

Checklists improve experts' diagnostic decisions

Citation for published version (APA):

Sibbald, M., de Bruin, A. B. H., & van Merriënboer, J. J. G. (2013). Checklists improve experts' diagnostic decisions. *Medical Education*, 47(3), 301-308. <https://doi.org/10.1111/medu.12080>

Document status and date:

Published: 01/03/2013

DOI:

[10.1111/medu.12080](https://doi.org/10.1111/medu.12080)

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.

Checklists improve experts' diagnostic decisions

Matthew Sibbald,^{1,2} Anique B H de Bruin³ & Jeroen J G van Merriënboer³

CONTEXT Checklists are commonly proposed tools to reduce error. However, when applied by experts, checklists have the potential to increase cognitive load and result in 'expertise reversal'. One potential solution is to use checklists in the verification stage, rather than in the initial interpretation stage of diagnostic decisions. This may avoid expertise reversal by preserving the experts' initial approach. Whether checklist use during the verification stage of diagnostic decision making improves experts' diagnostic decisions is unknown.

METHODS Fifteen experts interpreted 18 electrocardiograms (ECGs) in four different conditions: undirected interpretation; verification without a checklist; verification with a checklist, and interpretation combined with verification with a checklist. Outcomes included the number of errors, cognitive load, interpretation time and interpretation length. Outcomes were compared in two analyses: (i) a comparison of verification conditions with and without a checklist, and (ii) a comparison of all four conditions. Standardised scores for each outcome were used to calculate the efficiency of a checklist and to weigh its relative benefit against its

relative cost in terms of cognitive load imposed, interpretation time and interpretation length.

RESULTS In both analyses, checklist use was found to reduce error (more errors were corrected in verification conditions with checklists [0.29 ± 0.77 versus 0.03 ± 0.61 errors per ECG], and fewer net errors occurred in all conditions with checklists [0.39 ± 1.14 versus 1.04 ± 1.49 errors per ECG]; $p < 0.01$ for both). Checklists were not associated with increased cognitive load (verifications with and without checklists: 3.7 ± 1.9 and 3.3 ± 2.0 , respectively; conditions with and without checklists: 4.0 ± 1.8 versus 3.9 ± 2.0 , respectively [$p =$ not significant for both]). Checklists resulted in greater interpretation times and lengths ($p < 0.01$ for all). However, checklists were efficient in terms of the cognitive load invested, interpretation time and interpretation length ($p < 0.01$ for all).

CONCLUSIONS Among ECG interpretation experts, checklist use during the verification stage of diagnostic decisions did not increase cognitive load or cause expertise reversal, but did reduce diagnostic error.

Medical Education 2013; 47: 301–308
doi:10.1111/medu.12080

Discuss ideas arising from this article at
[www.mededuc.com 'discuss'](http://www.mededuc.com/discuss)



¹Ho Ping Kong Center for Excellence in Education and Practice, University Health Network, Toronto, Ontario, Canada

²Division of Cardiology, Department of Medicine, University of Toronto, Toronto, Ontario, Canada

³Department of Educational Development and Research, Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, the Netherlands

Correspondence: Dr Matthew Sibbald, Toronto Western Hospital, East Wing 8-420, Division of Internal Medicine, Department of Medicine, University of Toronto, 399 Bathurst Street, Toronto, Ontario M5T 2S8, Canada. Tel: 00 1 416 603 6412; E-mail: mattsibbald@gmail.com

INTRODUCTION

To make a medical diagnosis, a large number of interacting variables must be integrated into a summative decision. Given that working memory has a finite capacity, integrating such a large number of variables can quickly exhaust cognitive resources.¹ Therefore, the cognitive load involved in making medical diagnoses is often high.

Experts are able to lower the cognitive load involved in making diagnostic decisions.² Dual processing theory offers unique insights into how this is accomplished. Dual processing refers to two parallel systems of making decisions: intuitive, subconscious thinking (system 1), and analytic, conscious thinking (system 2).^{3–5} Experts use more system 1 processing.^{6,7} Rather than holding all relevant variables in working memory, experts recognise patterns they have seen before using subconscious system 1 processing. This involves less cognitive load than an attempt to analyse all interacting variables using system 2. In addition, experts have more efficient system 2 processes. They favour domain-specific strategies (e.g. using a schema) over the higher cognitive load domain-general strategies (e.g. testing one hypothesis at a time) used by novices.⁸

Other insight into how experts reduce cognitive load can be found in the expertise literature. Experts apply knowledge templates constructed from previous experiences, termed 'illness scripts'.⁹ These scripts lower cognitive load by reducing the large number of variables involved in diagnostic decisions to a few key variables relevant to a specific circumstance. How system processing relates to illness scripts has not been formally studied. However, it is likely that both system 1 and system 2 processing are involved in the use of an illness script. Selecting a script is likely to represent a system 1-driven process.¹⁰ By contrast, the application of a script probably involves system 2 processing to check key variables using domain-specific strategies.

Despite this ability to lower cognitive load, experts still make errors. In the cognitive psychology literature base, these errors are often viewed as a consequence of over-reliance on system 1 processing.^{7,11,12} However, forced use of system 2 processing has also been associated with error.¹³ In the literature on decision making, errors are often traced to systematic biases in how variables are considered. For example, we tend to suppress incongruences¹⁴ and ignore missing information.¹⁵ In addition, individuals asked to collect information about a hypothesis favour information that confirms their beliefs.¹⁶

Checklists are a potentially ideal tool with which to combat diagnostic error.¹⁷ A checklist composed of key variables might be used as a decision aid as it can mimic expert illness scripts. However, whereas illness scripts can be idiosyncratic and individual, a checklist ensures all key variables are assessed.⁹ In addition, checklists encourage system 2 processing and can force independent re-examination of all relevant information.¹⁴ Prior evidence suggests that this improves summative decision making in different contexts such as pilot responses to in-flight emergencies, personnel decisions on hiring the right person, and the modification of complex building plans during construction.^{14,17}

However, it is unclear whether a checklist approach can be applied to medical experts. Asking an expert to use a checklist risks increasing the cognitive load of the decision-making process. It might force the expert to abandon his or her own expert processes, ironically resulting in 'expertise reversal' or worsened performance.^{18,19} However, whether or not expertise reversal occurs may depend on *when* a checklist is used in the decision-making process. The decision-making process can be divided into two stages: interpretation, and verification. Checklist use in the interpretation stage is likely to result in increased cognitive load and expertise reversal.^{13,18,19} Whether such expertise reversal also occurs when a checklist is used during the verification stage, after the expert has had a chance to use his or her own expert processes, is unknown. Furthermore, merely *suggesting* that a checklist should be used in the verification stage might derail an expert's approach to the initial interpretation stage (even if the checklist is not meant to be applied in the interpretation stage).

If checklists do improve performance, it is unclear whether they can do this efficiently. An ideal diagnostic decision process should result in a correct and error-free decision, impose the least amount of cognitive load and use the least amount of time. It should also result in a decision that can be communicated with the least amount of written description possible (i.e. interpretation length). Checklists are likely to add to the cognitive load, time spent and written length of the interpretation. If checklists do improve expert performance, does this improvement outweigh any increases in cognitive load, time spent and interpretation length?

Calculating efficiency, such as by weighing a performance benefit against an increase in cognitive load, is one method of assessing the relative trade-offs

between two measured variables. Within the cognitive load literature, the calculation of efficiency was originally described to compare learning tools.²⁰ However, it can also be applied to the study of errors to compare trade-offs in error reduction against expected increases in cognitive load (cognitive efficiency), time spent (time efficiency) and interpretation length (length efficiency).

This study sought to understand whether the use of a checklist during the verification stage would improve or harm expert diagnostic decisions. We hypothesised that checklist use would reduce the number of errors and not result in increased cognitive load or expertise reversal if the checklist was used in a separate verification stage. By contrast, we hypothesised that making experts aware of the need to verify using a checklist before the interpretation stage would result in increased cognitive load, expertise reversal and a greater number of errors. If the use of a checklist did reduce the number of errors, we planned to determine whether this improvement was efficient in terms of cognitive load, time and interpretation length.

METHODS

Population

Fifteen experts (cardiology fellows with 8–11 years of experience in electrocardiogram [ECG] interpretation) participated in this study in February and March 2012.

ECG model

Eighteen ECGs from the 'difficult section' of an ECG textbook were selected.²¹ Two experts chose the ECGs based on their difficulty and the unambiguity of a defensible answer. The ECGs were presented in random order on an online survey tool as previously reported.²²

Checklist

Key variables in ECG interpretation were derived from several textbooks^{23,24} and collated as a checklist. Two experts reviewed and revised the checklist. The checklist included: (i) calculate the rate; (ii) scan the entire strip to confirm the rhythm; (iii) consider chamber hypertrophy, one chamber at a time; (iv) look for acute or chronic ischaemia: q waves (or tall R wave in V1), ST changes, and (v) check the intervals: PR, QRS, QT.

ECG interpretations

Subjects were asked to provide a summative interpretation of 18 different ECGs under each of four conditions: (i) undirected (ECGs 1–12); (ii) verification without a checklist (ECGs 1–6); (iii) verification with a checklist (ECGs 7–12), and (iv) interpretation and verification with a checklist (ECGs 13–18).

In condition 1, experts were asked to interpret the first 12 ECGs using their usual decision and verification process. In condition 2, experts were again shown the first six of these ECGs and asked to verify their interpretation. This measures the benefit of prolonging or drawing attention to the verification phase without explicitly prescribing the use of a checklist. In condition 3, experts were shown the remaining six ECGs from condition 1 and asked to verify their interpretation using a checklist. In condition 4, experts were asked to interpret and verify with a checklist six previously unexamined ECGs. Whereas conditions 3 and 4 both involved systematic verification with a checklist, condition 4 was included to identify whether expertise reversal would occur when checklist use was prescribed prior to initial decision making. Although the ECGs were presented in random order, experts always progressed through conditions 1–4 in numerical order to avoid the inadvertent use of the checklist in a non-checklist condition.

Outcome measures

Subjects rated their cognitive load using a previously validated 9-point scale after interpreting each ECG.²⁵ Each interpretation was timed electronically. Two experts blinded to the condition reviewed each interpretation and counted errors. Each diagnosis that was omitted and each incorrect diagnosis listed was scored as a mistake. Interpretation length was calculated by counting the number of characters written.

Analysis

All analyses were performed using SPSS Version 20 (IBM Corp., Armonk, NY, USA). Conditions 2 and 3 involved ECGs already interpreted in condition 1. Therefore, counted errors and interpretation lengths in these conditions represented the difference between condition 2 or 3 and condition 1. Independent *t*-tests were used to compare error detection, and changes in interpretation length, cognitive load and interpretation time between conditions 2 and 3.

Standardised scores were created for all four variables: error correction; cognitive load; interpretation time, and interpretation length. These scores were used to calculate cognitive load efficiency, time efficiency and length efficiency using a previously described formula.²⁶ For instance, cognitive efficiency is: $Z_{\text{performance}} - Z_{\text{cognitive load}}/\sqrt{2}$. Equal increases in performance and cognitive load would represent an intervention with neutral cognitive efficiency. By contrast, a disproportionate increase in performance compared with cognitive load would represent a cognitively efficient intervention. These efficiencies can be quantified using plots comparing standardised scores for performance (y-axis) against standardised scores for cognitive load or other variables (x-axis) (Fig. 1). The diagonal line at which standardised scores of performance equal standardised scores for cognitive load represents neutral efficiency. Points above this line indicate efficiency and points below indicate inefficiency. Mathematically, efficiency is the distance from any given point to this line. Independent *t*-tests were used to compare efficiencies.

All four conditions were compared using generalised linear modelling. Two binary variables were introduced as fixed factors: task type, and checklist use. Task type was coded as '1' in conditions which involved interpretation ('interpretation tasks' [conditions 1 and 4]) and as '0' in conditions without interpretation ('verification tasks' [conditions 2 and 3]). Checklist use was coded as '1' in conditions which involved checklists (conditions 3 and 4) and as '0' in conditions without a checklist (conditions 1 and 2). Errors in verification conditions (2 and 3) were calculated as the difference between errors made in

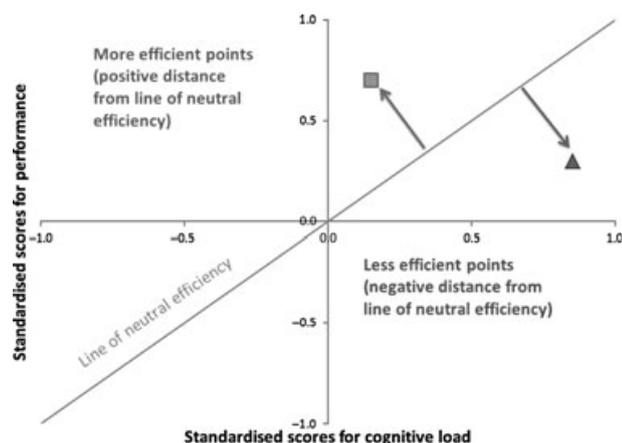


Figure 1 Mathematical determination of efficiency using standardised scores and distance measured from a line of neutral efficiency

these conditions and errors made in the undirected condition 1. Models were constructed for errors, cognitive load, interpretation time, cognitive efficiency, time efficiency and length efficiency using three factors: task type; checklist use, and the interaction term.

Survey data

When they had interpreted all ECGs, subjects were surveyed about their routine use of a checklist-like approach. They were asked how often they used this approach, and to estimate how often it resulted in error detection and how much time it involved. Ethical approval was obtained from the review board of the University of Toronto.

RESULTS

Across the 12 ECGs presented in condition 1, each of which was analysed by 15 experts (178 interpretations analyzed and 2 blank interpretations excluded), 282 diagnostic errors were made (mean = 1.58 per ECG interpretation).

Verification tasks (comparison of conditions 2 and 3)

Experts corrected errors when verifying their ECG interpretations (Table 1). However, a net benefit was present only when checklists were used. On average, experts corrected 19.2% of their errors (0.29 of the 1.51 errors per ECG) with a checklist compared with 1.8% (0.03 of the 1.66 errors per ECG) without a checklist ($t_{167} = 2.5$, $p < 0.01$). Surprisingly, cognitive load did not differ between verifications with and without checklists (3.7 ± 1.9 versus 3.3 ± 2.0 ; $t_{176} = -1.5$, $p = 0.14$). However, checklist use took more time (63 ± 67 seconds versus 35 ± 49 seconds; $t_{162} = -3.2$, $p < 0.01$) and resulted in greater interpretation length (24 ± 60 versus 4 ± 12 extra characters per interpretation; $t_{95} = -3.1$, $p < 0.01$). Verification with checklists was associated with higher cognitive load efficiency (0.2 ± 1.2 versus -0.2 ± 1.0 ; $t_{170} = 2.5$, $p < 0.01$), time efficiency (0.3 ± 1.0 versus -0.3 ± 1.2 ; $t_{169} = 3.6$, $p < 0.01$) and length efficiency (0.3 ± 1.2 versus 0.3 ± 0.7 ; $t_{140} = 3.9$, $p < 0.01$) compared with verification without checklists (Fig. 2).

Comparison of all conditions

Checklist use was associated with fewer errors (net number: 70 versus 279; 0.39 ± 1.14 versus 1.04 ± 1.49 errors per ECG; $F_{1,441} = 11$, $p < 0.01$) (Table 2).

Table 1 Error detection in conditions with and without the use of a checklist

	Condition 2: ECGs verified without checklist	Condition 3: ECGs verified with checklist
ECG interpretations, <i>n</i>	89	89
Errors in undirected interpretation (condition 1), mean \pm SD	148, 1.66 \pm 1.52 per ECG	134, 1.51 \pm 1.51 per ECG
Errors in verification, mean \pm SD	145, 1.63 \pm 1.53 per ECG	108, 1.21 \pm 1.32 per ECG
Errors corrected per ECG, <i>n</i>		
+ 3	0	2
+ 2	4	7
+ 1	4	10
0	74	66
- 1	5	4
- 2	2	0
Net errors corrected, mean \pm SD	3, 0.03 \pm 0.61 per ECG	26, 0.29 \pm 0.77 per ECG

ECG = electrocardiogram; SD = standard deviation

Verification tasks were associated with fewer additional errors compared with interpretation tasks (net number: - 29 versus 378; - 0.16 \pm 0.71 versus 1.41 \pm 1.39 errors per ECG; $F_{1,441} = 171$, $p < 0.01$). There was no interaction between task type and checklist use ($F_{1,441} = 1.2$, $p = 0.28$).

Cognitive load did not differ between conditions with and without checklist use (4.0 \pm 1.8 versus 3.9 \pm 2.0; $F_{1,441} = 1.9$, $p = 0.17$). However, cognitive load was lower for verification tasks than for interpretation tasks (3.5 \pm 1.9 versus 4.2 \pm 1.9; $F_{1,441} = 17$, $p < 0.01$). There was no interaction between task type and checklist use ($F_{1,441} = 0.8$, $p = 0.38$).

Checklist use was associated with greater interpretation and verification times (94 \pm 84 seconds versus 83 \pm 74 seconds; $F_{1,441} = 10.4$, $p < 0.01$). Time spent on interpretation tasks was greater than time spent on verification tasks (113 \pm 79 seconds versus 50 \pm 60 seconds; $F_{1,441} = 87$, $p < 0.01$). Again, no interaction was found between task type and checklist use ($F_{1,441} = 0.7$, $p = 0.41$).

Checklist use was associated with longer interpretations (48 \pm 53 versus 46 \pm 46 characters; $F_{1,441} = 4.9$, $p = 0.03$). Far fewer characters were added in verification tasks compared with interpretation tasks (8 \pm 18 versus 70 \pm 47 characters; $F_{1,441} = 258$, $p < 0.01$). No interaction was found between task type and checklist use ($F_{1,441} = 0$, $p = 0.96$).

Efficiencies

Checklist use was associated with greater cognitive efficiency, time efficiency and length efficiency. Verification tasks were associated with greater cognitive load efficiency and a trend toward greater time and length efficiency. No interaction was found between task type and checklist use in any of these three models (Fig. 2).

Survey data

When asked about their routine practices outside this study, experts said they used a checklist-like approach for a mean \pm standard deviation (SD) of 67 \pm 27% of ECG interpretations. Experts estimated that they found errors in a mean \pm SD of 20 \pm 18% of their ECG interpretations. In addition, experts estimated that they usually spent a mean \pm SD of 26 \pm 19% of their time on verifying their decision.

DISCUSSION

To our knowledge, this is the first study of the use of cognitive checklists in the interpretation of ECGs. Among experts, checklists afforded a clear benefit. Experts corrected one error for every 3.4 ECGs when checklists were applied. In fact, experts were aware of this benefit. They told us that they routinely used a checklist-like approach and expected to find an error in one in five ECGs.

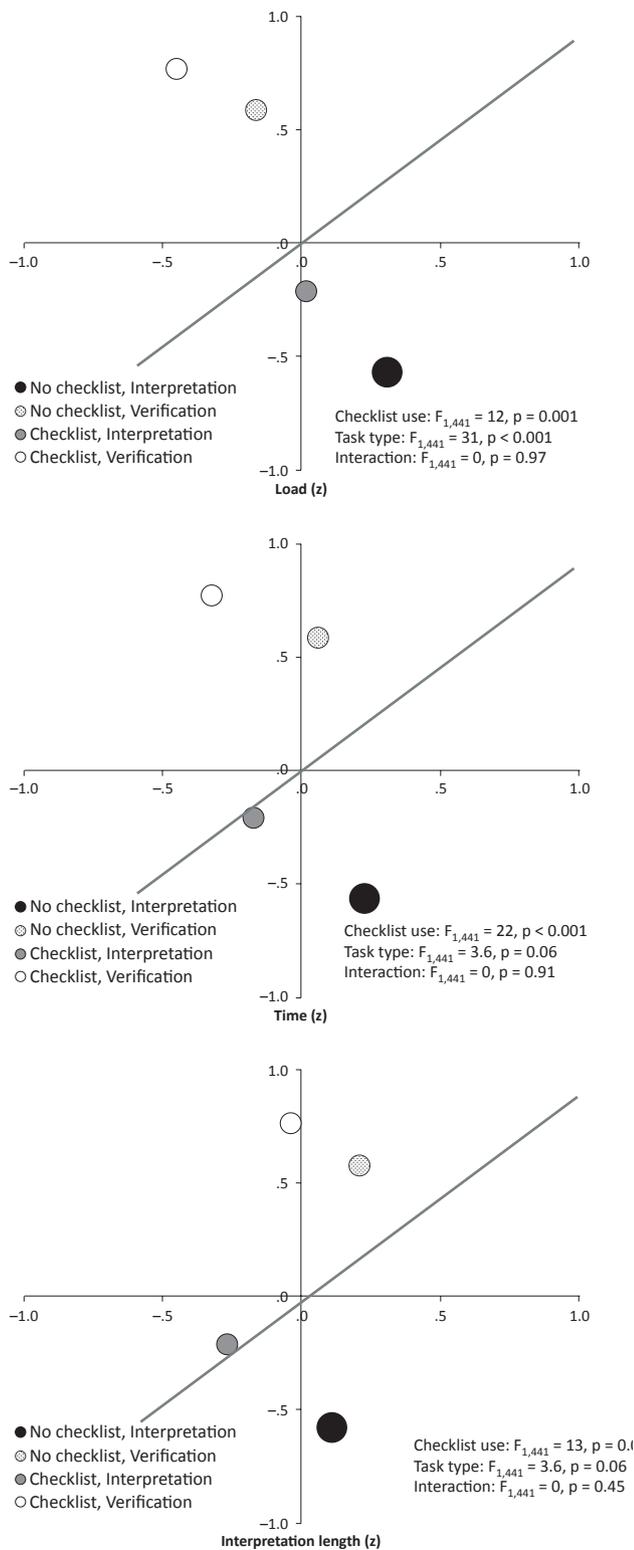


Figure 2 Checklist use and impact on (a) cognitive load, (b) interpretation time and (c) interpretation length efficiencies

Verification without a checklist was not associated with benefit, suggesting the effect requires more than just prolonging the verification stage of decision making. Potentiating system 2 processing is likely to be important. A previous study of ECG interpretation by intermediate-level trainees found that errors were detected only when system 2 processing was used, not when system 1 processing was encouraged.²² These findings could be extrapolated to experts: that is, experts are unlikely to detect an error unless they are verifying their interpretation using system 2 processing. In addition, the content of the checklist is likely to be important. The checklist might act as an ‘alternative’ illness script that experts can use in the verification stage. If an expert has not detected an error using his or her own illness script, it is unlikely that reapplying the same illness script in a verification phase will be helpful. However, a checklist offers an alternative approach because the variables are different, or the order of the variables is different or the list of variables is more comprehensive than that in the expert’s own illness script. Future development and study of checklists should be conducted with these hypothesised mechanisms in mind.

Surprisingly, checklist use did not increase cognitive load. The checklist we chose contained familiar variables. As a result, it was probably very easy for experts to adopt it. Furthermore, experts qualitatively valued the checklist approach, which suggests that we had sampled a group of willing participants who were familiar and maybe even expert with a checklist approach. As a result, checklist use in this context was not associated with expertise reversal. Whether this benefit translates to other diagnostic tests or to a context involving experts who place less value on systematic checking cannot be inferred from these results.

There were some disadvantages to the use of the checklist. Use of the checklist did increase verification time. On average, use of the checklist resulted in a 12% increase in verification time of approximately 10 seconds. Interestingly, this was less than estimated by the experts. However, our model does not account for the time taken by experts for systematic checking without prompting, and therefore is likely to represent an underestimate. Although the relative value of errors versus time is unlikely to be easily mathematically summarised, our use of time efficiency suggests that at the very least the time invested resulted in a disproportionate increase in error detection.

The use of a checklist also resulted in longer interpretations. However, although the increase in

Table 2 Interpretation errors in each condition

Conditions	Interpretations, <i>n</i>	Total number of errors*	Errors in initial undirected interpretation (condition 1) [†]	Net number of errors	Net errors per ECG, mean ± SD
1: undirected interpretation	178	282	NA	282	1.58 ± 1.51
2: verification without a checklist	89	145	148	- 3	- 0.03 ± 0.61
3: verification with a checklist	89	108	134	- 26	- 0.29 ± 0.77
4: combined interpretation and verification with a checklist	89	96	NA	96	1.08 ± 1.04
Checklist versus non-checklist conditions					
Checklist conditions (3 and 4)	178	204	134	70	0.39 ± 1.14
Non-checklist conditions (1 and 2)	267	427	148	279	1.04 ± 1.49
Verification versus interpretation tasks					
Verification tasks (conditions 2 and 3)	178	253	282	- 29	- 0.16 ± 0.71
Interpretation tasks (conditions 1 and 4)	267	378	NA	378	1.41 ± 1.39

* The total number of errors exceeds the number of interpretations as more than one error was made on most ECG interpretations

[†] Only applicable to conditions 2 and 3, in which ECGs were interpreted twice

ECG = electrocardiogram; SD = standard deviation; N/A = not applicable

interpretation length was measurable, it was relatively trivial, at two to 25 characters depending on the condition. Longer interpretations do have consequences in the health care system as they put a burden on consumers of the information. However, this small increment in interpretation length is unlikely to be clinically meaningful.

Limitations

First and foremost, expertise is content- and context-specific. Whether these findings apply in other contexts is unclear. Secondly, the benefit of a checklist is likely to depend on its content, its familiarity to experts and the way in which it is applied. Until the underlying principles of checklist efficacy are more firmly established, checklists should be trialled before use in each context. Thirdly, this study attempted to measure the trade-off between the advantages and disadvantages of using a checklist by adapting measures of efficiency from the cognitive load literature. However, this does not take into account the relative value of each of these measures and should not be applied too literally. How much time should be spent on detecting a life-threatening error on an ECG? We are not suggesting that such a question can be

answered by calculating efficiencies. Rather, we have included these measures in order to recognise potential disadvantages and include a relative gauge of effect sizes.

In summary, these results suggest there is substantial benefit to be derived by encouraging the use of checklists among experts in ECG interpretation. Interestingly, this study suggests there is still value to be gained by encouraging greater checklist use among experts who routinely use a systematic or checklist-based approach. Practically, this benefit should be shared not only with practising doctors, but also with the doctors in training who will become our future experts. Finally, participating experts told us that they used checklists in only two thirds of cases. The factors that determine whether an expert will use a checklist in any given case are unknown. Understanding the barriers against the usage of a checklist and content-specific triggers for its avoidance will be important in bringing this benefit to practice.

Contributors: all authors contributed to the conception and design of the study, and to the acquisition, analysis and interpretation of data. MS drafted the article. JJGvM and

ABHdB contributed to its critical revision. All authors approved the final manuscript for publication.

Acknowledgements: none.

Funding: MS received support from the Ho Ping Kong Center for Excellence in Education and Practice.

Conflicts of interest: none.

Ethical approval: this study was approved by the University of Toronto's independent ethics review board.

REFERENCES

- 1 van Merriënboer JJG, Sweller J. Cognitive load theory in health professional education: design principles and strategies. *Med Educ* 2010;**44** (1):85–93.
- 2 Sweller J, Ayres P, Kalyuga S. *Cognitive Load Theory*. New York, NY: Springer 2011;155–69.
- 3 Stanovich K. *Who is Rational? Studies of Individual Differences in Reasoning*. Mahwah, NJ: Lawrence Erlbaum Associates 1999;296.
- 4 Norman G. Dual processing and diagnostic errors. *Adv Health Sci Educ* 2009;**14**:37–49.
- 5 Evans J. Dual-processing accounts of reasoning, judgement, and social cognition. *Annu Rev Psychol* 2008;**59**:225–78.
- 6 Norman G, Young M, Brooks L. Non-analytical models of clinical reasoning: the role of experience. *Med Educ* 2007;**41** (12):1140–5.
- 7 Eva K, Cunningham J. The difficulty with experience: does practice increase susceptibility to premature closure? *J Contin Educ Health Prof* 2006;**26**:192–8.
- 8 Coderre S, Mandin H, Harasym PH, Fick GH. Diagnostic reasoning strategies and diagnostic success. *Med Educ* 2003;**37** (8):695–703.
- 9 Charlin B, Boshuizen HPA, Custers EJ, Feltovich PJ. Scripts and clinical reasoning. *Med Educ* 2007;**41** (12):1178–84.
- 10 Ross K, Shafer J, Klein G. Professional judgements and naturalistic decision making. In: Ericsson K, Charness N, Feltovich P, Hoffman R, eds. *The Cambridge Handbook of Expertise and Expert Performance*. Cambridge: Cambridge University Press 2006;403–19.
- 11 Norman G, Eva KW. Doggie diagnosis, diagnostic success and diagnostic reasoning strategies: an alternative view. *Med Educ* 2003;**37**:676–7.
- 12 Norman G, Eva K. Diagnostic error and clinical reasoning. *Med Educ* 2010;**44** (1):94–100.
- 13 Regehr G, Cline J, Norman G, Brooks L. Effect of processing strategy on diagnostic skill in dermatology. *Acad Med* 1994;**10** (Suppl):34–6.
- 14 Kahneman D. *Thinking, Fast and Slow*. Toronto, ON: Doubleday 2011;499.
- 15 Brenner L, Koehler D, Tversky A. On the evaluation of one-sided evidence. *J Behav Decis Mak* 1996;**9** (1):59–71.
- 16 Gilbert D, Krull D, Malone P. Unbelieving the unbelievable: some problems in the rejection of false information. *J Pers Soc Psychol* 1990;**59** (3):601–13.
- 17 Gawande A. *The Checklist Manifesto: How to Get Things Right*. New York, NY: Metropolitan Books 2009;209.
- 18 Kalyuga S. Expertise reversal effect and its implications for learner-tailored instruction. *Educ Psychol Rev* 2007;**19** (4):509–39.
- 19 Kalyuga S, Ayres P, Chandler P, Sweller J. The expertise reversal effect. *Educ Psychol* 2003;**38** (1):23–31.
- 20 Paas F, van Merriënboer J. The efficiency of instructional conditions: an approach to combine mental effort and performance measures. *Hum Factors* 1993;**35** (4):737–43.
- 21 Mattu A, Brady W. *ECGs for the Emergency Physician 2*. Malden: Blackwell Publishing 2008;119–63.
- 22 Sibbald M, de Bruin A. Feasibility of self-reflection as a tool to balance clinical reasoning strategies. *Adv Health Sci Educ Theory Pract* 2012;**17** (3):419–29.
- 23 Gertsch M, ed. *The ECG: A Two-Step Approach*. Berlin: Springer 2004.
- 24 Surawicz B, Knilans T, eds. *Chou's Electrocardiography in Clinical Practice*. Philadelphia, PA: Saunders Elsevier 2008.
- 25 Paas F. Training strategies for attaining transfer of problem-solving skill in statistics: a cognitive-load approach. *J Educ Psychol* 1992;**84** (4):429–34.
- 26 van Gog T, Paas F. Instructional efficiency: revisiting the original construct in educational research. *Educ Psychol* 2008;**43** (1):16–26.

Received 25 May 2012; editorial comments to authors 31 July 2012, 28 August 2012, 11 September 2012; accepted for publication 18 September 2012