

Finding and fixing mistakes: do checklists work for clinicians with different levels of experience?

Citation for published version (APA):

Sibbald, M., De Bruin, A. B. H., & van Merriënboer, J. J. G. (2014). Finding and fixing mistakes: do checklists work for clinicians with different levels of experience? *Advances in Health Sciences Education*, 19(1), 43-51. <https://doi.org/10.1007/s10459-013-9459-3>

Document status and date:

Published: 01/03/2014

DOI:

[10.1007/s10459-013-9459-3](https://doi.org/10.1007/s10459-013-9459-3)

Document Version:

Publisher's PDF, also known as Version of record

Document license:

Taverne

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Finding and fixing mistakes: do checklists work for clinicians with different levels of experience?

Matthew Sibbald · Anique B. H. De Bruin · Jeroen J. G. van Merriënboer

Received: 10 January 2013 / Accepted: 17 April 2013 / Published online: 27 April 2013
© Springer Science+Business Media Dordrecht 2013

Abstract Checklists that focus attention on key variables might allow clinicians to find and fix their mistakes. However, whether this approach can be applied to clinicians of varying degrees of expertise is unclear. Novice and expert clinicians vary in their predominant reasoning processes and in the types of errors they commit. We studied 44 clinicians with a range of electrocardiography (ECG) interpretation expertise: novice, intermediate and expert. Clinicians were asked to interpret 10 ECGs, self-report their predominant reasoning strategy and then check their interpretation with a checklist. We found that clinicians of all levels of expertise were able to use the checklist to find and fix mistakes. However, novice clinicians disproportionately benefited. Interestingly, while clinicians varied in their self-reported reasoning strategy, there was no relationship between reasoning strategy and checklist benefit.

Keywords Reasoning strategies · Checklists · Expertise · System processing · ECG interpretation · Diagnostic errors

Introduction

Finding a mistake and fixing it is not easy. If the mistake is a wrong diagnosis, then Groopman in his popular book “How Doctors think”, suggests revisiting the decision in a careful analytic way (Groopman 2007). This process of revisiting decisions to find and fix mistakes can be formalized using checklists, an approach which has been widely

M. Sibbald (✉)
Ho Ping Kong Center for Excellence in Education and Practice, Toronto Western Hospital,
University Health Network, East Wing 8-420 399, Bathurst St, Toronto, ON M5T 2S8, Canada
e-mail: mattsibbald@gmail.com

M. Sibbald
Division of Cardiology, University of Toronto, Toronto, ON, Canada

M. Sibbald · A. B. H. De Bruin · J. J. G. van Merriënboer
School of Health Professions Education, Maastricht University, Maastricht, The Netherlands

advocated, but not well studied (Gawande 2009). This differs from the more traditional use of checklists during decision-making, where they function as a decision aid (for example see Berbaum et al. 2006). Instead, checklist use after decision-making function as a proofreading tool to detect errors. Designing such a checklist to facilitate error detection requires an understanding of how diagnostic decision-making leads to errors. Diagnostic decisions are a form of global or summative assessment requiring the integration of multiple key variables (Kahneman 2011). Errors can occur when clinicians overlook key variables, are swayed by irrelevant variables, or lack necessary knowledge to formulate a correct hypothesis. A checklist focusing clinician attention on key variables can provide an opportunity to correct errors caused by overlooking key variables or overweighing irrelevant variables. While preliminary data suggests benefits to this approach among experts, it is unclear whether this benefit applies equally across all levels of expertise. There are two concerns in applying this approach across all levels of expertise. First, there are obvious differences in how novices and experts approach diagnostic problems (Ericsson 2004; Patel and Groen 1991). Second, novices and experts likely commit errors for different reasons (Groves et al. 2003).

A well-studied difference between a novice and expert approach is the predominant type of reasoning processing involved. Reasoning processes can be categorized into two systems: automatic, intuitive, effortless processes termed system 1 and methodical, analytical, effortful processes termed system 2 (Evans 2010). A useful analogy is to think of system 1 processes as a search engine that jumps to decisions, whereas system 2 processes are akin to a calculator that arrives at a decision through incremental steps. For instance, when trying to retrieve a lost wallet, you could search your brain and realize where the wallet was left using a system 1 process, or systematically retrace your steps using a system 2 process. While the distinction between these two thought processes is somewhat arbitrary, system 1 can be recognized in think-aloud protocols, short decision times, and even by clinician's self-reflection. In a study of clinicians interpreting ECGs, self-report of system 1 processes was correlated with short decision times and objective assessment of reasoning processes in post hoc think-aloud protocols (Sibbald and De Bruin 2011). In these post hoc think-aloud protocols, system 1 processing was objectively discerned when clinicians did not deliberate, but described immediate recognition of a diagnosis or jumping to an answer without referring to any intermediate steps.

As clinicians acquire expertise, the proportion of system 1 or automatic processing increases (Eva and Cunningham 2006). Because system 1 processes are more prone to influence by external factors (i.e., biases), encouraging system 2 processes among experts has been frequently proposed to reduce error (Croskerry 2003). Checklists may be a useful tool for this purpose. Using a checklist after a decision to check decision-making is an inherently analytic task, likely to encourage system 2 processing. Predictably, this type of checklist improved diagnostic decisions amongst experts by approximately 10 %. However, novices and intermediates tend to use system 1 processes less often, and may even benefit from instruction to encourage system 1 processes. Prior study suggested that novices learn better when encouraged to use both systems (Ark et al. 2006, 2007). While learning and checking are two different tasks, these findings cast doubt on whether non-experts will benefit from the use of checklists to find and fix mistakes.

Experts and novices are also prone to different types of errors. While classifying errors has been criticized because it is prone to assumption and subjective interpretation, differences are nevertheless apparent. Experts are more likely to overlook key data, but less likely to make a mistake in generating a hypothesis. In contrast, novices are more likely to formulate a faulty hypothesis and less likely to overlook data (Groves et al. 2003). While

novices are less likely to overlook data, they have more difficulty separating relevant from irrelevant data (Gegenfurtner et al. 2011). Overweighing irrelevant data may partly explain the novices' tendency to generate faulty hypotheses.

Whether a checklist can help detect both types of errors is unclear. We hypothesize that checklists comprised of key variables might help both novices and experts, but in different ways. For experts, this type of checklist encourages review of all relevant data, potentially reducing the errors noted in data collection. For novices, this type of checklist may limit attention to key variables thereby enabling better hypothesis generation. However, the literature surrounding this type of strategy includes conflicting reports of success. In one experiment, a computer program suggesting diagnoses based on user-entered consistent and inconsistent clinical features resulted in a disproportionate benefit amongst novices (Friedman et al. 1999). However, whether the benefit occurred because the intervention focused attention on relevant variables or because alternate hypotheses were suggested is unknown. Checklists can only do the former and not the latter. Trainees asked to manually generate hypotheses after listing consistent and inconsistent clinical features, did not improve their diagnostic accuracy (Kulatunga-Moruzi et al. 2001). However, this study did not force novices to focus on key variables, which might be a necessary ingredient in formulating correct hypotheses.

Therefore, the purpose of this study is to investigate the benefit of checklists in finding and fixing diagnostic errors across the spectrum of expertise. The secondary purpose of the study was to understand whether the benefit of checklists was dependent on the predominant type of thought process initially used in making a decision. We hypothesized that (1) checklists that focused attention on key variables would enable clinicians to detect and fix errors regardless of their level of expertise; (2) checklists that focused attention on key variables would have a disproportionate benefit when system 1 predominant processing was used.

Methods

Participants

Novices, intermediates and experts with 2–3 years, 4–5 years and 8–11 years of ECG interpretation experience were recruited by email. All participants were recruited in 2012 from the University of Toronto. Novices were recruited from the second and third year medical school class, intermediates from the internal medicine residency program and experts from the cardiology fellowship program.

Material

10 ECGs were selected from a common ECG textbook to cover a broad range of material (O'Keefe et al. 2010). Diagnoses are listed in Table 1.

The checklist was based on six key variables in ECG interpretation. It was originally derived by two experts using two ECG textbooks and was used in a prior study. The checklist items included: 1. Did I get the rate right? 2. Did I scan the entire strip to confirm the rhythm? 3. Is the axis really normal? 4. Did I check each chamber (right atrium, left atrium, right ventricle, left ventricle) for hypertrophy? 5. Did I check each territory (anterior, posterior, inferior, lateral, septal) for ischemia? 6. Did I miss any abnormal intervals: PR, QRS, QT?

Protocol

The study was constructed using an online survey platform (SurveyGizmo, www.surveymgizmo.com). Participants were introduced to a self-reflective reasoning scale to judge whether system 1 or 2 processes were predominantly used in interpreting an ECG. The instrument asks “How did you interpret this ECG?” with four possible answers: (1) it was a pattern I immediately recognized, (2) mostly through pattern recognition, (3) mostly through analyzing its different aspects, (4) I analyzed each of its components. Participants studied a worked example and then completed a sample ECG interpretation using the instrument. This instrument was used previously in a similar experimental setup with intermediates (Sibbald and De Bruin 2011). In this previous study, two observers independently applied the same four-point scoring system based on post hoc think-aloud protocols with good agreement between observers and participants. Interpretation times were also recorded, showing an expected 40 % increase with self-report of analysis.

Participants were asked to interpret each of the 10 ECGs with the following instruction: “Interpret this ECG as you normally would”. Participants completed a self-reflective reasoning scale after interpreting each ECG. Participants were then shown each ECG again with their original interpretation. Participants were asked to check their interpretation with the checklist. Time taken to interpret and check the interpretation were measured.

Outcome

An ECG expert, blinded to participant group and self-reported reasoning, reviewed each interpretation and counted errors. Each omitted diagnosis and each additional incorrect diagnosis was counted as an error.

Errors were summed for each participant pre and post checklist use and compared with a repeated ANOVA using reasoning strategy as a within subject variable and experience level as a between subject variable with a full factorial design. Interpretation time and checklist time were also compared with repeated ANOVAs using individual ECGs as within subject variables and experience level as a between subject variable. Analysis was conducted with SPSS 20 (IBM, Redmond).

Ethical approval was obtained by the Health Sciences Research Ethics Board, University of Toronto.

Results

Forty four participants analyzed 10 ECGs each: 16 novices, 20 intermediates and 8 experts. The mean number of errors per ECG was 2.9 (range 0–7, SD = .9) and per participant was 29 (range 6–41, SD = 9.1). Errors differed by experience level ($F_{2,41} = 12.1$, $\eta_p^2 = .37$, $p < .0001$, Table 2). Experts made fewer errors than both novices and intermediates (mean difference \pm standard error -15 ± 3 and -13 ± 3 respectively, both $p < .0001$).

The mean time to interpret each ECG was 103 s (SE = 12 s). Interpretation time tended to vary by experience level ($F_{2,41} = 2.9$, $\eta_p^2 = .14$, $p = .06$). Novices took more time than either intermediates or experts (mean difference \pm standard error 51 ± 24 and 63 ± 31 s respectively, both $p < .05$). Intermediates took similar time compared with experts (mean difference \pm standard error 13 ± 30 s, $p = .67$).

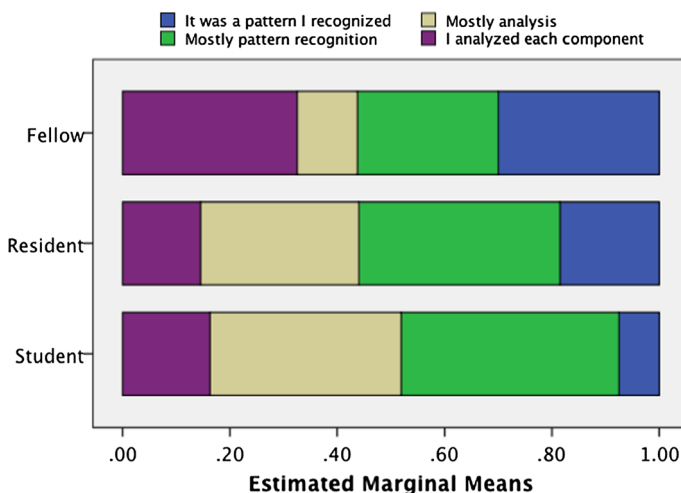
Self-reported reasoning varied by experience level ($F_{6,41} = 3.2$, $\eta_p^2 = .12$, $p = .02$, Fig. 1). As experience increased, so did self-report of “it was a pattern I immediately

Table 1 ECG diagnoses

ECG	Diagnoses
1	Sinus arrest with junctional escape, prolonged QT, premature ventricular contractions in a bigeminal pattern and nonspecific ST changes. Consider digoxin toxicity
2	Sinus rhythm, right bundle branch block, right axis deviation, right ventricular hypertrophy with strain pattern
3	Sinus rhythm, DDD pacer
4	Short run of narrow complex tachycardia (most likely atrioventricular reentry tachycardia) followed by sinus rhythm with left atrial enlargement
5	Sinus arrhythmia, accessory (Wolf Parkinson White) pathway
6	Sinus tachycardia, anterior and inferior ST elevation consistent with acute MI, right bundle branch block and left posterior fascicular block
7	Atrial flutter with 2:1 block and left anterior fascicular block
8	Sinus rhythm, Brugada pattern
9	Pre-excited atrial fibrillation with very rapid rates up to 300
10	Sinus rhythm, Mobitz type 1 s degree AV block, junctional escape and acute inferior ST elevation MI

recognized” ($F_{2,41} = 7.1$, $\eta_p^2 = .26$, $p = .002$). Post hoc testing with LSD correction revealed significant differences between all three groups. Experts reported this type of reasoning more often than intermediates and novices (mean difference \pm standard error $.11 \pm .05$, $p = .03$; $.23 \pm .06$, $p = .001$ respectively). Intermediates reported this type of reasoning more often than novices (mean difference \pm standard error $.12 \pm .06$, $p = .06$).

Self-report of “mostly through analyzing its components” also varied by experience level ($F_{3,41} = 4.5$, $\eta_p^2 = .181$, $p = .02$). Post hoc testing with LSD correction revealed experts were less likely to report this type of reasoning than intermediates and novices (mean difference \pm standard error $-.24 \pm .08$, $p = .01$; $-.18 \pm .08$, $p = .03$ respectively). However,

**Fig. 1** Reasoning strategy by experience level

the difference between intermediates and novices was not significant (mean difference \pm standard error $-.06 \pm .06, p = .34$).

Self-report of “mostly through pattern recognition” and “I analyzed its components” did not vary by experience level ($F_{3,41} = 1.1, p = .35$ and $F_{3,41} = 1.6, p = .22$ respectively).

The number of errors varied by self-reported reasoning ($F_{3,39} = 5.9, \eta_p^2 = .31, p = .002$). Self-report of “it was a pattern I immediately recognized” and “I analyzed each component” were both associated with fewer mistakes than self-report of “mostly through pattern recognition” and “mostly through analyzing its components” (all p 's $< .03$; Table 3).

Participants fixed a mean number of 1.6 errors when using the checklist (SE = .44, $F_{2,41} = 12.2, \eta_p^2 = .37, p = .001$, Table 2). This varied inversely with experience (mean \pm standard error $2.6 \pm .68, 1.6 \pm .61$ and $.05 \pm .97$ for novices, intermediates and experts respectively), though these differences were not statistically significant ($F_{2,41} = 1.7, \eta_p^2 = .08, p = .20$). The number of errors fixed did not vary by self-reported reasoning used for initial interpretation ($F_{2=3,39} = .21, \eta_p^2 = .11, p = .21$). Time spent using the checklist was 80 ± 17 s for novices, 57 ± 15 s for intermediates and 20 ± 24 s for experts ($F_{2, 40} = 2.1, \eta_p^2 = .10, p = .13$).

Discussion

In this study, we confirmed our hypothesis that checklists would be helpful to clinicians regardless of expertise level. We found that novices, intermediates and experts were all able to use a checklist to find and fix mistakes in ECG interpretation. The magnitude of the benefit was not small. On average, one mistake was detected using the checklist for every 8 ECG interpretations that were scrutinized. This benefit is similar to that described among experts previously, as well as the benefit found among intermediates using with a similar style checklist for cardiac physical exam (Sibbald et al. 2013a, b). This benefit stands in

Table 2 Errors, self reported reasoning and experience level

	N	Errors (mean \pm standard error)			
		Initial interpretation	Corrected with checklist	% corrected with checklist	Post checklist use
Self reported reasoning (by ECG)					
It was a pattern I immediately recognized	73	4.2 \pm 0.7	0.3 \pm 0.1	7.1 \pm 2.4	3.9 \pm 0.7
Mostly through pattern recognition	161	10.9 \pm 1.1	0.8 \pm 0.3	7.3 \pm 2.8	10.1 \pm 1.1
Mostly through analyzing its components	125	9.0 \pm 1.0	0.5 \pm 0.2	5.6 \pm 2.2	8.5 \pm 1.0
I analyzed each component	81	4.5 \pm 0.9	0.2 \pm 0.1	4.4 \pm 2.2	4.3 \pm 0.9
Level of expertise (by participant)					
Novice	16	32.4 \pm 1.8	2.6 \pm 0.7	8.0 \pm 2.2	29.8 \pm 1.9
Intermediate	20	30.5 \pm 1.6	1.6 \pm 0.6	5.2 \pm 2.0	28.9 \pm 1.7
Expert	8	16.5 \pm 2.5	0.5 \pm 1.0	3.0 \pm 6.1	16.0 \pm 2.8
Total	44	26.5 \pm 1.2	1.6 \pm 0.4	6.0 \pm 1.5	24.9 \pm 1.2

contrast to prior study of checklists during decision making, where they function as a decision aid rather than a proofreading tool (e.g., Berbaum et al. 2006). While these differences may be due to the skill or population studied, we argue that the timing of checklist application is the most important variable (for an in-depth discussion of this issue, see Sibbald et al. 2013a, b).

Importantly, novices benefited just as much if not more than intermediates and experts. The simplest explanation for this finding is that novices were more likely to find mistakes because they were more likely to make mistakes. However, other explanations are possible. For instance, checklists may help experts and novices differently. This study did not investigate what participants were actually doing when they were checking their interpretations with the checklist. Nevertheless, the differences in time spent on the task between experts and non-experts suggest that experts and non-experts were completing the task differently. Investigating how experts and non-experts use a checklist may be fruitful ground for future research. In particular, eye tracking technology may be a useful modality to explore exactly what experts and non-experts are doing when using a checklist. Such data would provide significant insight into the mechanism of checklist benefit.

Our data on reasoning strategies also provide an explanation for the disproportionate benefit among novices. Experts reported combination reasoning (mostly pattern or mostly analysis) less often than pure reasoning (only pattern recognition or only analysis). This provides evidence that experts are tailoring their approach to the ECG, selecting pure analytic or pattern recognition strategies when it is best suited to the content. This finding resembles a form of reflective expertise, where expert behavior includes the ability to efficiently mix problem solving strategies to optimize success (van Merriënboer 1997). By tailoring their reasoning strategy, experts may reduce the need for a checklist. This is also a suggested subject for future research. While tailoring a reasoning strategy might be a surrogate marker for expertise, it may also be a teachable skill that could accelerate the development of expertise.

Interestingly, our second hypothesis that checklists would have a disproportionate benefit where system 1 strategies were used was not supported by our data. The benefit of the checklist did not vary across different self-reported reasoning strategies. This suggests the checklist was not simply a tool to “balance” reasoning strategies. If this were the case, a checklist presumed to facilitate system 2 processes should have disproportionate benefit where predominantly system 1 or pattern recognition processes caused the errors. It is also possible that the checklist facilitates both types of processes. However, a previous study with a similar model suggested only analytic review of ECG interpretations was helpful (Sibbald and De Bruin 2011). Alternately, our use of retrospective reporting of reasoning may be insufficiently sensitive to categorize different reasoning strategies.

Several limitations of this study are worth mentioning. First, only a single skill was assessed. This strategy of checklist use to find and fix mistakes needs to be tested in other areas before widespread adoption can be promoted. The approach may be better suited to diagnostic test interpretation where key variables are readily isolated and all information is present in front of the participant. Second, we did not determine whether the benefit of checklists use was related to the additional time spent checking the interpretation or the checklist itself. Prior study compared checking with a checklist versus checking without specific direction and found benefit only with use of a checklist (Sibbald et al. 2013a, b). This suggests an added benefit to checklist use, but is a worthwhile question for future study. Third, while participants varied in experience level, they were all from a single center. Because attitudes towards checklists may play a role in their efficacy, replication in another setting seems prudent.

Table 3 Comparison of errors between different reasoning strategies

All mean difference (I–J) ± standard error	Reasoning strategy (J)			
	It was a pattern I immediately recognized	Mostly through pattern recognition	Mostly through analyzing its components	I analyzed each component
Reasoning strategy (I)				
It was a pattern I immediately recognized	NA	−5.2 ± 1.4 <i>p</i> = 0.001	−3.5 ± 1.2 <i>p</i> = 0.01	−0.2 ± 1.2 <i>p</i> = 0.91
Mostly through pattern recognition	5.2 ± 1.4 <i>p</i> = 0.001	NA	1.7 ± 1.6 <i>p</i> = 0.29	5.1 ± 1.8 <i>p</i> = 0.01
Mostly through analyzing its components	3.5 ± 1.2 <i>p</i> = 0.01	−1.7 ± 1.6 <i>p</i> = 0.29	NA	3.3 ± 1.6 <i>p</i> = 0.05
I analyzed each component	0.02 ± 1.2 <i>p</i> = 0.91	−5.1 ± 1.8 <i>p</i> = 0.01	−3.3 ± 1.6 <i>p</i> = 0.05	NA

In summary, checklists comprised of key variables were helpful to all levels of expertise finding and fixing mistakes in ECG interpretations. This benefit was particularly pronounced among novices, and did not vary based on initial reasoning strategy. Adopting a checklist strategy to confirm visual test interpretations holds significant promise to reduce error.

Acknowledgments M.S. received support for this work from the Ho Ping Kong Center for Excellence in Education and Practice and the Peter Munk Cardiac Center, University Health Network.

References

- Ark, T. K., Brooks, L. R., & Eva, K. W. (2006). Giving learners the best of both worlds: Do clinical teachers need to guard against teaching pattern recognition to novices? *Academic Medicine*, *81*(4), 405–409.
- Ark, T. K., Brooks, L. R., & Eva, K. W. (2007). The benefits of flexibility: The pedagogical value of instructions to adopt multifaceted diagnostic reasoning strategies. *Medical Education*, *41*(3), 281–287.
- Berbaum, K., Franken, E. A., Jr, Caldwell, R. T., & Scharz, K. M. (2006). Can a checklist reduce SOS errors in chest radiography? *Academic Radiology*, *13*(3), 296–304.
- Croskerry, P. (2003). Cognitive forcing strategies in clinical decisionmaking. *Annals of Emergency Medicine*, *41*(1), 110–120.
- Ericsson, K. (2004). Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Academic Medicine*, *79*(10), S70–S81.
- Eva, K., & Cunningham, J. (2006). The difficulty with experience: Does practice increase susceptibility to premature closure. *Journal of Continuing Education Health Professions*, *26*, 192–198.
- Evans, J. (2010). *Thinking twice: Two minds in one brain*. Oxford: Oxford University Press.
- Friedman, C., Elstein, A., & Wolf, F. (1999). Enhancement of clinicians: Diagnostic reasoning by computer-based consultation: A multisite study of 2 systems. *JAMA, the Journal of the American Medical Association*, *282*(19), 1851–1856.
- Gawande, A. (2009). *The checklist manifesto: How to get things right*. New York: Metropolitan Books.
- Gegenfurtner, A., Lehtinen, E., & Säljö, R. (2011). Expertise differences in the comprehension of visualizations: A meta-analysis of eye-tracking research in professional domains. *Educational Psychology Review*, *23*(4), 523–552.
- Groopman, J. (2007). *How doctors think*. Boston: Houghton Mifflin Company.
- Groves, M., O'Rourke, P., & Alexander, H. (2003). Clinical reasoning: The relative contribution of identification, interpretation and hypothesis errors to misdiagnosis. *Medical Teacher*, *25*(6), 621–625.

- Kahneman, D. (2011). *Thinking, fast and slow*. Toronto, ON: Doubleday.
- Kulatunga-Moruzi, C., Brooks, L., & Norman, G. (2001). Coordination of analytic and similarity-based processing strategies and expertise in dermatological diagnosis. *Teaching and Learning in Medicine*, 13(2), 110–116.
- O’Keefe, J., Hammill, S., Freed, M., & Pogwidz, S. (2010). *The ECG criteria book* (2nd ed.). Sudbury: Jones and Bartlett Publishers.
- Patel, V., & Groen, G. (1991). The general and specific nature of medical expertise: A critical look. In K. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise: Prospects and limits* (pp. 93–125). Cambridge: Cambridge University Press.
- Sibbald, M., & De Bruin, A. B. (2011). Feasibility of self-reflection as a tool to balance clinical reasoning strategies. *Advances in Health Science Education*. doi:10.1007/s10459-011-9320-5.
- Sibbald, M., de Bruin, A. B., Cavalcanti, R. B., & van Merriënboer, J. G. (2013a). Do you have to re-examine to reconsider your diagnosis? Checklists and cardiac exam. *BMJ Quality Safety*, 22(4), 333–338. doi:10.1136/bmjqs-2012-001537.
- Sibbald, M., de Bruin, A. B., & van Merriënboer, J. J. G. (2013b). Do checklists improve experts’ diagnostic decisions? *Medical Education*, 47(3), 301–308.
- van Merriënboer, J. J. G. (1997). *Training complex cognitive skills: A four-component instructional design model for technical training*. New Jersey: Educational Technology Publications.