Sun *et al. BMC Medical Education* (2025) 25:546
https://doi.org/10.1186/s12909-025-07134-6

BMC Medical Education

RESEARCH

Open Access

Quantify difference between physicians and medical students in clinical reasoning: evidence from eye-tracking

Lijun Sun[1,2], Yao Zhang[2] and Bin Zheng[2*]

Abstract

Background The assessment of clinical reasoning in health trainees is vital yet poses challenges. We tracked the eye movements of participants while they were reviewing a neurological case with the goal of finding behavioral evidence to improve health education.

Methods Eleven medical students and seventeen expert physicians were required to read a neurological case within a 150-second timeframe. The case included descriptive text, a brain CT scan, and an electrocardiogram (ECG). Participants completed a multiple-choice questions (MCQs) test after reading the case. Eye movements of participants in case reading on eleven patient-related information areas (PRIAs) were compared between experts and novices, contrasted with the remaining areas.

Results Experts spent significantly more time fixating on PRIAs during case reading than novices (42.1% vs. 29.2%, adjusted $p = 0.010$). Experts demonstrated significantly fewer gaze shifts between Text and CT images (2.0 times) and between CT and ECG images (2.4 times) compared to novices (6.2 and 5.4 times), with adjusted p-values of 0.002 and 0.019, respectively. A positive correlation was found between the fixation rate on PRIAs and MCQs outcome ($r = 0.402$, $p = 0.034$).

Conclusion Eye-tracking provides rich and reliable data reflecting physicians' ability to gather patient-relevant information during patient assessment.

Keywords Clinical reasoning, Eye-tracking, Skill assessment

Introduction

Making clinical diagnoses is not an easy task for healthcare trainees. Mistakes in this process will threaten patient safety and quality of care. A conservative estimate found that 5% of U.S. adults who seek out patient care each year experience a diagnostic error [1]. Annually, an estimated 40,000 to 80,000 deaths in the United States are attributed to the errors in patient diagnoses [2–5], prompting the National Academy of Sciences to declare this issue a national emergency [1]. Mistakes in diagnosis may rise from incomplete information taking, incorrect judgement, and deficiency in knowledge [2, 6, 7]. This

*Correspondence:
Bin Zheng
bin.zheng@ualberta.ca
[1]Department of Neurology, The Affiliated Hospital of Inner Mongolia Medical University, Hohhot, Inner Mongolia, China
[2]Surgical Simulation Research Lab, Department of Surgery, University of Alberta, Edmonton, Canada

cognitive exercise is often referred to as clinical reasoning [8–10]. Faulty clinical reasoning is considered a key contributor to diagnostic errors [11, 12]. Improving clinical reasoning competence will significantly reduce diagnostic errors [1, 2]. In addition, one study discovered that 57.7% of medical dispute cases (legal or formal grievances brought by patients (or their families) against healthcare providers or institutions), were attributed to clinical reasoning errors [11]. Any errors or delays in clinical reasoning will potentially lead to tragical outcomes for patients, including physical, mental, and life-threatening risks, alongside elevating healthcare costs. Consequently, training to enhance clinical reasoning - a foundation of clinical diagnoses - has become an increasingly central focus in medical education [9].

In order to improve clinical reasoning, medical education has shifted from lecture-based curricula to exploration learning formats, such as Problem-Based Learning (PBL), Case -Based Learning (CBL) and Team-Based learning (TBL) [13–15]. These approaches have helped learners interpret patient information and make clinical management judgments more effectively. However, quantifying clinical reasoning is still challenging due to the inherent vagueness of clinical information, the complex, multi-tiered nature of patient conditions, and the individual variances of practitioners' background knowledge [10, 16, 17]. Furthermore, because clinical reasoning is largely an individual cognitive process, it is often invisible and difficult to observe in practice [18]. Currently, evaluations of clinical reasoning largely rely on end-point assessments. There are other efforts to evaluate of clinical reasoning by analyzing trainees' answers to multiple-choice questions (MCQs), oral case presentation, and written clinical notes [1, 16, 19]. Trainees' performance in during OSCE (Objective structured clinical examinations) can also be analyzed for assessing their clinical reasoning [19]. These methods largely measure the final outcome rather than the reasoning steps that led to it, leaving an important gap in understanding how information is collected and processed for diagnostic decision-making [20].

Recent advances in sensor technology now offer fresh opportunities to observe cognitive activity in real time, potentially providing deeper insights into clinical reasoning [21–24]. Eye tracking, a method of recording eye movements, serves as a valuable tool for examining human attention, cognitive processing, and information processing strategies [23, 24]. Utilized across various domains, including psychology, education, and medical research, eye tracking offers an unparalleled perspective on medical professionals' decision-making processes [23–26]. By capturing how clinicians engage with patient information and diagnostic materials, it can shed light on the cognitive basis of clinical reasoning, despite the inherent challenge of making this mental process visible [27]. Indeed, previous eye-tracking studies have shown how experts focus on key areas of radiographs or textual information to detect abnormalities more efficiently [28].

Yet, most eye-tracking research in clinical reasoning has explored how individuals read either images or text alone. We are interested in investigating how physicians read both text and image together during their assessment of clinical cases. To address this gap, we employ an expert-novice paradigm, a common approach in medical education research, to investigate how physicians process both text and imagery (CT and ECG) during the assessment of neurological cases. In line with the cognitive theory of visual expertise [29] and the information-reduction hypothesis [30], experts are expected to deploy rich mental schemas that rapidly filter out irrelevant content and highlight diagnostic cues. Specifically, we ask participants to read a neurological case featuring a short text description, a CT scan, and an ECG image within 150 s, then answer multiple-choice questions. We hypothesize that experts (experienced neurologists), with their more refined clinical reasoning, will show (a) more intensive engagement (e.g., longer fixations) on key patient information, (b) stronger connectivity among key information areas (e.g., increased saccade frequency across the text and images), and (c) tighter correlations between eye-movement metrics and diagnostic outcomes than novices (MD students). In this way, our study seeks to deepen the understanding of clinical reasoning processes and expand upon prior eye-tracking research in the field.

Methods

This research was conducted at the University of Alberta's Surgical Simulation Research Lab in collaboration with the Department of Neurology of the Inner Mongolia Medical University. Prior to initiating participant recruitment and data collection, we obtained ethical approval from the Health Research Ethics Review Board of the University of Alberta (Pro00102074).

Participants

To evaluate clinical reasoning skills among participants, we conducted an experimental study involving 28 individuals, divided into two groups: 17 experts, each possessing a minimum of six years of clinical experience in neurology, and 11 novices, consisting of third-year medical students enrolled in an MD program.

Task & procedure

Participants were seated 75 cm away from a 24-inch high-quality monitor (ASUS LCD 60 Hz, ASUS Computer Inc, Taipei, Taiwan), with a screen-based eye tracker (Tobii Pro Nano, Tobii Technology, Stockholm, Sweden) positioned at the bottom (Fig. 1A). The eye

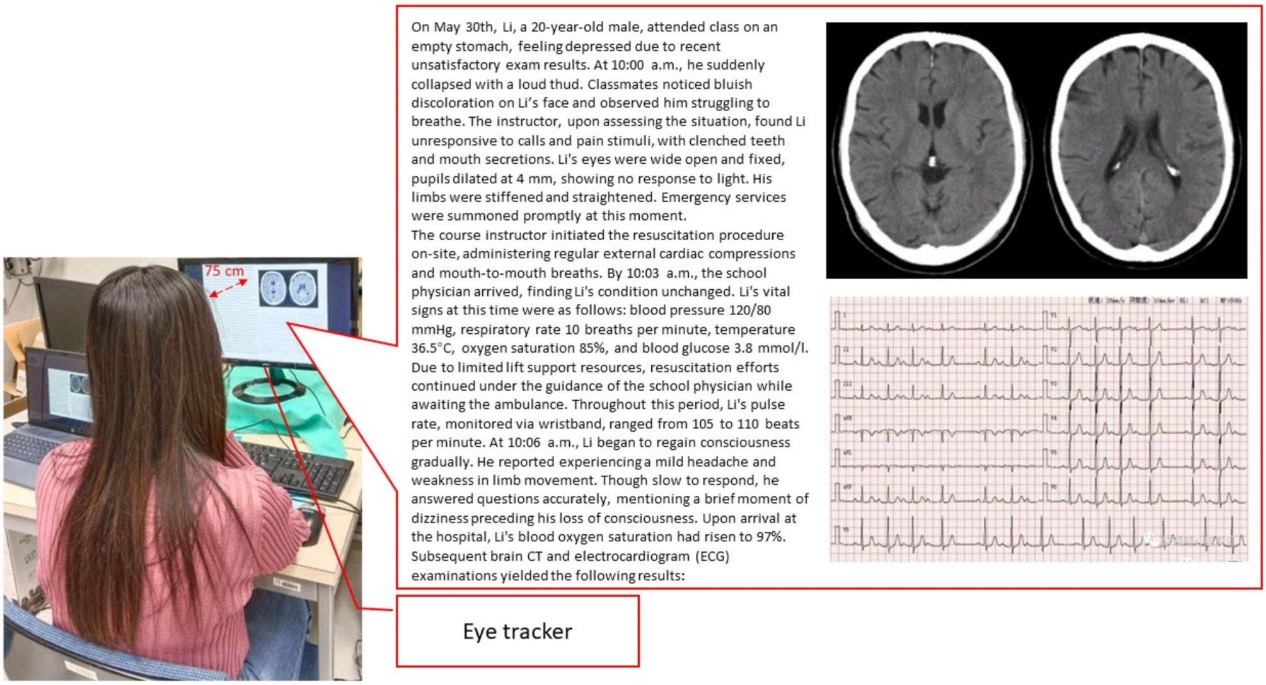

Fig. 1 Experimental setup

tracker has a sampling rate of 60 Hz and an accuracy of 0.3 degree of visual angle and precision of 0.15 degree. Displayed on the monitor was a page containing a medical record paragraph describing the case of a 20-year-old male patient who experienced a sudden fall, brief loss of consciousness, limb stiffness, and pupil dilation, suggestive of an epileptic episode. The patient description page also featured associated CT scans and ECG results obtained from the patient (Fig. 1B).

Participants were instructed to read the patient information carefully (text, CT image, and ECG), consider possible diagnoses, and be prepared to answer 10 multiple-choice questions (MCQs) about the clinical scenario. We assessed the internal consistency of our 10 MCQ test using Cronbach's alpha. We calculated a Cronbach's alpha of 0.81, demonstrates good internal consistency.

Each participant was given 150 s to review all the provided information. After this reading time, the system automatically transitioned to a page presenting 10 multiple-choice questions (MCQs) directly correlated with the reviewed materials. Participants were instructed to read and verbally respond to each question within a 300-second timeframe. An identical set of patient information pages and MCQs was presented to all participants, with an experimenter recording their verbal responses. The experimental setup is depicted in Fig. 1.

Data

Before commencing the experiment, participants completed a preliminary survey to gather demographic information, including age, sex, and either the year of clinical service for experts or the year of medical study for students.

Participants' clinical reasoning outcome on the patient information was evaluated at the end of the case reading based on their performance in a multiple-choice question (MCQ) test. Each correct answer in the MCQ test contributed one point to the cumulative score, with a maximum achievable score of 10.

Throughout the experiment, a screen-based eye tracker recorded participants' eye movement trajectories in conjunction with the on-screen content. These recordings were subsequently analyzed, with three main *areas of interest* (AOIs) identified on the displayed patient information page: AOI 1 (Descriptive Text) represented the descriptive text, AOI 2 (CT Scan) encompassed the CT image, and AOI 3 (ECG) focused on the ECG image. Furthermore, the entire medical information page was segmented by an expert neurologist (consultant and co-investigator on the project, did not participate as a subject in the expert group), identifying nine *patient-related information areas* (PRIAs) within the text and two within the images that are critical for clinical reasoning based on patient's condition (Supplementary Fig. 1).

A list of eye metrics is presented in Table 1. Frequency and duration of fixation and saccade are commonly utilized to describe general eye scanning patterns of human operator in reading tasks. The fixation rate will unveil features of visual attention across various AOIs and PRIAs. Additionally, we assessed the frequency of 'gaze

Table 1 Outcome of clinical reasoning and eye metrics

Outcome assessment	
MCQ Test Score	Cumulative score of MCQ test
Eye Metrics	
Fixation frequency (counts/s)	The total number of fixations for each trial divided by reading time (150 s)
Fixation duration (%)	The summation of the total duration of fixations for each trial divided by reading time (150 s)
Saccade frequency (counts/s)	The total number of saccades for each trial divided by reading time (150 s)
Fixation rate on Descriptive Text (%)	The summation of the total duration of fixations landing inside the AOI 1 divided by reading time (150 s)
Fixation rate on AOI 2 (%)	The summation of the total duration of fixations landing inside the AOI2 divided by reading time (150 s)
Fixation rate on AOI 3 (%)	The summation of the total duration of fixations landing inside the AOI 3 divided by reading time (150 s)
Fixation rate on PRIA (%)	The summation of the total duration of fixations landing inside the PRIA divided by reading time (150 s)
Number of gaze shift AOI 1 - AOI 2 (count)	The number of gaze shift between AOI 1 and 2
Number of gaze shift AOI 1 - AOI 3 (count)	The number of gaze shift between AOI 1 and 3
Number of gaze shift AOI 2 - AOI 3 (count)	The number of gaze shift between AOI 2 and 3

Note: MCQ: multiple-choice question; AOI: area of interest; AOI 1: descriptive text; AOI 2: brain CT scan image; AOI 3: electrocardiogram

Table 2 Participant demographics

	Sex (Male: Female)	Age (Mean ± SD)	Years of Experience (Mean ± SD)
Expert	2:15	35.4 ± 10.6	11.0 ± 8.5
Novice	5:6	21.0 ± 1.3	3.0 ± 0.0

shifts' - instances where the participant's gaze transitioned between different AOIs, aiding us in investigating how participants connect information presented at different AOIs.

Statistical analysis

To identify significant differences in eye behavior between students and experts, we employed t-tests. These tests allowed us to compare the two groups with respect to various eye-tracking metrics listed in Table 1, thereby highlighting any statistically significant disparities in their visual attention patterns.

Furthermore, we conducted linear regression analyses on the eye metrics associated with PRIAs and the participants' MCQ test scores. This approach was intended to explore potential relationships between participants' eye behaviors, specifically, how they engaged with the identified PRIAs, and their performance on the MCQ tests. By correlating these eye metrics with test scores, we aimed to discover whether certain patterns of visual attention could predict or relate to higher or lower levels of performance in clinical reasoning tasks.

Results

The study comprised 28 participants. Participant demographics, including sex, age, and years of experience, are detailed in Table 2. There were 17 experts (Mean age: 35.41, 2 males and 15 females) and 11 novices (Mean age: 21, 5 males and 6 females).

On average, experts have 11 years of clinical experience in neurology. Conversely, all the novices were third year medical students, starting to learn clinical course with minimal experience in neurology.

Outcome of clinical judgement

The outcome of clinical judgement was quantitatively assessed by the MCQ test after the case reading for each participant. The expert group achieved a higher average score of 9.1 ± 1.4 out of 10, compared to the novice group, which received an average score of 4.7 ± 1.3 out of 10. The t-test applied to compare these scores revealed a statistically significant difference between the expert and novice groups, with a p value < 0.001.

Eye movement behaviors

We began our analysis by examining general eye scanning behaviors of participants, including fixation frequency (counts/s), Fixation duration (%) and Saccade frequency (counts/s). As shown in Table 3, no significant differences were observed between the two groups in these three scene-independent variables. However, a closer look at the data for each scene-dependent variables revealed notable differences.

Fixation, a key metric in eye-tracking studies, represents a prolonged focus on a specific area (scene) displayed in the screen, reflecting participants' control of their visual attention while reading message from medical sheet. Analysis of fixation rates on PRIAs showed that experts spent a significant portion of their reading time (42.1% ± 7.9%) focusing on patient-related information, compared to novices (29.2% ± 10.6%). A t-test confirmed this difference as statistically significant (adjusted $p = 0.010$). Furthermore, fixation rates on AOI 2, a representation of a normal brain CT scan, varied significantly between groups. Experts allocated a significantly smaller percentage of their fixations (7.5% ± 6.3%) to AOI 2 (CT SCAN) than novices (20.8% ± 14.2%), with an adjusted p value of 0.023.

Table 3 Eye metrics results

Eye Metrics	Mean ± SD (Expert)	Mean ± SD (Novice)	t value	p value	p-value (Bonferroni-adjusted)	95% CI (Lower, Upper)	Cohen's d
Fixation frequency (counts/s)	2.8 ± 0.3	2.9 ± 0.4	-0.86	0.401	1	(-0.37, 0.15)	-0.3
Fixation duration (%)	81.4% ± 6.9%	80.3% ± 7.3%	0.39	0.705	1	(-0.05, 0.07)	0.2
Saccade frequency (counts/s)	2.3 ± 0.4	2.3 ± 0.4	-0.06	0.956	1	(-0.33, 0.31)	0.0
Fixation rate on AOI 1 (%)	54.1% ± 18.8%	43.9% ± 15.8%	1.49	0.149	1	(-0.04, 0.24)	0.6
Fixation rate on AOI 2 (%)	7.5% ± 6.3%	20.8% ± 14.2%	-3.39	0.002*	0.023*	(-0.21, -0.05)	-1.3
Fixation rate on AOI 3 (%)	19.5% ± 14.9%	15.0% ± 9.4%	0.89	0.384	1	(-0.06, 0.15)	0.3
Fixation rate on PRIA (%)	42.1% ± 7.9%	29.2% ± 10.6%	3.70	0.001*	0.010*	(0.06, 0.20)	1.4
Number of gaze shift AOI 1 - AOI 2 (count)	2 ± 1.4	6.2 ± 3.6	-4.36	0.000*	0.002*	(-6.15, -2.21)	-1.7
Number of gaze shift AOI 1 - AOI 3 (count)	3.1 ± 2.0	4 ± 2.6	-1.09	0.285	1	(-2.71, 0.83)	-0.4
Number of gaze shift AOI 2 - AOI 3 (count)	2.4 ± 1.4	5.4 ± 3.1	-3.46	0.002*	0.019*	(-4.71, -1.20)	-1.3

Note: PRIA: patient-related information area; AOI: area of interest; AOI 1: descriptive text; AOI 2: brain CT scan image; AOI 3: electrocardiogram

In addition, we assessed the number of gaze shifts between AOIs to understand navigational patterns. Experts demonstrated significantly fewer gaze shifts between AOI 1 (Descriptive Text) and 2 (CT Scan) (2 ± 1.4 times) and between AOI 2 (CT Scan) and 3 (ECG) (2.4 ± 1.4 times) compared to novices (6.2 ± 3.6 times between AOI 1 (Descriptive Text) and 2 (CT Scan); 5.4 ± 3.1 times between AOI 2 (CT Scan) and 3 (ECG)), with adjusted p values of 0.002 and 0.019, respectively.

Presented in Table 3, these results suggest that experts navigate content more efficiently than novices, focusing less on irrelevant information. This implies a more structured approach to information gathering and processing.

Correlation between performance and eye behaviors

We explored the correlation between eye movement behaviors and performance on MCQ test through testing Pearson's correlation coefficients (r), as depicted in Fig. 2. Our findings indicated a positive correlation between the fixation rate on PRIAs and MCQ scores, evidenced by a correlation coefficient (r) of 0.402 and a p value of 0.034. This suggests that a longer duration of fixation on PRIAs is associated with higher MCQ scores.

In contrast, the fixation rate on AOI 2, which represents a normal CT scan without critical diagnostic information, was negatively correlated with MCQ scores (r = -0.499, p = 0.007). This negative correlation implies that experts, who typically offer fewer glances to such non-relevant information, demonstrate a better grasp of the case and scenario, as less attention to AOI 2 (CT SCAN) correlates with higher MCQ scores.

Additionally, our analysis showed a negative correlation between the frequency of gaze shifts between AOI 1–2 and AOI 2–3 and MCQ scores (r = -0.500, p = 0.007 for AOI 1–2; r = -0.436, p = 0.020 for AOI 2–3). This finding suggests that participants who navigate between areas more efficiently, by reducing unnecessary gaze shifts, tend to achieve higher scores on the MCQ test.

Examining subgroup correlations for experts and novices separately can clarify whether the observed associations rise from between-group or within-group differences. Therefore, we conducted a post-hoc analysis on a per-group basis.

For the expert group, the correlation between PRIAs and MCQ was weaker (r = -0.093, p = 0.721), likely due to the reduced variance in both MCQ scores (most experts scored near the top) and gaze patterns (experts were relatively consistent in focusing on PRIAs). The correlation between CT fixation and MCQ remained small (r = -0.232, p = 0.371). Both results are not statistically significant.

For the novice group, we found a moderate, negative correlation between fixation on PRIAs and MCQ scores (r = -0.398, p = 0.225). Fixation on the CT did not correlate strongly with MCQ performance (r = 0.043, p = 0.901). Both results are not statistically significant.

Discussion

We are pleased to observe the evidence collected by the eye tracker, particularly the scene-dependent measures reveal differences between experts and novices when reading medical information sheets. Within a fixed period of time (150 s), experts dedicated more time to key patient-related information areas than novices did. Specifically, in a case where a CT image of the patient appeared normal, without significant changes following the patient's epileptic episode, experts seemed to quickly notice this fact. Consequently, they spent less time fixating on these areas, resulting in fewer revisits to the CT area during the case reading, as shown in Table 3. In contrast, novices focused more on the CT area, spending longer periods and frequently revisiting the CT images during the case reading. These findings suggest that experts do not evenly distribute their visual attention across the entire medical information sheet. Instead, they efficiently locate key information scattered throughout the sheet. Similar observations were observed in

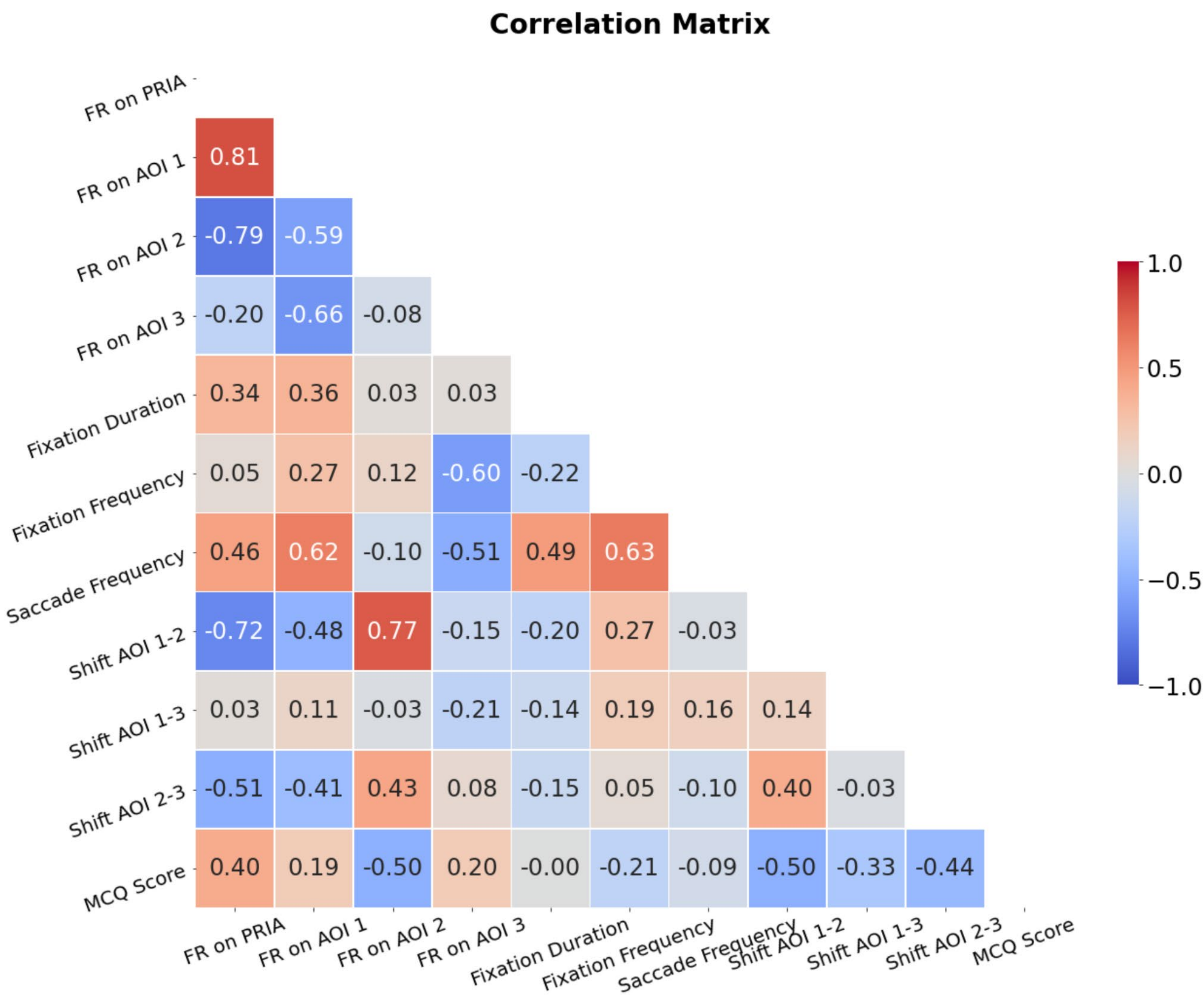

Fig. 2 Correlation matrix with all variables

pathologist [31, 32]. When reading pathological slides, the experts spent more time to the highest diagnostic relevant regions, which enable them to make diagnosis quick and more accurate. Conversely, novices spent more time on visually salient regions without relevance to the diagnosis [31, 32].

When describing clinical experts' reasoning processes, scientists often characterize them as intuitive, quick, and accurate [10, 33, 34]. We believe this type of reasoning is built upon their efficiency in gathering key information from the environment and integrating it with their previous knowledge, facilitating rapid and accurate decision-making, including diagnosis and management plans. In line with Kahneman's "System 1" (fast, intuitive) and "System 2" (slow, analytical) framework [35], experts may rely more on System 1 processes for identifying normal or irrelevant findings whereas novices engage in a more laborious, System 2 approach that involves scrutinizing every element. In terms of top-down and bottom-up processing, experts appear to use top-down strategies, guided by schema-based knowledge, to decide whether a region warrants additional scrutiny, while novices depend more on bottom-up processing, examining each area more evenly due to less developed schema.

Our findings can also be better understood by integrating them into established cognitive theories of expertise. The information-reduction hypothesis [30] posits that experts rapidly categorize information and filter out less relevant cues by drawing on well-developed mental representations. This aligns with our observation that experts quickly dismissed the normal CT scan, thereby allocating more attention to crucial patient-related details. However, some studies [36] have reported overall differences in fixation count and duration between trained and untrained learners, suggesting that factors like task complexity or participant characteristics can influence whether global measures significantly distinguish between groups.

Moreover, our reliance on AOIs aligns with an expanding research base that uses AOI-based approaches to examine clinical reasoning [37–39]. While earlier eye-tracking studies often reported only global fixations or saccade counts, more recent work [40] reveals the added value of dissecting scene-dependent interactions to capture subtle differences in how participants process critical portions of medical cases. Reporting only overall fixation frequency and duration, as well as saccade frequency, failed to reveal significant differences between experts and novices. In contrast, scene-dependent eye measures offered meaningful evidence. Therefore, we strongly recommend incorporating scene-dependent eye measures in future studies, despite the labor-intensive process of identifying AOIs from video-based eye-tracking data.

To further support our focus on scene-dependent measures, we noted a positive and significant correlation between participants' fixation on PRIA and their scores MCQ tests. Although we did not quantify the specific contribution of fixation on PRIA to the MCQ score in this study from regression analysis, we aim to further investigate the association between healthcare providers' eye scanning behaviors during chart reading and their decision-making in clinical practice. The post-hoc analysis suggest that the overall correlation reflects a group-level distinction rather than a simple within-group mechanism in which any novice who looks longer at key details automatically achieves higher scores. The small sample size and heterogeneous strategies among novices may have contributed to the non-significant correlation. Future research should incorporate a larger novice sample or additional intermediate levels to better understand how increased attention to critical information might eventually translate to improved performance as learners advance in clinical expertise.

From a medical education perspective, our results carry several practical implications. First, eye-tracking methods could reveal where novices spend excessive time on irrelevant or normal findings, suggesting targeted teaching interventions. Second, gaze training, where novices learn to emulate experts' scanning strategies, could hasten the acquisition of top-down reasoning skills. Finally, identifying high vs. low-yield elements in a case allows educators to focus on schema-building exercises, aiming to reduce future diagnostic errors by guiding students toward the information experts consider essential.

Our study has several limitations. Although our introduction and overall motivation highlight real-world medical errors, we concede that reading a medical chart and answering questions is an oversimplified representation of clinical reasoning. We are developing more advanced simulation models to capture the complexity of actual practice. The case presentation was adopted from a neurologic patient; therefore, caution is necessary when extrapolating results to other specialties where patient descriptions and associated vital medical images may vary. Additionally, we recognize that idle eye movements occurred in some participants, particularly among experts who finished scanning early. In this study, we did not exclude those idle periods from the total reading time, which represents a key limitation. We will refine our data processing to filter out these non-productive segments by adjusting how we define and measure fixation reading periods in future studies. What is more, we identified PRIAs via a neurologist consultant. We acknowledge that this may introduce a certain bias. Experts often rely on implicit, procedural knowledge, and may omit or undervalue details that are critical to learners. Future work could triangulate these PRIAs with empirical data on diagnostic accuracy or with multiple experts to mitigate the bias. Last but not least, incorporating an intermediate group, such as residents may allow us to observe transitions between expert and novice behaviors, which should be considered in our future studies.

Conclusion

Eye-tracking is a reliable tool for collecting significant evidence to reveal differences between experts and novices when reading medical information sheets. When analyzing eye-tracking data, scene-dependent variables surpass scene-independent variables. We advocate for the incorporation of scene-dependent analysis in future studies utilizing eye-tracking technologies.

Identifying differences between novices and experts is the first step toward our goal of improving clinical reasoning in health education. Leveraging evidence collected from eye-tracking as an educational tool can help novices learn to identify key medical information as effectively as experts during their early learning phases.

Abbreviations

ECG	Electrocardiogram
MCQs	Multiple-Choice Questions
PRIAs	Patient-Related Information Areas
AOIs	Areas of Interests

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12909-025-07134-6.

Supplementary Material 1: Fig. 1. Patient-related information areas identified by experts.

Acknowledgements

Not applicable.

Author contributions

The conception and design of the study was by LJS and BZ. LJS and YZ collected the data, which was analyzed by LJS, YZ and BZ. The article was drafted by LJS and critically revised by YZ and BZ. The final version to be published was approved by LJS, YZ and BZ.

Funding
This work was supported by International Collaboration Funds from the Affiliated Hospital of Inner Mongolia Medical University to Dr. Lijun Sun.

Data availability
The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate
Prior to initiating participant recruitment and data collection, we obtained ethical approval from the Health Research Ethics Review Board of the University of Alberta. Written consent was obtained from each of the participants before conducting the experiment.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 24 May 2024 / Accepted: 7 April 2025
Published online: 16 April 2025

References

1. Medicine Io. Improving diagnosis in health care. Washington (DC): National academies press (US) copyright 2015 by the National academy of sciences. All rights reserved.; 2015.
2. Berwick DM, Leape LL. Reducing errors in medicine. BMJ. 1999;319(7203):136–7.
3. Leape LL. Error in medicine. JAMA. 1994;272(23):1851–7.
4. Leape LL. Reporting of medical errors: time for a reality check. Qual Health Care. 2000;9(3):144–5.
5. Leape LL. Errors in medicine. Clin Chim Acta. 2009;404(1):2–5.
6. Berner ES, Graber ML. Overconfidence as a cause of diagnostic error in medicine. Am J Med. 2008;121(5 Suppl):S2–23.
7. Medford-Davis L, Park E, Shlamovitz G, Suliburk J, Meyer AN, Singh H. Diagnostic errors related to acute abdominal pain in the emergency department. Emerg Med J. 2016;33(4):253–9.
8. Bowen JL. Educational strategies to promote clinical diagnostic reasoning. N Engl J Med. 2006;355(21):2217–25.
9. Rencic J, Trowbridge RL Jr., 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):1242–6.
10. Van den Brink N, Holbrechts B, Brand PLP, Stolper ECF, Van Royen P. Role of intuitive knowledge in the diagnostic reasoning of hospital specialists: a focus group study. BMJ Open. 2019;9(1):e022724.
11. Lee CY, Lai HY, Lee CH, Chen MM. Medical dispute cases caused by errors in clinical reasoning: an investigation and analysis. Healthc (Basel) 2022, 10(11).
12. Connor DM, Durning SJ, Rencic JJ. Clinical reasoning as a core competency. Acad Med. 2020;95(8):1166–71.
13. Alamoudi AA, Al Shawwa LA, Gad H, Tekian A. Team-based learning versus traditional didactic lectures in teaching clinical biochemistry at King Abdulaziz university; learning outcomes and student satisfaction. Biochem Mol Biol Educ. 2021;49(4):546–59.
14. Thistlethwaite JE, Davies D, Ekeocha S, Kidd JM, MacDougall C, Matthews P, Purkis J, Clay D. The effectiveness of case-based learning in health professional education. A BEME systematic review: BEME guide 23. Med Teach. 2012;34(6):e421–44.
15. Trullas JC, Blay C, Sarri E, Pujol R. Effectiveness of problem-based learning methodology in undergraduate medical education: a scoping review. BMC Med Educ. 2022;22(1):104.
16. Brentnall J, Thackray D, Judd B. Evaluating the clinical reasoning of student health professionals in placement and simulation settings: A systematic review. Int J Environ Res Public Health 2022, 19(2).
17. Stolper E, Van Royen P, Jack E, Uleman J, Olde Rikkert M. Embracing complexity with systems thinking in general practitioners' clinical reasoning helps handling uncertainty. J Eval Clin Pract. 2021;27(5):1175–81.
18. Higgs J, Burn A, Jones M. Integrating clinical reasoning and evidence-based practice. AACN Clin Issues. 2001;12(4):482–90.
19. Daniel M, Rencic J, Durning SJ, Holmboe E, Santen SA, Lang V, Ratcliffe T, Gordon D, Heist B, Lubarsky S, et al. Clinical reasoning assessment methods: A scoping review and practical guidance. Acad Med. 2019;94(6):902–12.
20. Charlin B, Lubarsky S, Millette B, Crevier F, Audétat MC, Charbonneau A, Caire Fon N, Hoff L, Bourdy C. Clinical reasoning processes: unravelling complexity through graphical representation. Med Educ. 2012;46(5):454–63.
21. Zheng B, Jiang X, Atkins MS. Detection of changes in surgical difficulty: evidence from pupil responses. Surg Innov. 2015;22(6):629–35.
22. Zheng B, Jiang X, Tien G, Meneghetti A, Panton ON, Atkins MS. Workload assessment of surgeons: correlation between NASA TLX and blinks. Surg Endosc. 2012;26(10):2746–50.
23. Beesley T, Pearson D, Le Pelley M. Chap. 1 - Eye Tracking as a Tool for Examining Cognitive Processes. In: *Biophysical Measurement in Experimental Social Science Research*. edn. Edited by Foster G: Academic Press; 2019: 1–30.
24. Eckstein MK, Guerra-Carrillo B, Miller Singley AT, Bunge SA. Beyond eye gaze: what else can Eyetracking reveal about cognition and cognitive development? Dev Cogn Neurosci. 2017;25:69–91.
25. Mele ML, Federici S. Gaze and eye-tracking solutions for psychological research. Cogn Process. 2012;13(Suppl 1):S261–265.
26. Atkins MS, Tien G, Khan RS, Meneghetti A, Zheng B. What do surgeons see: capturing and synchronizing eye gaze for surgery applications. Surg Innov. 2013;20(3):241–8.
27. Harezlak K, Kasprowski P. Application of eye tracking in medicine: A survey, research issues and challenges. Comput Med Imaging Graph. 2018;65:176–90.
28. McLaughlin L, Bond R, Hughes C, McConnell J, McFadden S. Computing eye gaze metrics for the automatic assessment of radiographer performance during X-ray image interpretation. Int J Med Inf. 2017;105:11–21.
29. Chun MM, Jiang Y. Contextual cueing: implicit learning and memory of visual context guides Spatial attention. Cogn Psychol. 1998;36(1):28–71.
30. Haider HFP. Information reduction during skill acquisition: the influence of task instruction. J Experimental Psychol Appl. 1999;5(2):129–51.
31. Jaarsma T, Jarodzka H, Nap M, van Merrienboer JJ, Boshuizen HP. Expertise under the microscope: processing histopathological slides. Med Educ. 2014;48(3):292–300.
32. Jaarsma T, Jarodzka H, Nap M, van Merrienboer JJ, Boshuizen HP. Expertise in clinical pathology: combining the visual and cognitive perspective. Adv Health Sci Educ Theory Pract. 2015;20(4):1089–106.
33. Brush JE Jr., Sherbino J, Norman GR. How expert clinicians intuitively recognize a medical diagnosis. Am J Med. 2017;130(6):629–34.
34. Jarodzka H, Balslev T, Holmqvist K, Nyström M, Scheiter K, Gerjets P, Eika B. Conveying clinical reasoning based on visual observation via eye-movement modelling examples. Instr Sci. 2012;40(5):813–27.
35. Krämer W. Kahneman, D. (2011): thinking, fast and slow. Stat Pap. 2013;55(3):915–915.
36. Darici D, Masthoff M, Rischen R, Schmitz M, Ohlenburg H, Missler M. Medical imaging training with eye movement modeling examples: A randomized controlled study. Med Teach. 2023;45(8):918–24.
37. Kok EM, Jarodzka H, de Bruin AB, BinAmir HA, Robben SG, van Merrienboer JJ. Systematic viewing in radiology: seeing more, missing less? Adv Health Sci Educ Theory Pract. 2016;21(1):189–205.
38. Manning D, Ethell S, Donovan T, Crawford T. How do radiologists do it? The influence of experience and training on searching for chest nodules. Radiography. 2006;12(2):134–42.
39. Södervik I, Hanski L, Boshuizen HPA, Katajavuori N. Clinical reasoning in pharmacy: what do eye movements and verbal protocols tell Us about the processing of a case task? Adv Health Sci Educ Theory Pract. 2024;29(1):45–65.
40. Dogus Darici CR, Markus M. Webcam-based eye-tracking to measure visual expertise of medical students during online histology training. GMS J Med Educ 2023, 40(5).

Publisher's note