Available online at www.sciencedirect.com
ScienceDirect
Health Professions Education 2 (2016) 99–105
www.elsevier.com/locate/hpe

The Influence of Time Pressure and Case Complexity on Physicians' Diagnostic Performance

Dalal A. ALQahtani[a,*], Jerome I. Rotgans[b], Nasr Eldin Ahmed[c], Ibrahim A. Alalwan[c], Mohi Eldin M. Magzoub[d]

[a]Department of Oral Medicine and Diagnostic Sciences (DDS), College of Dentistry – King Saud University, P.O. Box 60169, Riyadh 11545, Saudi Arabia
[b]Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore
[c]King Saud bin Abdulaziz University for Health Sciences, Riyadh, Saudi Arabia
[d]Department of Medical Education, College of Medicine, King Saud bin Abdulaziz University for Health Sciences, Riyadh, Saudi Arabia

Received 17 January 2016; accepted 18 January 2016
Available online 4 May 2016

Abstract

Purpose: Practicing medicine is a cognitively demanding task that consists of the ability to assess the patient, judge the nature of his or her complaints, and make an appropriate diagnosis. A number of factors have the potential to affect the physician's diagnostic performance negatively. Two of these factors are time pressure and case complexity. However, the empirical evidence that supports this negative influence is scant. This study experimentally investigated the effect of time pressure and the complexity of clinical cases on diagnostic accuracy.
Method: Thirty-seven senior internal medicine residents participated in this study. These residents were randomly allocated to two experimental groups (with time pressure vs. without time pressure). These residents were instructed to diagnose 8 case scenarios (4 straightforward and 4 complex cases) presented on a computer by using E-Prime® 2.0. The time pressure group received feedback after each case that they were behind schedule, whereas the control group did not receive such information. The dependent variables were the mean diagnostic accuracy and the mean processing time spent on each case during diagnosis.
Results: Participants under time pressure spent nearly the same time as the group without time pressure in diagnosing the clinical cases. The diagnostic accuracy scores did not differ significantly between the experimental and control group ($F(1,35)=0.07$, $P=0.79$, and $\eta^2=0.002$). Conversely, a main effect of case complexity was found ($F(1,35)=203.19$, $P<0.001$, and $\eta^2=0.85$). Participants processed straightforward cases faster and more accurately compared with complex cases. No interaction was found between time pressure and case complexity on diagnostic accuracy ($F(1,35)=0.003$, $P=0.96$, and $\eta^2<0.001$).
Conclusions: Time pressure did not impact the diagnostic performance, whereas the complexity of the clinical case negatively influenced the diagnostic accuracy. Further studies with the enhanced experimental manipulation of time pressure are needed to reveal the effect of time pressure, if any, on a physician's diagnostic performance.

Keywords: Time pressure; Case complexity; Clinical reasoning; Physicians; Diagnostic error

*Corresponding author. Tel.: +966 114677423; fax: +966 14679018.
E-mail address: dalalq@ksu.edu.sa (D.A. ALQahtani).
Peer review under the responsibility of King Saud bin Abdulaziz University for Health Sciences.

http://dx.doi.org/10.1016/j.hpe.2016.01.006

1. Introduction

Physicians can make mistakes. According to a report by the Institute of Medicine (IOM), 44,000–98,000 people die each year in the United States alone as a result of medical errors.[1] These errors include medication mistakes, surgical errors, the neglect of serious conditions, and diagnostic errors which form a large part of such mistakes. It is estimated that the death rate caused by the incorrect diagnosis is higher than for any other type of medical error.[2,3] A Canadian study[4] reported the incidence of adverse events among hospitals and reported that 10.5% of adverse events was related to diagnostic errors. In 2008, Berner and Graber[5] published an extensive review of studies that focus on diagnostic error. The researchers recognized that the diagnostic error rate in clinical specialties is higher (a maximum of 10–15%) compared with perceptual specialties such as radiology, dermatology, and pathology (less than 5%). Understanding the etiology of diagnostic error in clinical practice is important because the causes of diagnostic errors involve both environmental influences and cognitive factors.[6]

Practicing medicine is a cognitively demanding task that requires the ability to assess a patient, to judge the significance of signs and symptoms, and to arrive at the appropriate diagnosis. In certain clinical situations, these tasks are not easily performed, particularly when under time pressure. Physicians usually see, per visit, a high volume of cases of varying difficulty level that need diagnosis and treatment planning. Having to deal with many cases, in a limited amount of time, exerts time pressure on the physicians, which may eventually affect the quality of care provided.[7,8] Given that time pressure is a reality in medical practice, and has been linked to stress, fatigue, low job satisfaction, and suboptimal patient care,[9,10] it is important to investigate whether it also has a negative effect on the diagnostic performance of a physician.

Besides time pressure, the nature of the case has also an important influence on the diagnostic reasoning process. Studies have shown that the level of case difficulty influences diagnostic reasoning and accuracy.[11,12] It has been found that complex cases often result in medical error.[13] Combining both conditions, time pressure may hypothetically interact with case difficulty, exacerbating the probability of error. This assumption has, however, not been subjected to detailed investigation and requires further testing.

In addition to the above, it is important to realize that the diagnostic process involves a complex form of thinking, referred to as clinical reasoning, which involves multiple levels of cognition and metacognition.[14] According to Schmidt et al.[15] 'illness scripts' play an important role, which are mental representations of a disease and develop from continuous exposure to similar cases. Once an illness script is formed, it can be applied, rather effortlessly, to treating new patients. This heuristic process has been coined "non-analytical reasoning," whereas the diagnostic process involving systematic, effortful analysis of a case is referred to as "analytical reasoning" (or system 1 and system 2).[16]

It can be argued that when a physician is under time pressure, he or she has to rely more on non-analytical thinking because there is limited time for slow analytical reasoning. This is particularly a problem if the case is perceived as complex, that is, the physician does not have a well-developed illness script and needs to fall back on systematic analysis of the case.[17] But even if the physician has a developed illness script regarding the case at hand, heuristics are sometimes prone to result in cognitive biases and errors.[18,19]

To examine the extent of the potential issue of time pressure and case complexity on the accuracy of medical diagnoses, more studies are needed. The objective of the present study is to explore the effect of time pressure and case complexity, while diagnosing a clinical case, on physicians' diagnostic accuracy. We hypothesized that physicians under time pressure would spend less time in diagnosing the cases than physicians without time pressure, both for straightforward and for complex cases. Moreover, we hypothesized that the more non-analytical diagnostic mode would reduce the diagnostic accuracy scores obtained by physicians under time pressure in complex cases (but not in straightforward cases) in comparison with physicians who do not experience time pressure.

2. Method

2.1. Design

The experiment employed a 2×2 experimental study, with 'time pressure' ('under time pressure' vs. 'without time pressure') as a between-subjects factor, and 'case complexity' (straightforward cases vs. complex cases) as within subject factor. The dependent variables were the mean diagnostic accuracy scores and the mean response time for each case. The ethical approval to conduct the study was granted by the Institutional Review Board (IRB) of the National Guard Health Affairs Riyadh, Saudi Arabia.

2.2. Setting

An internal medicine residency training program in Riyadh was chosen to recruit the research project's participants. This program is considered as one of the largest programs in Saudi Arabia for training physicians. This program is accredited and operated by the Saudi Commission for Health Specialties (SCFHS), which was established in 1992.[20] The program is divided into two stages: junior residency of two years, named R1 and R2, and two years of senior residency, called R3 and R4. Through the program, the residents are exposed to a wide range of cases that cover general internal medicine and all subspecialties. The program also provides the residents with the chance to be introduced to the related specialties of dermatology, neurology, and diagnostic medical specialties.[21]

2.3. Participants

Thirty seven senior residents specializing in the internal medicine training program offered by SCFHS (2011–2012) in Riyadh were enrolled in this study. To maximize the homogeneity of the study's population, the inclusion criteria were: a) level of training: the residents were at stage R3 or R4 of their training; and b) age: less than 35 years old. Repeater residents were excluded from the study. The estimated size of the population was 100 residents. The sample was recruited from three main hospitals (King Abdulaziz Medical City, King Khalid University Hospital, and King Saud Medical City). The participants' involvement was voluntary, and informed consent was obtained from each resident. At the beginning of the study, we did not disclose the full purpose of the study because this may cause participants to think or act in ways 'during solving cases' that would yield inaccurate data. However, at the end of the study, both experimental and control participants were debriefed regarding the true objective of the experiment. Participants who completed the required task received a small financial incentive, which is equivalent to one working hour in the local context.

2.4. Materials

Eight written clinical cases, four straightforward cases and four complex cases were used for this research (see Table 1). Each case consists of a brief description of a patient's medical history, signs and symptoms, and the results of the investigations. The cases were designated into two categories based on their level of complexity: A) four straightforward cases, in that they represented problems frequently encountered by internal medicine residents. B) Four complex cases, in that they were characterized by their uncommonness and rarely seen by residents or may show an atypical presentation of diseases. The cases were written by experts in internal medicine and were used in previous research of which the data demonstrated that the eight selected cases were indeed complex or straightforward respectively.[22,23]

Table 1
Diagnosis of the clinical cases used in the experiment.

Straightforward cases
1. Community acquired pneumonia
2. Acute pericarditis
3. Liver cirrhosis
4. Addison's disease
Complex cases
1. Thyrotoxicosis
2. Septic shock secondary to pneumonia with abdominal aortic aneurysm rupture
3. Inflammatory bowel disease
4. Cushing syndrome secondary to small cell carcinoma

2.5. Procedure

This study was conducted over two months in a computer lab with residents in hospitals. Each session was 60 min long. The cases were presented to the participants using E-Prime 2.0 (Psychology Software Tools, Inc. Pittsburgh, Pennsylvania). E-Prime is a programming package for designing and running psychological experiments.[24] Upon arriving at the lab, residents were randomly allocated, either to the "with time pressure" experimental group or to the "without time pressure" control group. This allocation was performed by assigning participants alternatively to either group.

Each resident was seated in front of a computer screen and signed the informed consent form. Then, the participants were instructed to log into the computer program and work in silence without interruption. Upon logging in, the program provided further instructions. The group under time pressure was informed of the following by the initial instructions provided by the computer program:

1. You have a set of clinical cases to diagnose.
2. The available time for diagnosis is short.
3. You will be informed after each case is diagnosed, how much time remains and what proportion of the complete task remains to be done.

Time pressure perception was manipulated in this experiment by providing the participants with feedback after each case, which was composed of two bars in different colors: a green bar indicated the amount of time remaining, and a red bar indicated the number of the cases yet to be diagnosed. This feedback was independent of the participants' performance and gave them the impression that they are always behind schedule. Conversely, the group without time pressure was informed that they had a set of cases to be diagnosed and that the time allocated for the task has been proved to be sufficient.

Prior to the experiment, both groups were given two example cases to get familiarized with the procedure. The actual cases were presented to the participants in random order. The software recorded the response time for each case in seconds.

2.6. Analysis

The diagnoses provided by the participants were scored by two experts in internal medicine in a blind (i.e., without knowing the experimental condition under which the responses were given) and independent (i.e., without discussing with each other during the scoring) manner. By following a standardized procedure,[25,26] the diagnosis was judged as accurately correct, partially correct/partially incorrect or incorrect, receiving scores of 1, 0.5, and 0, respectively. A diagnosis was considered correct when the main component of the diagnosis (i.e., the main/core diagnosis) appears in the diagnosis indicated by the participant, for example: writing "Endocarditis" in the case of acute bacterial endocarditis or "hepatitis" in the case of acute viral hepatitis. A diagnosis was considered partially correct/partially incorrect when one of the constituent elements of the diagnosis appears in the diagnosis written by the participant; however, the main diagnosis was not cited. For example: writing "sepsis" as the diagnosis in the case of "pneumonia with sepsis" or "Myopathy" as the diagnosis in the case of "Hyperthyroidism". A diagnosis was considered incorrect when it did not correspond to the main diagnosis, and none of its constituent elements appears in the diagnosis written by the participant (that means, it did not fall into one of the previously noted categories). For example: writing "Acute myocardial infarction" in the case of "Aortic dissection". All statistical analyses were performed using IBM SPSS version 20.0 (Armonk, NY: IBM Corp.).

A 2×2 repeated-measures ANOVA (significance level: $P = 0.05$) with the experimental condition (time pressure vs. without time pressure) as a between-subjects factor and case complexity (complex cases vs. straightforward cases) as a within-subjects factor was conducted on the mean diagnostic accuracy scores and the mean response time obtained in the two experimental conditions. This analysis tested the hypothesis that the time pressure would reduce the diagnostic accuracy scores obtained by the group under time pressure in complex cases (but probably not in straightforward cases) in comparison to the group without time pressure.

3. Results

Thirty-nine residents participated in this study. The descriptive statistics revealed that there were two outliers for which the response time was significantly longer. One outlier was from the experimental group during the solving of difficult cases (mean response time = 520 s) and one from the control group during the solving of easy cases (mean response time = 343 s). These data points are more than 1.5 interquartile ranges (IQRs) above the third quartile. We believe these values are considered as unusual response times, which may affect the mean response time. Thus, we decided to exclude them.

The remaining 37 participants (18 experimental and 19 control) exhibited the demographics presented in Table 2. After randomization, both the experimental and control groups did not show any significant differences in age or clinical practice.

3.1. Response time

Table 3 shows the means and standard deviations of the response time of the straightforward and complex cases for both experimental conditions.

A 2×2 repeated-measures ANOVA was performed to test the effect of the time pressure (time pressure vs. without time pressure) as between-subject factor and the case complexity (complex vs. straightforward) as within-subject factor and their interaction on response time. The test revealed non-significant differences of time pressure on response time ($F(1,35) = 0.72$, $P = 0.40$, $\eta^2 = 0.02$). However, a large significant effect was found for case complexity on response time ($F(1,35) = 114.36$, $P < 0.001$, and $\eta^2 = 0.77$). Straightforward cases were diagnosed more rapidly than complex ones. This outcome constitutes supportive evidence for the validity of the difficulty level of the chosen cases. There was no significant interaction effect of time pressure and case complexity on response time ($F(1,35) = 0.04$, $P = 0.85$, $\eta^2 = 0.001$). Participants under time pressure diagnosed the cases with nearly the same speed as the group that was not under time pressure for both the straightforward and complex cases.

Table 2
Demographic information for both groups.

Conditions	Demographic features	Values
Experimental group (Under time pressure)	Number of subjects	18
	Age mean in years (range)	29.3 (26–40)
	Missing data:3	
	Gender	
	M	12
	F	6
	Clinical practice mean in years, (range)	3.9 (1–12)
Control group(Without time pressure)	Number of subjects	19
	Age mean in years (range)	28.6 (27–33)
	Missing data:1	
	Gender	
	M	14
	F	5
	Clinical practice mean in years (range)	3.6 (2–7)

Table 3
Means and standard deviations obtained for response time (in seconds) during diagnosis of the clinical cases as a function of case complexity and experimental condition.

	Under time pressure (Experimental)			Without time pressure (Control)		
	n	mean	SD	*n*	mean	SD
Straightforward	18	97.24	33.71	19	107.62	35.17
Complex	18	191.2	69.11	19	205.02	58.09

3.2. Diagnostic accuracy scores

Table 4 shows the means and standard deviations of the accurate diagnosis generated for the straightforward and complex cases under the two experimental conditions.

A 2 × 2 repeated-measures ANOVA was performed to test the effect of the time pressure (time pressure vs. without time pressure) as between-subject factor and the case complexity (complex vs. straightforward) as within-subject factor and their interaction on diagnostic accuracy. The results revealed that there was no significant between-subjects effect of time pressure on diagnostic accuracy ($F(1,35)=0.07$, $P<0.79$ and $\eta^2=0.002$). However, the results also revealed a significant within-subject effect of case complexity on the diagnostic accuracy ($F(1,35)=203.19$, $P<0.001$, $\eta^2=0.85$); which suggests that case difficulty significantly affected the diagnostic accuracy. Finally, there was no significant interaction effect of time pressure and case complexity on diagnostic accuracy ($F(1,35)=0.003$, $P=0.96$, $\eta^2<0.001$).

Table 4
Means and standard deviations obtained for means of accurate diagnosis as a function of case complexity and experimental condition.

	Under time pressure (Experimental)			Without time pressure (Control)		
	n	mean	SD	*n*	mean	SD
Straightforward	18	0.80	0.16	19	0.81	0.11
Complex	18	0.25	0.18	19	0.26	0.21

4. Discussion

In this study, we investigated the effect of time pressure and case complexity on the diagnostic performance of physicians. We hypothesized that, when doctors perform under time pressure, their diagnostic skills would be negatively affected when dealing with complex cases but not when dealing with straightforward cases. The assumption is that the induced time pressure would limit the time available to process information through deliberate analytical reasoning,[27] making the participant more depend on non-analytical, heuristic reasoning to process the case. Thus, physicians would spend less time and commit more diagnostic errors, particularly with complex cases. To test these assumptions, we conducted an experiment involving senior internal medicine residents. Straightforward and complex cases were diagnosed under time pressure or without time pressure.

Contrary to our prediction, the results suggest that doctors produced similar diagnostic accuracy scores under time pressure and without time pressure. Although there was not a significant difference between the two groups in terms of diagnostic accuracy, our results revealed large significant differences in terms of the complexity of the case. Participants spent more time—nearly twice as much—on diagnosing complex cases compared with straightforward ones. Apparently, complex problems presented uncommon features, which needed more exploration and elaboration associated with reflective practice (analytical reasoning) in medicine.[28] Conversely, straightforward cases were processed faster because they presented familiar features. This finding is consistent with those of Mamede et al.[22] who found that case ambiguity affected the diagnostic reasoning of internal medicine residents. Another study found that task difficulty activated reflection and therefore has an influence on the reasoning strategies used.[29] Interestingly, although complex cases took a longer time to diagnose compared with the straightforward ones, it did not necessarily improve the diagnostic accuracy. Most of the complex cases were not diagnosed accurately by the participants. For instance, the "septic shock secondary to pneumonia" case was only diagnosed correctly by one participant. This was despite the fact that our residents were senior; some were even eligible to take the certifying board exam. This result suggests that the time available to generate a diagnosis is not the only factor that may have impacted the diagnostic reasoning process. It appears that the level of expertise, and thus knowledge (i.e., availability of illness scripts) is a crucial factor in generating a correct diagnosis.[6,27] In other words, if cases are complex and thus knowledge is lacking, no matter how much more time was spent on reasoning about them, it did not result in a correct diagnosis because the knowledge is missing to deal with the cases.[13,23,30]

Overall, our finding that there was no significant effect of the treatment on response time nor diagnostic accuracy can mean two things. First, it is possible that time pressure has no significant effect on diagnostic reasoning. This would be in line with the conclusions by Norman and colleagues. Norman et al. divided second year internal medicine participants into two groups and requested that they diagnose 20 clinical cases; one group was requested to be fast but accurate, and the second group was requested to be slow and careful.[31] The researchers found no difference in diagnostic accuracy between the two groups.

A second possibility is that the experimental manipulation was too subtle to cause any significant effect of time pressure on diagnostic performance. The mean difference in response time between both conditions was non-significant and a meager 12 s. This may be too little to cause sufficient "damage" to the diagnostic reasoning process. As such, the instructions may have not been sufficiently clear with regard to the limited time available for the entire session and the emphasis on being quick. Although we used two colored bars to visually convey the message that the participants were behind schedule, this visual cue may have been insufficient to result in working faster. This is a limitation of the study. For future studies, it may be more effective to provide additional negative feedback to help the participants interpret the status of the time and progress bars. Providing additional negative feedback, which is independent of the actual progress, is something that needs further investigation.

Funding/Support

King Abdullah International Medical Research Center (KAIMRC) (Grant no: RC10/122) funded this study.

Other disclosures

None.

Ethical approval

This study received ethical approval (RC10/122) from the Institutional Review Board (IRB) of the National Guard Health Affairs, Riyadh, Saudi Arabia.

Previous presentations

None.

Acknowledgments

The authors would like to thank the program directors and residents of internal medicine residency program on the following hospitals: King Abdulaziz Medical City, King Khalid University Hospital, King Saud Medical City.

References

1. *To Err Is Human: Building a Safer Health System*. Washington D.C.: National Academies Press; 2000.
2. Leape LL, Brennan TA, Laird, N, et al. The nature of adverse events in hospitalized patients. Results of the Harvard Medical Practice Study II. *N Engl J Med.* 1991;324(6):377–384.
3. Thomas EJ, Studdert DM, Burstin, HR, et al. Incidence and types of adverse events and negligent care in Utah and Colorado. *Med Care.* 2000;38(3):261–271.
4. Baker GR, Norton PG, Flintoft, V, et al. The Canadian adverse events study: the incidence of adverse events among hospital patients in Canada. *Can Med Assoc J=J l'Assoc Med Can.* 2004;170(11):1678–1686.
5. Berner ES, Graber ML. Overconfidence as a cause of diagnostic error in medicine. *Am J Med.* 2008;121(5 Suppl):S2–23.
6. Graber ML, Franklin N, Gordon R. Diagnostic error in internal medicine. *Arch Intern Med.* 2005;165(13):1493–1499.
7. Linzer M, Gerrity M, Douglas JA, McMurray JE, Williams ES, Konrad TR. Physician stress: results from the physician worklife study. *Stress Health.* 2002;18(1):37–42.
8. Wetterneck TB, Linzer M, McMurray, JE, et al. Worklife and satisfaction of general internists. *Arch Intern Med.* 2002;162(6): 649–656.
9. Bovier PA, Perneger TV. Predictors of work satisfaction among physicians. *Eur J Public Health.* 2003;13(4):299–305.
10. Spickard JA. MId-career burnout in generalist and specialist physicians. *JAMA: J Am Med Assoc.* 2002;288(12):1447–1450.
11. Mamede S, Schmidt HG, Rikers R. Diagnostic errors and reflective practice in medicine. *J Eval Clin Pract.* 2007;13(1): 138–145.
12. Mamede S, Schmidt HG, RMJP Rikers, Penaforte JC, Coelho-Filho JM. Influence of perceived difficulty of cases on physicians' diagnostic reasoning. *Acad Med J Assoc Am Med Coll.* 2008;83(12):1210–1216.

13. Jacobs S, O'Beirne M, Derfiingher LP, Vlach L, Rosser W, Drummond N. Errors and adverse events in family medicine. *Can Fam Physician.* 2007;53(2):270–276.
14. Simmons B. Clinical reasoning: concept analysis. *J Adv Nurs.* 2010;66(5):1151–1158.
15. Schmidt HG, Norman GR, Boshuizen HP. A cognitive perspective on medical expertise: theory and implication. *Acad Med J Assoc Am Med Coll.* 1990;65(10):611–621.
16. Norman G. Research in clinical reasoning: past history and current trends. *Med Educ.* 2005;39(4):418–427.
17. Evans JSBT, Handley SJ, Bacon AM. Reasoning under time pressure. A study of causal conditional inference. *Exp Psychol.* 2009;56(2):77–83.
18. Evans JSBT, Curtis-Holmes J. Rapid responding increases belief bias: evidence for the dual-process theory of reasoning. *Think Reason.* 2005;11(4):382–389.
19. De Neys W. Dual processing in reasoning: two systems but one reasoner. *Psychol Sci.* 2006;17(5):428–433.
20. SCFHS-Postgrduate Programs. Available at: ⟨http://www.scfhs.org.sa/en/Pages/default.aspx⟩; Accessed 14.01.16.
21. SCFHS-Program Manual of Internal Medicine Residency Training Program. Available at: ⟨⟨http://www.scfhs.org.sa/en/MESPS/TrainingProgs/TrainingProgsStatement/InternalMedicine/Pages/ProgBook.aspx⟩; Accessed 14.01.16.
22. Mamede S, Schmidt HG, Rikers RMJP, Penaforte JC, Coelho-Filho JM. Breaking down automaticity: case ambiguity and the shift to reflective approaches in clinical reasoning. *Med Educ.* 2007;41(12):1185. 192.
23. Mamede S, Schmidt HG, Penaforte JC. Effects of reflective practice on the accuracy of medical diagnoses. *Med Educ.* 2008;42(5):468–475.
24. Hairston WD, Maldjian JA. An adaptive staircase procedure for the E-Prime programming environment. *Comput Methods Programs Biomed.* 2009;93(1):104–108.
25. Mamede S, Schmidt HG, RMJP Rikers, EJFM Custers, TAW Splinter, van Saase JLCM. Conscious thought beats deliberation without attention in diagnostic decision-making: at least when you are an expert. *Psychol Res.* 2010;74(6):692.
26. 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.
27. Evans JSBT. The heuristic-analytic theory of reasoning: extension and evaluation. *Psychon Bull Rev.* 2006;13(3):378–395.
28. Mamede S, Schmidt HG. The structure of reflective practice in medicine. *Med Educ.* 2004;38(12):1302–1308.
29. Heemskerk L, Norman G, Chou S, Mintz M, Mandin H, McLaughlin K. The effect of question format and task difficulty on reasoning strategies and diagnostic performance in internal medicine residents. *Adv Health Sci Educ Theory Pract.* 2008;13(4):453–462.
30. Weinberg NS, Stason WB. Managing quality in hospital practice. *Int J Qual Health Care: J Int Soc Qual Health Care/ISQua.* 1998;10(4):295–302.
31. Norman G, Sherbino J, Dore, K, et al. The etiology of diagnostic errors: a controlled trial of system 1 versus system 2 reasoning. *Acad Med.* 2014;89(2):277–284.