubric judging students' summary statements on five characteristics: SQ use, factual accuracy, transformation of information into abstract concepts, narrowing of the Ddx, and subjective global quality.[10] While inter-rater reliability for the entire rubric was acceptable, individual experts using the rubric diverged in how they judged global quality.[10] These findings suggest that even experts have different notions of what composes high-quality PR.

While a gold standard for PR could be established by experts via Delphi procedures or interviews, such results would be opinion-based and may not correlate with diagnostic performance.[12,13] An approach based on observing expert behavior in real-life or simulated situations may more likely ascertain what factors constitute expert-level PR.[12,13] Because the goal of understanding high-quality PR is supporting instruction for trainees, the natural comparison for expert PR creation is novice PR creation.

We sought to elucidate the characteristics that distinguish PRs created by attending physicians (termed "experts") from those created by physicians early in their training (termed "novices"). We also explored which characteristics correlated with diagnostic accuracy.

METHODS

Subjects and recruitment

We recruited internal medicine residents and inpatient teaching faculty from two academic institutions in different locations in the United States (site directors C.N.M. and J.L.). Residents met inclusion criteria if they had completed 10–15 months of post-graduate training. We included trainees at this level to lessen the impact of limited clinical experience on participants' abilities to diagnose the vignettes in the study. We operationalized expertise by experience level assuming that, on average, even less experienced attending physicians represent relative expertise in clinical medicine and PR compared with early residents.

At the University of Pittsburgh (UoP), resident clinical reasoning education includes required online modules covering basic concepts plus formal instruction on topics like PR during case conferences. At the University of Minnesota (UoM), resident clinical reasoning education includes a didactic on key concepts during their academic half-day schedule and practical instruction on PR during case conferences and morbidity and mortality conferences.

Recruitment occurred via email. Participants were given time to complete the study during regularly scheduled educational sessions. Participation was voluntary and incentivized by the chance to win a gift card.

This study was deemed exempt after IRB review at both institutions.

Study design

We presented study participants with two vignettes with an undisclosed diagnosis and asked them to write a "summary statement, also called a PR or a one liner" and provide an unranked three-item Ddx for each. No further instructions were given for what details to include or how to structure PRs. Participants could view the cases while completing these tasks.

Participants then rated their agreement with three statements on a 5-point Likert scale. These statements explored participants' diagnostic confidence and how writing a PR affected their differential. Finally, we collected demographic data for participants.

The vignettes contained ambiguity while also providing enough information to achieve a single diagnosis (Supporting Information: Digital Appendix 1). To standardize case difficulty and support statistical analysis, we constructed the vignettes with identical numbers of total historical elements (e.g., past medical history and surgical history) and pertinent positive and negative findings. The list of comorbidities for each case was crafted to contain two items relevant to the diagnosis and five irrelevant items. Vignettes were developed by two internal medicine faculty with clinical reasoning expertise (C.N.M. and E.B.) and were piloted by four chief medical residents. Vignettes took approximately 5–8 min each to complete. Case A described a patient presenting with right-sided heart failure from pulmonary hypertension and Case B a patient with sepsis from acalculous cholecystitis. We used two cases to decrease the likelihood of content-specific findings while balancing subjects' total participation time.

Qualitative data analysis

We qualitatively analyzed PRs using deductive content analysis to enable statistical analysis comparing the frequencies of characteristics by expertise level and by diagnostic accuracy. We developed an initial codebook based on characteristics of PRs studied in previous works[1–6,14–18] and iteratively refined it throughout the coding process (Supporting Information: Digital Appendix 2).

We coded a variety of PR characteristics, including structure (e.g., word count and sentence number), patient demographics (e.g., number and relevance of comorbidities), SQ use, degree of encapsulation, inclusion of components of a history and physical, and whether a diagnosis was included.

We identified SQs using the strategies outlined by Connell et al., which count only mutually exclusive semantic categories and avoid repetition.[15] For example, a summary including both "acute" pain and "subacute" dyspnea counted once toward the acute/subacute/chronic semantic axis. The degree of encapsulation[16,17] was divided into three categories based on the highest level of data transformation: 0 represented repetition of information (e.g., "white blood cells 14"), 1 represented renaming (e.g., "leukocytosis"), and 2 represented synthesizing features into larger concepts (e.g., combining fever, tachycardia, and infection into "sepsis"). PR word count was obtained in Microsoft Excel (Microsoft Corporation) after removing articles, conjunctions, prepositions, and hyphenation and spelling out abbreviations. We quantified diagnostic accuracy based on including the correct diagnosis in the differential. Answers could be incorrect, partially correct, or completely correct. For example, for Case A, "heart failure" would be partially correct, while "right-sided heart failure," "pulmonary hypertension" or an equivalent term would be completely correct.

Two authors (C.N.M. and M.G.S.) blinded to participant expertise coded 25% of the data ($n = 21$ participants, 42 PRs). Inter-rater reliability was acceptable, with 90% total percent agreement for binary and ordinal variables and 91% interclass correlation for continuous variables. Disagreements were resolved by consensus. One author (C.N.M.) coded the remaining PRs.

Quantitative data analysis

Participant and PR characteristics were summarized descriptively and stratified by expertise level and diagnostic accuracy. Means and standard deviations (SDs) were computed for continuous variables. Frequencies and percentages were computed for categorical variables. We compared case-invariant participant characteristics by expertise level using chi-square tests or Fisher's exact tests, for low expected cell counts. To compare case-level outcomes by group, we employed mixed-effects models to account for clustering of cases within participants. For continuous outcomes, linear mixed-effects regression models were fitted. Mixed-effects logistic regression models were used for binary outcomes, and mixed-effects proportional odds models were employed for ordinal outcomes. All models included a dichotomous fixed effect for group (e.g., expert vs. novice) and a random intercept for participants to account for repeated measures. A type 1 error rate of 0.05 was assumed. No adjustments were made for multiplicity, given the study's exploratory nature. All statistical analyses were completed in SAS version 9.4 (SAS Institute).

For the secondary outcome, diagnostic accuracy, we calculated a target sample size to detect a medium-sized difference with 80% power at 5% significance. Over a range of potential expert-to-novice participant ratios, we determined that we needed $n = 76$ experts and $n = 59$ novices to declare significance with an odds ratio (OR) as small as 2.48. All power calculations were performed using PASS version 14.

RESULTS

Participants and cases

A total of 167 PRs from 84 participants, including 54 experts and 30 novices, were analyzed (Table 1). Each included participant completed both cases except for 1 expert who completed only Case A. One expert was excluded due to incomplete data. Sixty-two participants (38 experts and 24 novices) were from UoP and 22 (16 experts and 6 novices) were from UoM.

Participants were 52% women ($n = 43$), 46% men ($n = 40$), and 2% preferring not to say ($n = 1$), without significant differences between experts and novices ($p = .052$).

There was no difference in diagnostic accuracy between groups ($p = .48$) or between institutions ($p = .259$). Diagnostic accuracy was statistically different between cases ($p = .02$). For Case A, 44% (37/84) submitted a Ddx with a completely correct answer and 46% (39/84) included a partially correct answer. For Case B, 27% (22/83) submitted a Ddx with a completely correct answer and 52% (43/83) included a partially correct answer.

TABLE 1 Characteristics of participants.

	Female	Male	Prefer not to say
University of Pittsburgh			
24 Novices	11 (46%)	13 (54%)	
38 Experts	22 (58%)	15 (39%)	1 (3%)
University of Minnesota			
6 Novices	3 (50%)	3 (50%)	
16 Experts	7 (44%)	9 (56%)	
Expert years in practice			
2 or fewer years	16	30%	
3–5 years	14	26%	
6–10 years	11	20%	
11–20 years	7	13%	
More than 20 years	6	11%	

Differences in PR characteristics by expertise level

Data were analyzed for categorical characteristics (Table 2) and continuous characteristics (Table 3) by expertise level.

Structure, demographics, and clinical background

Most participants (104/167 PRs, 62%) followed the three-part PR structure with no differences between groups ($p = .31$). Every PR included the patient's age, and most (154/167, 92%) included gender with no difference between groups. Experts included on average 0.8 fewer comorbidities/PR (1.8 vs. 2.6, $p < .01$) and were less likely to include irrelevant comorbidities (OR 0.40, $p < .01$). For example, in Case A, an expert likely noted the patient's history of hypertension only, while a novice likely included anxiety and polycystic ovarian syndrome.

Time course and SQs

There was no difference between groups in the frequency of including a temporal description of the patient's presentation ($p = .49$), with it appearing in 112/167 (67%) PRs. The total number of SQs was similar between groups, with 5.8 SQs/PR (SD = 2.3, $p = .51$).

Experts demonstrated a higher degree of encapsulation than novices. On the three-part encapsulation scale, experts scored 0.4 points higher on average than novices (1.3 vs. 0.9, $p < .01$). Experts synthesized findings in 51/107 (48%) PRs and renamed findings in 33/107 (33%) PRs. Novices synthesized findings in 17/60 (28%) PRs and renamed findings in 20/60 (33%) PRs.

TABLE 2 Differences in problem representation (PR) between experts and novices, categorical characteristics.

PR characteristic	Odds ratio (95% confidence interval)	*p* Value
Follows the 3-Part PR format	0.81 (0.28–2.32)	.69
Relevant comorbidities	0.17 (0.02–1.43)	.10
Irrelevant comorbidities	0.40 (0.21–0.78)	.008
Temporal semantic qualifiers	0.75 (0.37–1.50)	.49
Pertinent negatives	1.60 (0.55–4.72)	.39
Proposing a diagnosis	0.32 (0.12–0.82)	.02
Inaccurate information	0.54 (0.24–1.22)	.13
Overall diagnostic accuracy	0.79 (0.42–1.51)	.48

Note: Statistical significance indicated by $p < .05$. All of the PR characteristics in this table were categorical variables, present or not present. Odds ratios compare experts to novices. An odds ratio of 1.0 means experts were just as likely to include a characteristic as novices. Diagnostic accuracy was scored into three discrete categories: completely incorrect, partially correct, and incorrect. An odds ratio <1 corresponds to lower expert accuracy, while an odds ratio >1 corresponds to a higher expert accuracy.

Description of the clinical syndrome

All PRs included pertinent positives, while 41/167 (25%) included pertinent negative findings, with no differences by group. Experts and novices included similar numbers of findings from the history of present illness ($p = .16$) and labs/other studies ($p = .60$) vignette sections. Experts included 0.70 fewer past historical items ($p < .01$) and 0.6 more physical examination findings ($p = .01$). The average total number of findings included was similar between groups, with 7.2 findings/PR (SD 3.0, $p = .74$). For example, in Case B, an expert likely noted the benign abdominal examination while the novice mentioned no examination findings and instead referenced the patient's nursing facility stay (a non-comorbidity, past historical finding). Finally, experts were less likely to propose a diagnosis in their PR (10% vs. 27%, OR = 0.37, $p = .02$).

Inaccuracies

The rate of including inaccuracies was similar between experts and novices (14% vs. 25%, $p = .13$) (Table 4). Inaccuracies were congruent with a participant's Ddx in 23/27 (85%) instances. For example, when changes in taste/smell were incorrectly reported, COVID-19 was on the Ddx even though the patient tested negative.

Post-vignette questions

There were no differences between experts' and novices' responses to the post-vignette questions (Supporting Information: Digital Appendix 3). Results were similar between cases. Across both cases, 150/167 (89%) participants agreed/strongly agreed that they had a diagnosis in mind before writing their PR, 130/167 (78%) agreed/strongly agreed that writing a PR clarified their Ddx, and 121/167 (72%) agreed/strongly agreed that they were confident in their Ddx.

Differences in PR characteristics by diagnostic accuracy

Writing a PR following the three-part structure was associated with diagnostic accuracy: OR = 3.3 ($p = .02$) for partially correct answers, OR = 5.0 ($p < .01$) for fully correct answers. Including temporal SQs was independently associated with accuracy, regardless of whether the three-part PR structure was followed: OR = 2.5 ($p = .049$) for partially correct answers, OR = 4.9 ($p < .01$) for fully correct answers. Diagnostic confidence correlated with diagnostic accuracy. Average Likert scores were 3.5 (SD = 0.9) for incorrect answers, 3.9 (SD = 0.9, $p = .01$) for partially correct answers, and 4.0 (SD = 0.8, $p < .01$) for fully correct answers.

TABLE 3 Differences in problem representation (PR) between experts and novices, continuous characteristics.

PR characteristic	Average number per PR Experts	Novices	Mean difference (95% confidence interval)	*p* Value
Number of comorbidities	1.8	2.6	−0.8 (−1.3 to −0.3)	.001
Total semantic qualifiers	5.7	6.0	−0.3 (−1.3 to 0.6)	.51
Number of included items per case section				
History of present illness	2.2	2.5	−0.3 (−0.6 to 0.1)	.16
Past histories	1.9	2.6	−0.7 (−1.2 to −0.2)	.003
Examination	1.4	0.9	0.6 (0.1–1.0)	.01
Labs/other studies	1.6	1.5	0.1 (−0.5 to 0.8)	.60
Total number of included items	7.2	7.4	−0.2 (−1.5 to 1.0)	.74
Degree of encapsulation[a]	1.3	0.9	0.4 (0.1–0.6)	.009
Number of sentences	1.3	1.0	0.3 (0.04–0.5)	.02
Word count	33	36	−3 (−8 to 2)	.23
Post-vignette questions[b]	Average Likert score			
I had a diagnosis in mind before writing my differential diagnosis.	4.2	4.4	−0.2 (−0.6 to 0.1)	.22
Writing a PR clarified my differential diagnosis.	4.0	4.0	0 (−0.4 to 0.4)	.88
I am confident the correct diagnosis appears in my differential diagnosis.	3.8	3.9	−0.1 (−0.5 to 0.2)	.45

Note: Statistical significance indicated by $p < .05$. All of the PR characteristics in this table were continuous variables. Negative mean differences indicate experts including fewer instances of an item compared with novices, while positive mean differences indicate experts including more instances of an item.

[a]The degree of encapsulation was scored on a three-part scale: 0 for repetition of data, 1 for renaming of data, and 2 for synthesis of data into larger concepts.

[b]For the post-vignette questions, mean differences represent higher scores (positive values) or lower scores (negative values) on a 5-point Likert scale ranging from strongly disagree (1) to strongly agree (5).

TABLE 4 Type and number of included inaccuracies by case.

Case A—Pulmonary hypertension (n)	Case B—Sepsis from acalculous cholecystitis (*n*)
Including a history of weight gain (7)	Mischaracterizing the patient as a nursing home resident (5)
Wrong character of chest pain (7)	Including a history of dyspnea (4)
Wrong time course of symptoms (4)	Including a history of rigors (1)
Including a history of paroxysmal nocturnal dyspnea (1)	Including a history of smell/taste loss (1)
	Mischaracterizing surgical history (1)

Note: A total of 27 problem representations included inaccuracies. Because some problem representations included multiple inaccuracies, the numbers in this table sum to greater than 27.

DISCUSSION

In this mixed-methods clinical vignette study, we demonstrate that expert and novice physicians write PRs with consistent, substantive differences. Experts include less irrelevant data, focus more on physical examination, synthesize data into higher-order concepts like sepsis or respiratory failure, and propose a diagnosis less frequently than novices. Regardless of expertise, writing a PR following a three-part structure comprised of demographics, temporal course, and clinical syndrome was associated with diagnostic accuracy.

The association of diagnostic performance with PR organization underscores the importance of teaching the theory-derived three-part structure for PRs.[1,4] This result follows prior work emphasizing that knowledge organization, not just raw accumulated knowledge, enables sound decision-making.[11,18] Additionally, specifically including a temporal description of the illness correlated strongly with accuracy. Using temporal descriptions—whether acute/chronic, intermittent/constant, and so on—allows clinicians to efficiently hone their Ddx.[6,15] Because clinical reasoning education ultimately should strive to improve diagnostic accuracy, we suggest that future PR training efforts should emphasize using structured PR to organize and narrow a differential rather than just collecting important facts about a case.

Compared with past works, SQ use was common regardless of expertise level. This difference is likely due to the level of learner studied: the original SQ studies[5,6,11] evaluated differences between attendings and medical students, who are less experienced than the residents in our study. Subsequent studies[19,20] analyzing diagnostic discourses suggest that using SQs to contrast candidate diagnoses correlates better with accuracy than total SQ numbers. Surprisingly, including pertinent negative findings was uncommon among experts and novices. We suspect this is an artifact of the PR-writing process (e.g., striving for word efficiency or assuming unmentioned findings are normal) rather than judgment on the importance of negative findings to diagnosis.[1,3]

Experts included more physical exam findings and synthesized multiple data points into higher-order terms. The difference in reporting exam findings is reminiscent of concerns that novices rely less on physical examination than on laboratory/imaging studies.[21,22] The differences in data synthesis reflect prior work on knowledge encapsulation, where experts more frequently synthesize individual biomedical data like leukocytosis, fever, and suspected infection into a larger idea like sepsis.[17,18] The average total number of findings included in each PR was consistent with working memory theory, which states that working memory can hold six to eight total items.[23]

We also note how few experts proposed a diagnosis at the end of their PR, especially compared with novices. Finishing PRs with a favored solution may prematurely close intellectual exploration, negating the educational purpose of PR and potentially injecting bias into the search for a diagnosis.[1,4] We hope this finding emboldens educators to continue discouraging this practice while promoting robust differential diagnosis as a distinct step in solving clinical problems.[1]

The prevalence of inaccuracies in PRs was surprising: participants were able to review the cases to verify data while writing their PRs. We speculate that, even though our study had no time limit, subjective time pressure may have caused inaccuracies.[24,25] Alternately, because of how often inaccuracies were congruent with the Ddx and how frequently participants reported having a diagnosis in mind before writing their PRs, we propose these errors could represent confirmation bias, where memory gaps are filled with ideas confirming a hypothesis.[26] Educators must emphasize that PR is an opportunity to refine hypotheses by synthesizing supporting and conflicting data—similar to deliberate reflection[27]—rather than choosing data supporting a prevailing theory. Future investigations should examine how errors summarizing cases occur and why they cluster around particular details.

Knowing these results, educators could create curricula teaching expert-like PR to novices. We advise caution. Such curricula may cause novices to skip important developmental steps related to clinical problem-solving.[11,18] Novices may need to write longer, more diffuse summaries while they are refining their illness scripts and have high diagnostic uncertainty.[18,28] Additionally, we note that most of our attending participants had 5 years of experience or less. While the literature would suggest that PR skills continue to develop as experience accrues, all prior studies have examined trainees versus attendings, not attendings of different experience levels.[3,5,11,17] Truly expert-level PR may be both content- and context-dependent, regardless of total years of experience. For example, a resident freshly off an immersive ambulatory month may provide a better PR for subacute foot pain than an intensivist with 20 years of focusing only on critical care. Future studies should explore how attending PR improves with continued experience and what developmental steps characterize how trainees refine PR skills. In the meantime, we suggest that our findings could motivate conversations with trainees about encapsulating data points into larger ideas, discerning relevant risk factors from irrelevant comorbidities, and emphasizing the importance of exam findings to diagnosis.

LIMITATIONS

Although we analyzed many PRs from two geographically distinct academic centers, our participants may not reflect novices or experts in community settings with various levels of prior clinical reasoning instruction.[29] Additionally, the nomenclature of PR versus summary statement versus one-liners is heterogeneous, which could lead to differences in PR creation. Importantly, there was no difference in diagnostic accuracy between experts and novices. This may have been due to the cases' difficulty or selection bias related to our study being underpowered for diagnostic accuracy. Nonetheless, we elucidated important differences in how novices and experts create PRs and describe which facets of PRs correlate with accuracy. Larger studies with vignettes covering additional clinical areas should investigate if other aspects of PR contribute to diagnostic accuracy.

CONCLUSIONS

This study establishes concrete differences in how experts and novices construct PRs and emphasizes aspects of PR that aided clinicians in solving clinical vignettes. These findings can inform future research on using PR as a teaching tool to improve diagnosis. Future studies should determine the causes of novice-expert differences.

ACKNOWLEDGMENTS

The authors would like to thank Melissa McNeil, MD, MPH and William Follansbee, MD, without whose mentorship and encouragement this project would not have been possible. This work was supported by the Society to Improve Diagnosis in Medicine's Fellowship in Diagnostic Excellence and a grant from the Gordon and Betty Moore Foundation.

CONFLICT OF INTEREST STATEMENT

C.N.M. has been a consultant for Glass Health, Inc., since February 2024. The other authors declare no conflict of interest.

ETHICS STATEMENT

This study was deemed exempt after IRB review at both UoP and UoM.

ORCID

Casey N. McQuade http://orcid.org/0000-0002-6139-5531

TWITTER

Casey N. McQuade @CaseyMcQuadeMD
Michael G. Simonson @mgsimonson1
Andrew P. J. Olson @andrewolsonmd
Laura Zwaan @LauraZwaan81

REFERENCES

1. Bowen JL. Educational strategies to promote clinical diagnostic reasoning. *N Engl J Med*. 2006;355:2217-2225.
2. Olson A, Rencic J, Cosby K, et al. Competencies for improving diagnosis: an inter professional framework for education and training in health care. *Diagnosis*. 2019;6(4):335-341.
3. Chang RW, Bordage G, Connell KJ. The importance of early problem representation during case presentations. *Acad Med*. 1998;73(10): S109-S111.
4. Rencic JR, Durning SJ. *Teaching Clinical Reasoning*. American College of Physicians; 2015.
5. Bordage G, Lemieux M. Semantic structures and diagnostic thinking of experts and novices. *Acad Med*. 1991;66(9):S70-S72.
6. Bordage G, Connell KJ, Chang RW, Gecht MR, Sinacore JM. Assessing the semantic content of clinical case presentations: studies of reliability and concurrent validity. *Acad Med*. 1997;72(10):S37-S39.
7. Duca NS, Glod S. Bridging the gap between the classroom and the clerkship: a clinical reasoning curriculum for third-year medical students. *MedEdPortal*. 2019;15:10800. doi:10.15766/mep_2374-8265.10800
8. Heist BS, Kishida N, Deshpande G, Hamaguchi S, Kobayashi H. Virtual patients to explore and develop clinical case summary statement skills amongst Japanese resident physicians: a mixed methods study. *BMC Med Educ*. 2016;16:39.
9. Fleiszer D, Hoover ML, Posel N, Razek T, Bergman S. Development and validation of a tool to evaluate the evolution of clinical reasoning in trauma using virtual patients. *J Surg Educ*. 2018;75(3):779-786.
10. Smith S, Kogan JR, Berman NB, Dell MS, Brock DM, Robins LS. The development and preliminary validation of a rubric to assess medical students' written summary statements in virtual patient cases. *Acad Med*. 2016;91:94-100.
11. Nendaz MR and Bordage Ge. Promoting diagnostic problem representation. *Med Educ*. 2002;36:760-766.
12. Graber ML. Progress understanding diagnosis and diagnostic errors: thoughts at year 10. *Diagnosis*. 2020;7(3):151-159.
13. Klein G. *Sources of Power: How People Make Decisions*. MIT Press; 1998.
14. Baker EA, Connell KJ, Bordage G, Sinacore J. Can diagnostic semantic competence be assessed from the medical record? *Acad Med*. 1999;74(10):S13-S15.
15. Connell JK, Bordage G, Chang RW. *Assessing Clinicians' Quality of Thinking and Semantic Competence: A Training Manual*. University of Illinois at Chicago, Northwestern University Medical School; 1998.
16. ten Cate O, Durning SJ. Understanding clinical reasoning from multiple perspectives: a conceptual and theoretical overview. In: ten Cate O, Custers EJFM, Durning SJ, eds. *Principles and Practice of Case-based Clinical Reasoning Education: A Method for Preclinical Students*. Springer; 2017:35-46.
17. Schmidt HG, Boshuizen HPA. On the origin of intermediate effects in clinical case recall. *Mem Cognit*. 1993;21(3):338-351.
18. Schmidt HG, Rikers RM. How expertise develops in medicine: knowledge encapsulation and illness script formation. *Med Educ*. 2007;41: 11133-11139.
19. Dory V, Gagnon R, Charlin B, et al. In brief: validity of case summaries in written examinations of clinical reasoning. *Teach Learn Med*. 2016;28(4):375-384.
20. Valentine J, Tackett S, Bord S, Zink K, Botta J, Jung J. Problem representation and diagnostic accuracy during an OSCE. *Adv Health Sci Educ Theory Pract*. 2022;27:1021-1031.
21. Hyman P. The disappearance of the primary care physical examination—losing touch. *JAMA Intern Med*. 2020;180(11):1417-1418.
22. Russo S, Berg K, Davis J, et al. Incoming interns recognize inadequate physical examination as a cause of patient harm. *J Med Educ Curric Dev*. 2020;7:238212052092899.
23. Young JQ, Van Merrienboer J, Durning S, Ten Cate O. Cognitive load theory: implications for medical education: AMEE Guide No. 86. *Med Teach*. 2014;36:371-384.
24. Sherbino J, Dore KL, Wood TJ, et al. The relationship between response time and diagnostic accuracy. *Acad Med*. 2012;87: 785-791.
25. ALQahtani DA, Rotgans JI, Mamede S, Mahzari MM, Al-Ghamdi GA, Schmidt HG. Factors underlying suboptimal diagnostic performance in physicians under time pressure. *Med Educ*. 2018;52:1288-1298.
26. Norman GR, Monteiro SD, Sherbino J, Ilgen JS, Schmidt HG, Mamede S. The causes of errors in clinical reasoning: cognitive biases, knowledge deficits, and dual process thinking. *Acad Med*. 2017;92:23-30.
27. Mamede S, van Gog T, van den Berge K, et al. Effect of availability bias and reflective reasoning on diagnostic accuracy among internal medicine residents. *JAMA*. 2010;304(11):1198-1203.
28. Audétat MC, Laurin S, Sanche G, et al. Clinical reasoning difficulties: a taxonomy for clinical teachers. *Med Teach*. 2013;35(3):e984-e989.
29. Rencic J, Trowbridge RL, Fagan M, Szauter K, Durning S. Clinical reasoning education at US medical schools: results from a national survey of internal medicine clerkship directors. *J Gen Intern Med*. 2017;32(11):1242-1246.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: McQuade CN, Simonson MG, Lister J, et al. Characteristics differentiating problem representation synthesis between novices and experts. *J Hosp Med*. 2024;19: 468-474. doi:10.1002/jhm.13335