# The Etiology of Diagnostic Errors: A Controlled Trial of System 1 Versus System 2 Reasoning

Geoffrey Norman, PhD, Jonathan Sherbino, MD, Kelly Dore, PhD, Tim Wood, PhD, Meredith Young, PhD, Wolfgang Gaissmaier, PhD, Sharyn Kreuger, and Sandra Monteiro, MSc

## Abstract

### Purpose
Diagnostic errors are thought to arise from cognitive biases associated with System 1 reasoning, which is rapid and unconscious. The primary hypothesis of this study was that the instruction to be slow and thorough will have no advantage in diagnostic accuracy over the instruction to proceed rapidly.

### Method
Participants were second-year residents who volunteered after they had taken the Medical Council of Canada (MCC) Qualifying Examination Part II. Participants were tested at three Canadian medical schools (McMaster, Ottawa, and McGill) in 2010 (n = 96) and 2011 (n = 108). The intervention consisted of 20 computer-based internal medicine cases, with instructions either (1) to be as quick as possible but not make mistakes (the Speed cohort, 2010), or (2) to be careful, thorough, and reflective (the Reflect cohort, 2011). The authors examined accuracy scores on the 20 cases, time taken to diagnose cases, and MCC examination performance.

### Results
Overall accuracy in the Speed condition was 44.5%, and in the Reflect condition was 45.0%; this was not significant. The Speed cohort took an average of 69 seconds per case versus 89 seconds for the Reflect cohort (P < .001). In both cohorts, cases diagnosed incorrectly took an average of 17 seconds longer than cases diagnosed correctly. Diagnostic accuracy was moderately correlated with performance on both written and problem-solving components of the MCC licensure examination and inversely correlated with time.

### Conclusions
The study demonstrates that simply encouraging slowing down and increasing attention to analytical thinking is insufficient to increase diagnostic accuracy.

---

*Editor's Note: A commentary on this article by P. Croskerry, D.A. Petrie, J.B. Reilly, and G. Tait appears on pages 197–200.*

---

Current "dual process" theories of clinical reasoning suggest that two distinct psychological processes are at work when doctors reach a diagnosis. *System 1* is a rapid, unconscious, intuitive, and primarily pattern recognition process involving retrieval of previous specific experiences from long-term memory. *System 2*, by contrast, is slow, conscious, effortful, logical, systematic, and based on explicit rules such as those that govern clinical diagnosis.[1,2] Evidence for these two separable processes is substantial and extends to (1) anatomical localization based on functional magnetic resonance imaging studies,[3,4] (2) changes in System 1 versus System 2 processing based on glucose availability,[5,6] and (3) studies of individuals with neurological damage.[7]

Although it is often thought that tasks can be identified as requiring primarily System 1 or 2 reasoning, based on task characteristics, experimental instructions, or the outcomes,[8] it is more likely that any real and complex task such as clinical diagnosis involves both kinds of processes.[7] Over 20 years ago, Jacoby[9(p513)] stated that

> problems interpreting task dissociations have arisen from equating particular processes with particular tasks and then treating those tasks as if they provide pure measures of those processes.

Jacoby presumed instead that the two processes act together and showed that it is possible to identify the contribution of System 1 (automatic) and System 2 (intentional) processes to a particular task through the use of a "process dissociation" framework in which, under one condition, an interference task is used to load working memory and hence impede System 2 thinking.

Building on these ideas, a more useful theory may be that the diagnostic task involves both System 1 and System 2 thinking. But in what proportions? Because System 1 amounts to rapid retrieval from memory and is not under conscious control, it occupies a relatively small part of the total time of the interaction[1] and cannot be influenced by external demands such as instructions by the experimenter. Conversely, System 2 uses varying degrees of analytical resources depending on a variety of factors, including:

1. The clinician may feel confident in the diagnosis derived from primarily System 1, and may do minimal analytic verification. Conversely, the clinician may be uneasy with the diagnosis and may spend more time exploring alternatives. There is evidence that diagnosticians are able to self-assess diagnostic accuracy to some degree, and that this increases with expertise.[10] To the extent that accurate diagnoses are generated by System 1 and are recognized as such, this would lead to an inverse relationship between time and accuracy. Such a relationship was observed in a study we conducted

in 2011[11] using a sample of written cases presented to second-year residents, where we found a correlation between time and accuracy of 0.54, indicating that slower diagnoses were less accurate. This held true even after controlling for participants' knowledge.

2.  Greater expertise is known to lead to increased reliance on System 1 and less on System 2.[12] Experts can be both faster and more accurate than relative novices. The former tend to ignore information and test fewer but more relevant attributes, which is both efficient and successful.[13–15] Furthermore, with increasing experience, expert clinicians can increasingly rely on patterns of a patient's symptoms, rather than treating the symptoms as largely unconnected, verbatim entities.[16] This allows experienced physicians to make better decisions using less information[17] and, thus, potentially less time.

3.  External conditions encouraging speed may lead to lower processing time and reduced use of analytical resources. Conversely, external conditions that lead to increased uncertainty[18] or that encourage a more systematic approach should lead to greater use of analytical resources and longer processing times.

Two critical questions then emerge: the extent to which the kinds of manipulations just listed actually increase or decrease reliance on analytical resources, and whether this, in turn, influences diagnostic accuracy. With respect to the first question, the finding of increased processing time is a crude but suggestive measure of participants' additional processing. More direct evidence derives from eye-tracking studies by Horstmann et al,[19] who found that (1) more complex decision-making tasks and (2) instructions to use deliberation increased the total time of processing, the number of fixations, and the amount of information processed, which is evidence that instructions and task complexity do result in more intensive and extensive processing. However, the distribution of long and short saccades was constant, leading the authors to conclude that both intuitive (System 1) and deliberative (System 2) processes were operative under all conditions.

With respect to the second question, the literature outside of medicine is equivocal, sometimes finding that abstract, System 2 processes are superior[14] and sometimes finding that intuitive processes are superior, particularly with experts.[20–22] However, within medicine, there is a prevalent view that errors originate in System 1 and are corrected by System 2. For example, Croskerry[23(p32)] writes:

> Most errors occur with Type I and may to some extent be expected, whereas Type 2 [System 2] errors are infrequent and unexpected.

Below we report a controlled trial, in which the performance of the 2010 cohort described earlier[11] was contrasted with that of a new cohort that we studied in 2011 that was encouraged to be slow, systematic, and thorough. The participants were drawn from the same population on two successive years (candidates sitting the Medical Council of Canada (MCC) Qualifying Examination [MCCQE] Part II), the same cases were used, and procedures were identical except for the instructions and the time allowed to complete the cases, which was longer in 2011.

Thus, we directly tested an experimental manipulation designed to encourage rapid versus slow processing and, in turn, to influence reliance on analytical knowledge. The primary hypothesis is that the instruction to be slow and thorough will have no advantage in diagnostic accuracy over the instruction to proceed rapidly. A second hypothesis, consistent with the findings of the previous study, is that, within groups, making accurate diagnoses will take less time than making diagnostic errors.

Finally, we hypothesize that there will be a positive relation between accuracy and the MCCQE Part I and Part II scores related to diagnostic problem solving, confirming the validity of the diagnostic cases as a measure of competence.

## Method

### Study procedures

The design was a prospective controlled study. Participants in the speeded diagnosis (Speed) group (n = 96) were recruited into the study in 2010; those in the reflective diagnosis (Reflect)

group (n = 108) were recruited in 2011. All participants were recruited from the same population of residents, who had a minimum of 1.5 years of clinical postgraduate training. Detailed analysis of demographic and MCCQE performance data (presented later) showed the groups were equivalent. They were asked to read through the study instructions on a computer screen and view a practice case, which was not evaluated but instead served to orient the candidates to the computer program.

The oral instructions for the Speed cohort emphasized that participants should be as quick but accurate as possible, and noted a red timer button that showed elapsed time on each case.

Written instructions on the first screen of the computer test were:

> You are to make your diagnosis and type it in as quickly but as accurately as possible. Case information will appear on one screen, and you click on a button to go to the diagnosis screen. You may spend as much time as you wish reading the case information, but remember that you only have 30 minutes to complete all the cases.

For the Reflect cohort, oral instructions emphasized thoroughness and care, and considering all the data. The timer was replaced by a case counter, to deemphasize speed within the case. Written instructions on the first screen were:

> You are to consider all the data and then make your diagnosis and type it in. Case information will appear on one screen, and you click on a button to go to the diagnosis screen. You may spend as much time as you wish reading the case information, so make sure you've considered all the data.*

For both cohorts, when a participant had completed reading the case, he or she then moved to a new screen and typed in the diagnosis.

Thus, in the Speed cohort there were several aspects of the design that were designed to encourage rapid diagnosis. Conversely, the Reflect cohort was encouraged to be thorough and to carefully review the data before arriving

---

*In fact, the time was restricted to 120 minutes, and as a result, some candidates did not complete all cases. See the Results section for further information.

at a conclusion, consistent with careful analytical thinking.

## Case development

As described by Sherbino et al,[11] cases were developed by four academic emergency medicine specialists supervised by one of us (J.S.). They were designed to represent a range of difficulty of acute general medicine cases from straightforward to rare, atypical, or complex presentations. The number of words per case (and reading times) were equivalent across cases. Cases were further reviewed by four academic internists and emergency physicians and were then piloted on a convenience sample of 10 residents, from the first and third years of internal medicine and emergency medicine programs.

A typical case is shown in Appendix 1. All cases contained some investigations that required interpretation, either visual (chest X ray, ECG) or laboratory, as in a typical case. Cases were designed to be fairly complex, requiring synthesis of data from history, physical, and lab as well as interpretation of some lab data. Although this may have deemphasized pattern recognition, the likelihood of significant errors provided a more sensitive test of the ameliorating effect of analytical reasoning.

Forty cases were developed initially, and then reduced to 25 for the 2010 study by eliminating cases that were too easy[†] or too difficult (primarily the former) based on pilot data. The 2010 Speed cohort was presented with all 25 cases in a 30-minute period (an average of 72 seconds per case). The 2011 Reflect cohort was presented with the first 20 cases in a 40-minute period (allowing 120 seconds per case). An extra 5 cases were included at the end of the test in 2010 because participants were encouraged to proceed quickly, so they could potentially complete more cases in the allotted time. Analysis was restricted to the first 20 cases, which were the same in both cohorts.

---

[†]Because the goal is to see the impact of reasoning strategy on diagnostic errors, one has to have errors to look at. If one uses simple cases, there is the risk that everyone gets them right all the time; thus, one can conclude nothing. Error rates of 50% actually optimize the chance of seeing an effect of instructions.

Cases were presented on computers using custom RunTime Revolution software (version 2.8.1; Edinburgh, Scotland). The software recorded time spent on each case and the diagnosis (entered as free text). Cases were presented in the same fixed sequence to both cohorts.

## Participants

All participants were recruited from candidates sitting the MCCQE Part II, a high-stakes clinical skills examination used for licensure in Canada. All candidates had a minimum of 1.5 years of clinical postgraduate training and had passed the MCCQE Part I, which is a computer-based test of basic medical knowledge and clinical decision making taken at the end of medical school. The examination is administered twice a year on the same day to candidates at all 17 medical schools in Canada.

Candidates in the Speed cohort were recruited in 2010; as mentioned earlier, their data were analyzed separately to look at the relation between speed and accuracy.[11] Candidates in the Reflect cohort were enrolled in 2011.

Candidates who sit the morning examination at examination centers within eastern Canada are sequestered at the examination site until candidates attempting the examination in other time zones have begun. For this study, candidates at three sites were approached: McGill University Faculty of Medicine, Montreal; McMaster University Faculty of Health Sciences, Hamilton; and the University of Ottawa Faculty of Medicine, Ottawa. Candidates were approached at the start of the sequestered period and informed of the purpose of the study, offered an honorarium of $30, and asked to complete a consent form if they wanted to participate. At each site, volunteers exceeded the number of computers; enrollment proceeded only until there were sufficient volunteers for the available computers. Once candidates were identified and consent forms signed, they then went immediately to a computer lab, where they began the timed, invigilated session.

Candidates for the MCCQE2 Part II examination can opt to sit the exam in French or English, and then select a site accordingly. All sites used in this study were for English-language candidates. Each year, approximately 64 candidates

are sequestered at McMaster and McGill, and 32 at Ottawa. A total of 96 candidates participated in the study in 2010 and 108 in 2011. The study was approved by the medical research ethics boards at each institution and by the MCC.

In contrast to the 2010 study, which analyzed only Canadian graduates, the current study used all volunteers, both Canadian medical graduates (CMGs) and international medical graduates (IMGs). The reason was to maintain a conservative bias. As pointed out in the original report, inclusion of IMGs in the correlational analysis could inflate correlations because IMGs have longer reading times and lower overall performances. Conversely, because the present analysis examined differences between cohorts, IMGs would likely increase differences between participants within each cohort and reduce power. In any case, the main findings were replicated with the CMG cohort and results were consistent.

## Scoring and analysis

Diagnostic accuracy was scored on a scale where 0 = incorrect, 1 = partially correct, and 2 = exactly correct. For example, for Case 3, the correct diagnosis was either *malignant otitis externa* or *necrotizing otitis externa*; a diagnosis of *otitis externa* was awarded one point. Diagnoses were originally scored and tallied by a research assistant. Any diagnoses suggested by participants that were not included in the original codes were scored by a content expert (J.S.). Because data were collected from the two cohorts a year apart, blinding for study condition was not possible. For each case and candidate, time to diagnosis (i.e., time from initial case presentation to when a free-text diagnostic response is entered) was collected and an average time per case was computed. A total accuracy score was calculated using the number of correctly diagnosed cases as the numerator, and the number of cases each candidate completed (as opposed to the total number of cases) as the denominator. In this manner, the accuracy of slower candidates would not be reduced by receiving scores of 0 for cases that were not attempted. Although the 2010 cohort had a case bank of 25 cases versus 20 cases for 2011, analyses were confined to the 20 cases common to both cohorts. Finally, after completing all cases, candidates estimated how many cases like

each they had seen during their medical education on a five-point nonlinear scale, where 0 = never seen, 1 = once, 5 = 5 times, 10 = 10 times, and 15 = more than 10 times.

We also had available total scores from the MCCQE Part I, as well as the total score and three subscores (data gathering, problem solving, and patient interaction, which includes communication skills) from the MCCQE Part II. The release of this information met all requirements used by MCC for sharing confidential examination results.

A database containing candidate identifiers was provided to the MCC, which then added in the MCCQE I and II data and deidentified the data. All subsequent analyses were conducted with an anonymized database.

The primary analysis was a comparison of diagnostic accuracy and time to diagnosis between the two groups. Mean scores and times were compared with $t$ tests. Additional analysis of the relation between time and accuracy, and relation to experience and MCCQE (Part I and II) scores, was conducted with regression, correlation, and analysis of variance subprograms, using PASW Statistics 20.0 for Macintosh (IBM SPSS, Armonk, New York).

## Results

The characteristics of the two samples are shown in Table 1. Both samples included about 20% IMGs. About one-third of both samples were residents in family medicine. There were no significant

differences between groups in location of graduation (Canadian/international), family medicine residency, or average performance on the MCCQE Part I or Part II. Because of the recruiting methods, where candidates volunteered until the quota was reached and no "head count" was taken of applicants in the room, it was not possible to estimate how many decided not to participate.

In the Speed cohort, 75 of 96 (78%) completed all 20 cases, and 84 completed more than 15 cases; in the Reflect cohort, 95 of 108 (88%) completed all cases, and 104 completed more than 15.

The critical comparison examined the overall accuracy and reading time of the two instructional groups. Mean accuracy of the Speed group was 44.5% (SD = 13) and, for the Reflect group, 45.0% (SD = 12). The difference was not significant ($t = -0.255$, $df = 212$, $P = .80$, effect size [ES] = .04). Conversely, there was a significant difference in time to diagnosis; 68.7 seconds (SD, 21.5) for the Speed group and 89.7 seconds (SD, 26.0) for the Reflect group ($t = -6.2$, $P < .0001$, ES = 0.87). Examining only those who completed all cases, the results were similar (time: 59 seconds for the Speed cohort versus 84 seconds for the Reflect cohort, $P < .0001$; accuracy, 46.3% versus 45.3%, $P = .85$). Secondary analysis using only the CMG participants resulted in similar conclusions (accuracy: 45.1% for the Speed cohort, 45.8% for the Reflect cohort, $P = .55$; time: 63.3 seconds for the Speed cohort, 84.5 seconds for the Reflect cohort, $P < .01$). Thus, instructions directed at encouraging systematic and thorough inquiry, although resulting

in significantly slower responding, did not improve accuracy of diagnostic reasoning.

Various models of reasoning[24–27] presume that experts will, when faced with uncertain or difficult cases, slow down and become more deliberative. To assess this, we analyzed the relation between time and accuracy in the two conditions. For each case, the times to diagnosis for participants with incorrect, partially correct, and correct answers in each condition were computed. The data were subjected to a mixed-model ANOVA, with Case as subject, Condition as the between-subject factor, and Correctness as the within-subject factor. With this strategy, average times per case are compared within case, so time is not confounded with case difficulty.

The results are shown in Figure 1. Consistent with the results of the prior study,[11] there was an inverse relationship between time and accuracy in both conditions (F = 21.4, $P < .001$); that is, incorrect diagnoses were associated with longer times. The instruction to proceed slowly added 20 to 25 seconds to time to diagnosis at all levels of accuracy (F = 6.4, $P = .02$). There was no interaction (F = 0.031, $P = .86$). Thus, regardless of whether the participant was encouraged to go slow or fast, correct diagnoses were associated with shorter reading times.

Although there was no overall advantage in accuracy of the Reflect condition, it is possible that this was masked by improved performance on difficult cases but reduced performance on easy cases, an effect observed in some prior studies.[26] To examine this, we reviewed the correlation between average accuracy in the two conditions case by case. The results are shown in Figure 2, with the line of identity superimposed. The correlation between the two mean scores/case in the two cohorts was 0.95 ($P < .001$); the slope of the best fit line was 0.98, and the intercept was 0.01. Case difficulty was nearly identical for the two cohorts.

We examined the correlation between accuracy and specific experience case by case across both cohorts. The average correlation was 0.17. Nineteen of 20 correlations were positive, and 12/20 were significant, suggesting a positive relation between specific case experience and accuracy.

## Table 1

**Characteristics of Study Participants, Three Canadian Medical Schools, 2010 and 2011\***

| Characteristic | Speed cohort (2010) | Reflect cohort (2011) | *P* value |
|---|---|---|---|
| Sample size: no. | 96 | 108 | NA |
| Canadian graduate: no. (%) | 76 (79) | 90 (84) | .44 |
| Family medicine: no. (%) | 34 (39) | 35 (36) | .46 |
| MCCQE Part I total score | 511.8 | 511.8 | .99 |
| MCCQE Part II total score | 511.5 | 510.5 | .93 |

\*All participants were recruited from candidates sitting the Medical Council of Canada Qualifying Exam (MCCQE) Part II, a high-stakes clinical skills examination used for licensure in Canada. All candidates had a minimum of 1.5 years of clinical postgraduate training and had passed the MCCQE Part I, a test of basic medical knowledge taken at the end of medical school. The table shows that the two cohorts are statistically equivalent in terms of background and licensing examination performance. The Speed cohort was instructed to be as quick as possible but not to make mistakes when determining diagnoses for 20 internal medicine cases; the Reflect cohort was instructed to be careful, thorough, and reflective when doing the same task for the 20 cases. NA indicates not applicable.
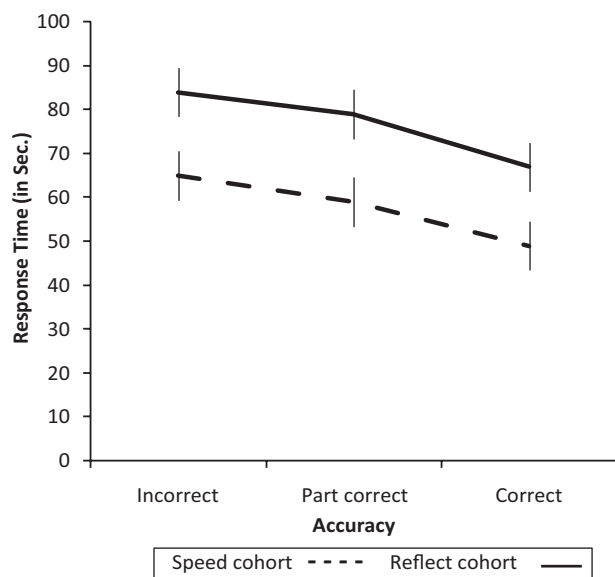
**Figure 1** Average response time per case (in seconds) for participants in the Speed (2010; n = 96) and Reflect (2011; n = 108) cohorts who were incorrect, partly correct, and correct. Error bars represent ± 1 standard error. There was an inverse relationship between time and accuracy in both conditions. The Speed cohort was instructed to be as quick as possible but not to make mistakes when determining diagnoses for 20 internal medicine cases; the Reflect cohort was instructed to be careful, thorough, and reflective when doing the same task for the 20 cases. Both cohorts were second-year residents in three Canadian medical schools.

Finally, we examined the relation between accuracy and MCCQE scores separately for the two cohorts, as shown in Table 2. Overall score was strongly positively correlated with MCCQE Part I scores (0.436 in the Speed cohort, 0.398 in the Reflect cohort) and with MCCQE Part II problem-solving scores (0.262 in the Speed cohort, 0.352 in the Reflect cohort), but not with patient interaction or data-gathering scores ($r < 0.15$). This provides some construct validation of the task, in that it is associated with problem-solving measures in both written and practical assessments. We then performed a multiple regression predicting the accuracy score with (1) knowledg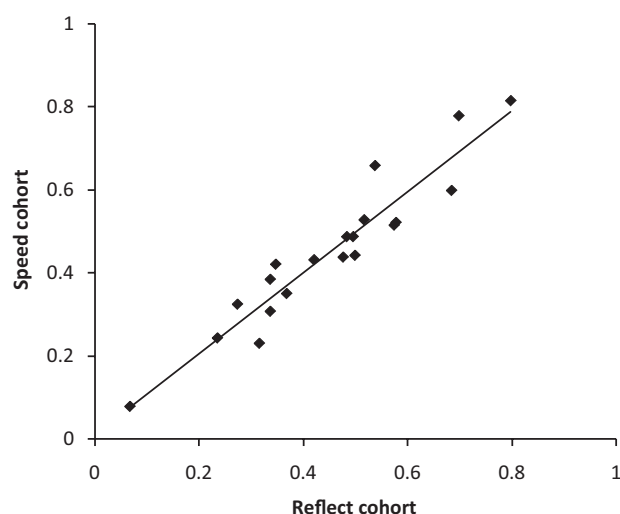e, assessed by the MCCQE Part I score, (2) overall response time (RT), (3) group (Speed versus Reflect), and (4) the group × RT interaction. Knowledge (beta = .41, $P < .001$) and RT (beta = −.71, $P < .001$) were independent predictors; group and the group × RT interaction were not significant, suggesting that rapid processing is a predictor of reasoning independent of formal knowledge.

## Discussion and Conclusions

In this controlled experiment, we demonstrated that residents who were encouraged to take more time to diagnose acute general medicine cases spent about 30% more time with each case than did those encouraged to proceed quickly through the same cases. However, this resulted in no improvement in diagnostic accuracy.

Furthermore, within cohorts, accurate diagnosis was associated with less, not more, reading time. Thus, there is no evidence from this study that rapid processing resulted in more diagnostic errors, or that instructions to be more careful, thorough, and to attend to all relevant data reduced diagnostic errors. A power calculation indicated that, with a sample size of 96 and 108 per group, we had sufficient power to detect a difference of 4.8% in accuracy.

The assumption that rapid diagnosis will lead to more errors is closely aligned with some dual processing models of reasoning, where it is assumed to result from System 1 thinking, which is vulnerable to cognitive biases, and hence error prone.[25,28] It then follows that instructions to slow down and be more thorough may encourage analytical thinking, which is assumed to compensate for System 1 errors. Our results are inconsistent with this view. An extra 20 seconds per case devoted to analysis resulted in no benefit in diagnostic accuracy. Moreover, within cases, inaccurate diagnosis was associated with more, rather than less, time.

This result contrasts with results of prior research[29,30] that have shown a benefit for a "reflective" intervention in increasing accuracy of diagnosis of difficult cases. However, in the previous studies, the initial experimental manipulation was deliberately designed to induce errors in System 1 thinking by showing the resident similar cases that had a different



**Figure 2** Correlation between case difficulty (proportion correct) for each case in the Speed (2010; n = 96) and Reflect (2011; n = 108) cohorts, showing cases were of equal difficulty for both cohorts. The Speed cohort was instructed to be as quick as possible but not to make mistakes when determining diagnoses for 20 internal medicine cases; the Reflect cohort was instructed to be careful, thorough, and reflective when doing the same task for the 20 cases. Both cohorts were second-year residents in three Canadian medical schools.

## Table 2

**Correlations Between Participants' Diagnostic Accuracy on 20 Computer-Based Cases and on Their Scores on the MCCQE Part I and Part II, Three Canadian Medical Schools, 2010 and 2011***

| MCCQE | Speed cohort (2010) | Reflect cohort (2011) |
|---|---|---|
| **Part I** | | |
| Multiple-choice | 0.468[†] | 0.394[†] |
| Clinical decision making | 0.151 | 0.269[†] |
| Overall score | 0.436[†] | 0.398[†] |
| **Part II** | | |
| Data gathering | 0.140 | 0 |
| Problem solving | 0.262[†] | 0.352[†] |
| Patient interaction | 0.146 | 0.108 |

*All participants were recruited from candidates sitting the Medical Council of Canada Qualifying Exam (MCCQE) Part II, a high-stakes clinical skills examination used for licensure in Canada. All candidates had a minimum of 1.5 years of clinical postgraduate training and had passed the MCCQE Part I, a test of basic medical knowledge taken at the end of medical school. The Speed cohort was instructed to be as quick as possible but not to make mistakes when determining diagnoses for 20 internal medicine cases; the Reflect cohort was instructed to be careful, thorough, and reflective when doing the same task for the 20 cases
[†]The diagnostic test was moderately correlated with the Part 1 multiple-choice and overall test score, but less with the clinical decision-making subtest, as might be expected from the reliability of the two MCC test components. For the Part 2 OSCE, the diagnostic test correlated significantly with the problem-solving score but not with the data-gathering or patient interaction score, supporting the construct validity of the diagnostic test.

diagnosis[29] or by giving the resident instructions designed to encourage superficial processing.[30] Thus, in some circumstances where a System 1 approach fails, there may be some advantage in encouraging reflection. However, it is not clear how such situations can be identified outside of a structured experimental setting.

Other studies have shown that residents and physicians, when instructed to do "reflective" reasoning, did better with complex cases but slightly worse with simple cases.[25,30] We found no evidence of this effect in the present study; average accuracy per case was nearly identical in the two conditions.

The current study has some limitations. A primary one is that the participants were beginning second-year residents, so they may not have been sufficiently expert. It would be interesting to conduct the same comparison with practicing physicians. However, it is unlikely that the results would favor reflective instructions, because practicing physicians have more experience than residents and rely more on System 1 thinking.[12]

The cases are representative of those in a busy emergency department, but were, on average, quite difficult for residents. If anything, one would suppose that these cases would have accentuated the advantage of reflective processing and would have biased the results to increasing System 1 errors in the Speed cohort. But our findings indicate otherwise.

Other limitations are methodological. The study did not involve true randomization to groups, and some cohort effects may be present. However, the two groups were shown to be equivalent on measures such as scores on the MCCQE Part I and II, and in fact, they were equivalent in performance, so it is difficult to see how this may alter interpretation. The study was conducted using written cases, and all the data on each case were available to the resident. Nevertheless, to achieve any measure that approximates reasoning time, one is forced into such contexts. We have shown that performance in this test of diagnostic skill was moderately related to licensing examination performance; other studies have shown that the written MCCQE Part I is, in turn, related to various aspects of practice performance.[31]

The results of the present study clearly show that diagnostic errors are not simply related to insufficient time for reflection or systematic analysis. By contrast, we showed that increased attention to analytical reasoning, through instructions to slow down and be more thorough, did not result in any increase in accuracy. Undoubtedly, some diagnostic errors are a product of human foibles—inattention, time pressures, distractions—and given more time and more careful attention, some could be prevented. But admonitions to always be slower, more reflective, and more systematic may not have the desired effect of reducing diagnostic errors.

Rather, the real challenge for physicians is to determine under which circumstances should clinicians trust their intuition, and under which circumstances should they slow down and deliberate?[32] And how can they tell one situation from the other? Future studies should therefore more carefully work out the relation between clinical situation, expertise, and diagnostic approach, and should ideally move beyond the vague distinction between System 1 and System 2 towards more precise models of diagnostic decision making.[33]

**Dr. Norman** is professor, Department of Clinical Epidemiology and Biostatistics, McMaster University, Ontario, Canada.

**Dr. Sherbino** is associate professor, Department of Medicine, McMaster University, Ontario, Canada.

**Dr. Dore** is assistant professor, Department of Medicine, McMaster University, Ontario, Canada.

**Dr. Wood** is associate professor, Department of Medicine, University of Ottawa, Ontario, Canada.

**Dr. Young** is assistant professor, Department of Medicine, McGill University, Quebec, Canada.

**Dr. Gaissmaier** is chief research scientist, Harding Center for Risk Literacy, Max Plank Institute for Human Development, Berlin, Germany.

**Ms. Kreuger** is research coordinator, Program for Educational Research and Development, McMaster University, Ontario, Canada.

**Ms. Monteiro** is a PhD candidate, Department of Psychology, Neuroscience and Behaviour, McMaster University, Ontario, Canada.

# References

1 Evans JS. Dual-processing accounts of reasoning, judgment, and social cognition. Annu Rev Psychol. 2008;59:255–278.

2 Stanovich KE, West RF. Individual differences in reasoning: Implications for the rationality debate. Behav Brain Sci. 2000;23:645–726.

3 Goel V, Buchel C, Frith C, Dolan RJ. Dissociation of mechanisms underlying syllogistic reasoning. Neuroimage. 2000;12:504–514.

4 Goel V, Dolan RJ. Explaining modulation of reasoning by belief. Cognition. 2003;87: B11–B22.

5 Masicampo EJ, Baumeister RF. Toward a physiology of dual-process reasoning and judgment. Psychol Sci. 2008;19:255–260.

6 Bos MW, Dijksterhuis A, van Baaren R. Food for thought? Trust your unconscious when energy is low. J Neurosci Psychol Econ. 2012;5:124–130.

7 Smith ER, deCoster J. Dual process models in social and cognitive psychology: Conceptual integrations and links to underlying memory systems. Pers Soc Psychol Rev. 2000;4:108–131.

8 Kahneman D. Thinking Fast and Slow. New York, NY: MacMillan; 2011.

9 Jacoby LL. A process dissociation framework: Separating automatic from intentional uses of memory. J Mem Lang. 1991;30:513–541.

10 Friedman CP, Gatti GG, Franz TM, et al. Do physicians know when their diagnoses are correct? Implications for decision support and error reduction. J Gen Intern Med. 2005;20:334–339.

11 Sherbino J, Dore KL, Wood TJ, et al. The relationship between response time and diagnostic accuracy. Acad Med. 2012;87: 785–791.

12 Eva KW, Link CL, Lutfey KE, McKinlay JB. Swapping horses midstream: Factors related to physicians' changing their minds about a diagnosis. Acad Med. 2010;85:1112–1117.

13 Shanteau J. How much information does an expert use? Is it relevant? Acta Psychol. 1992;81:75–86.

14 Ericsson KA, Prietula MJ, Cokely ET. The making of an expert. Harv Bus Rev. 2007;85:114–121.

15 Garcia-Retamero R, Dhami MK. Take-the-best in expert–novice decision strategies for residential burglary. Psychon Bull Rev. 2009;16:163–169.

16 Lloyd FJ, Reyna VF. Clinical gist and medical education: Connecting the dots. JAMA. 2009;302:1332–1333.

17 Reyna VF, Lloyd FJ. Physician decision making and cardiac risk: Effects of knowledge, risk perception, risk tolerance, and fuzzy processing. J Exp Psychol Appl. 2006;12:179–195.

18 Mamede S, Schmidt HG, Rikers RM, Penaforte JC, Coelho-Filho JM. Influence of perceived difficulty of cases on physicians' diagnostic reasoning. Acad Med. 2008;83:1210–1216.

19 Horstmann N, Ahlgrimm A, Glockner A. How distinct are intuition and deliberation? An eye-tracking analysis of instruction-induced decision modes. Judg Decis Mak. 2009;4:335–354.

20 Evans JS. In two minds: Dual-process accounts of reasoning. Trends Cogn Sci. 2003;7:454–459.

21 Klein GA, Calderwood R, Clinton-Cirocco A. Rapid decision making on the fire ground. In: Proceedings of the Human Factors and Ergonomics Society 30th Annual Meeting. Vol 1. Norwood, NJ: Ablex; 1986:576–580.

22 Ambady N. The perils of pondering: Intuition and thin slice judgments. Psychol Inq. 2010;21:271–278.

23 Croskerry P. Clinical cognition and diagnostic error: Applications of a dual process model of reasoning. Adv Health Sci Educ. 2009;14:27–35.

24 Redelmeier DA. Improving patient care. The cognitive psychology of missed diagnoses. Ann Intern Med. 2005;142:115–120.

25 Mamede S, Schmidt HG, Rikers RJM, Penaforte J, Coelho J. Breaking down automaticity: Case ambiguity and the shift to reflective approaches in clinical reasoning. Med Educ. 2007;41:1185–1192.

26 Mamede S, Schmidt HG, Rikers RM, Custers EJ, Splinter TA, van Saase JL. Conscious thought beats deliberation without attention in diagnostic decision-making: At least when you are an expert. Psychol Res. 2010;74:586–592.

27 Moulton CA, Regehr G, Lingard L, Merritt C, MacRae H. Slowing down to stay out of trouble in the operating room: Remaining attentive in automaticity. Acad Med. 2010;85:1571–1577.

28 Croskerry P. The importance of cognitive errors in diagnosis and strategies to minimize them. Acad Med. 2003;78:775–780.

29 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;304:1198–1203.

30 Ilgen JS, Bowen JL, Yarris LM, Fu R, Lowe RA, Eva K. Adjusting our lens: Can developmental differences in diagnostic reasoning be harnessed to improve health professional and trainee assessment? Acad Emerg Med. 2011;18(suppl 2):S79–S86.

31 Tamblyn R, Abrahamowicz M, Dauphinee D, et al. Physician scores on a national clinical skills examination as predictors of complaints to medical regulatory authorities. JAMA. 2007;298:993–1001.

32 Gigerenzer G, Gaissmaier W. Heuristic decision making. Annu Rev Psychol. 2011;62:451–482.

33 Marewski JN, Gaissmaier W, Gigerenzer G. We favor formal models of heuristics rather than lists of loose dichotomies: A reply to Evans and Over. Cogn Process. 2010;11:177–179.

## Appendix 1

**Sample Case: CO Poisoning\***

HPI: Mr. Patel is a 50-year-old male brought to the emergency department by ambulance after his landlord discovered him in his apartment. Ambulance personnel are concerned by his low oxygen saturation and respiratory distress as well as a decreased level of consciousness. Paramedics have established an IV. They are assisting ventilation with a bag-mask ventilation using 100% oxygen.

PMHx: According to family members, the patient has been depressed for the past year and has been taking medications for this. He also has a history of hypercholesterolemia, and hypertension.

Medication:

- Nitroglycerin 0.4 mg SL prn

- Hydrochlorothiazide 12.5 mg OD

- Ramipril 10 mg OD

- Rosuvastatin 10 mg OD

- Paroxetine 40 mg OD

Allergies: none.

Vitals: Temperature 36.4°C, heart rate 114 bpm, blood pressure 80/64; respiratory rate 30—assisted bag-mask ventilation; oxygen saturation on nonrebreather 86%.

Px: obtunded with some withdrawal to pain and moaning indiscriminately. The skin color is pink.

Cardiovascular exam: no murmurs or rubs, thready peripheral pulses.

Respiratory exam: bilateral air entry with coarse crepitations to the midlung fields.

Abdominal exam: scaphoid abdomen with normal bowel sounds, no obvious tenderness to palpation.

Head and neck exam: Pupils are 5 mm, equal and reactive to light, although slightly sluggish.

Ix: Blood is sent for a CBC, electrolytes, glucose, bicarbonate, BUN, creatinine, PT, PTT, CK, and troponin. A venous blood gas is also drawn. A portable chest x-ray is pending.

Hb = 128
WBC = 14.5
Plt = 265
Na = 149
K = 4.8
Cl = 108
Glc = 7.1
BUN = 6.5
Cr = 78

Venous blood gas:
pH = 7.52
$pO_2$ = 50
$pCO_2$ = 30
$HCO_3$ = 10

CarboxyHb: 29%

---

\*This is a typical case given to two cohorts of residents (n = 96 and 108) from three Canadian residencies who had a minimum of 1.5 years of clinical residency training and who had passed a test of basic knowledge at the end of medical school. One cohort was recruited in 2010; the other in 2011. For more information about the cohorts, see the report.