Reflection as a strategy to foster medical students' acquisition of diagnostic competence

Sílvia Mamede,[1] Tamara van Gog,[1] Alexandre S Moura,[2] Rosa M D de Faria,[2] José M Peixoto,[2] Remy M J P Rikers[1] & Henk G Schmidt[1]

OBJECTIVES Developing diagnostic competence in students is a major goal of medical education, but there is little empirical evidence on instructional strategies that foster the acquisition of this competence. The aim of this study was to investigate the effects of structured reflection compared with the generation of immediate or differential diagnosis while practising with clinical cases on learning clinical diagnosis.

METHODS This was a three-phase experimental study. During a learning phase, 46 Year 4 students diagnosed six clinical cases under different experimental conditions: structured reflection, immediate diagnosis, or differential diagnosis. This was followed by an immediate test and a delayed test administered 1 week later. Each test consisted of diagnosing four different cases of diseases presented in the learning phase. Performance in diagnosing these new cases was used as a measure of learning.

RESULTS Repeated-measures analysis of variance on the mean diagnostic accuracy scores (range: 0–1) showed a significant interaction between performance moment (i.e. performance in the learning phase and on each test) and instructions followed during the learning phase ($p = 0.003$). Follow-up analyses of this interaction showed that diagnostic performance did not differ between conditions in the learning phase. On the immediate test, scores in the reflection condition (mean = 0.48, 95% confidence interval [CI] 0.38–0.58) were significantly lower than scores in the differential diagnosis condition (mean = 0.62, 95% CI 0.54–0.70; $p = 0.012$) and marginally lower than those in the immediate diagnosis condition (mean = 0.61, 95% CI 0.52–0.70; $p = 0.04$). One week later, however, scores in the reflection condition (mean = 0.66, 95% CI 0.56–0.76) significantly outperformed those in the other conditions (differential diagnosis: mean = 0.48, 95% CI 0.37–0.58 [$p < 0.01$]; immediate diagnosis: mean = 0.52, 95% CI 0.43–0.60 [$p = 0.01$]). Comparisons within experimental conditions showed that performance from the immediate to the delayed test decreased in the immediate and differential diagnosis conditions (immediate diagnosis: $p = 0.042$; differential diagnosis: $p = 0.012$), but *increased* in the reflection condition ($p = 0.003$).

CONCLUSIONS Structured reflection while practising with cases appears to foster the learning of clinical knowledge more effectively than the generation of immediate or differential diagnoses and therefore seems to be an effective instructional approach to developing diagnostic competence in students.

Medical Education 2012: ***46****: 464–472*
doi:10.1111/j.1365-2923.2012.04217.x

Discuss ideas arising from this article at www.mededuc.com 'discuss'

[1]Department of Psychology, Erasmus University Rotterdam, Rotterdam, the Netherlands
[2]Department of Medical Education Development, Medical School, José do Rosário Vellano University (UNIFENAS), Belo Horizonte, Minas Gerais, Brazil

Correspondence: Sílvia Mamede MD, PhD, Department of Psychology, Erasmus University Rotterdam, Burgemeester Oudlaan 50, Rotterdam 3062 PA, the Netherlands. Tel: 00 31 10 408 9655; Fax: 00 31 10 408 9009; E-mail: mamede@fsw.eur.nl

INTRODUCTION

Optimal patient care is critically dependent on doctors' ability to arrive at correct diagnoses and therefore medical education places great emphasis on developing diagnostic competence in students. Little is known, however, about instructional strategies that would make the process of learning clinical diagnosis more effective; such strategies have been a source of much debate but little empirical evidence.[1,2] This article reports a study on the effects of using structured reflection as an instructional strategy on medical students' learning of clinical diagnosis.

Research has shown that students' diagnostic competence develops not only through knowledge expansion, but also through knowledge restructuring.[3,4] The way in which knowledge about diseases is organised in the mind changes throughout education. According to a well-accepted theory of medical expertise development, repeated application of knowledge acquired in the early stages of training and, especially, experiences with real patients are considered to lead to the formation of illness scripts.[3–6] These mental representations of diseases contain relatively little information about pathophysiological mechanisms, but rich clinical knowledge about the relationships among a particular set of symptoms, a particular disease, and the conditions under which that disease may emerge.[5,6]

Expert doctors have a rich collection of illness scripts available in memory, ranging from general representations of a disease category or prototypes, to representations of individual patients encountered previously.[6,7] Early in a clinical encounter, cues in a patient's history activate one or a few scripts in the doctor's mind and launch a subsequent search for additional information aimed at matching patient information to elements of the script in order to either confirm or exclude diagnostic hypotheses.[3,8] Accurate diagnostic performance critically depends, therefore, on the diversity and richness of the scripts of diseases that the doctor has in mind.[6,8]

To foster the acquisition of such knowledge, which is prerequisite for clinical diagnosis, students need to repeatedly encounter well-selected clinical cases in order to gain experience with many diseases and their different clinical presentations.[1,8] Whether and how students' learning from such cases might be enhanced, however, has remained largely unexplored. Teaching approaches for helping students to benefit more from their practice with clinical cases have been discussed, but little is known about their effectiveness.[1,9] Students are usually taught to think about alternative diagnostic hypotheses for the cases they encounter, rather than to rapidly provide a diagnosis they think is correct. It is not known, however, whether this is the most effective approach to foster the development of illness scripts and diagnostic competence.

Structured reflection on cases to be diagnosed might be a better approach to enhancing the learning of clinical presentations associated with a disease than simply teaching students to think of alternative diagnoses. Such structured reflection has been shown to be a very effective way of improving the diagnostic accuracy of doctors in complex problems[10,11] and of counteracting mistakes provoked by cognitive biases.[12] However, whether this might foster medical students' *learning* in order to enhance their diagnostic performance on similar cases in the future is an open question. This study addresses this question. We expect that structured reflection might foster learning, thereby improving diagnostic competence, operationalised here as performance in diagnosing written clinical cases. This positive effect of reflection would occur because it explicitly focuses students' attention on the findings in the case that speak for or against each of the alternative diagnoses, which may lead to the formation of deeper and richer mental representations of diseases, consequently enhancing students' competence in diagnosing other cases of the same disease.

We investigated this question in an experimental study with Year 4 medical students. In the learning phase, participants diagnosed clinical cases by following different procedures according to the condition to which they had been assigned. In subsequent tests, participants in all conditions were asked to diagnose different exemplars of the same diseases. We hypothesised that: (i) there would be no differences among the conditions in initial diagnostic performance during training, and (ii) on the tests, students who had reflected while diagnosing the cases during the training phase would outperform students in the other conditions.

METHODS

Design

The study consisted of three phases: a learning phase, an immediate diagnostic performance test, which took place immediately after the learning phase, and a delayed test administered 1 week later. In the learning phase, participants diagnosed six clinical cases under three different experimental conditions. Depending on their assigned condition, they did this by providing the first diagnosis that came to mind (i.e. an immediate diagnosis), or by providing an immediate diagnosis and then generating a differential diagnosis, or by providing an immediate diagnosis and then engaging in structured reflection. The tests required candidates to diagnose a set of six new clinical cases, of which four represented new exemplars of the diseases encountered in the learning phase and two represented fillers (cases of different diseases used to decrease the chance that participants would easily recognise the new set of cases as representing the diseases seen in the previous phases). Participants' previous clinical experience with the diseases was evaluated prior to the learning phase.

Setting and participants

A total of 46 Year 4 medical students (mean ± standard deviation [SD] age 22.9 ± 2.6 years; 29 women) at José do Rosário Vellano University (UNIFENAS) Medical School, in Belo Horizonte, Minas Gerais, Brazil, participated in this study. The school has a problem-based learning curriculum with a duration of 6 years, the last 2 years of which are dedicated to clerkships. We recruited Year 4 students because students at this point in their training have been exposed to theoretical knowledge about diseases but not yet to many patients. Therefore, they are not yet likely to have well-developed illness scripts. All 120 Year 4 students were invited to voluntarily participate. All those who volunteered were recruited. Written consent was obtained from 46 volunteers, a number of participants that was showed to provide sufficient power for the study.

Materials and procedure

Three sets of six different clinical cases were used in the study, one for each phase (Table 1). Each case consisted of a half-page description of a patient's medical history, present complaints, findings of a physical examination and results of laboratory tests. An example is presented in Table 2. Three experts in internal medicine (ASM, JMP, RMDdF) prepared the cases based on real patients. They subsequently revised each case until they agreed that the case findings supported a single best diagnosis, which was considered the correct diagnosis for the case.

Learning phase

The learning phase required the students to diagnose six clinical cases. Three cases concerned diseases in which chest pain is a prominent clinical manifestation, and three cases concerned diseases that present with jaundice and abnormal liver enzymes (Table 1). The cases were presented in a booklet in which one case per page was presented in a randomised order for each participant.

Participants were randomly assigned to one of the three conditions with varying instructions for the learning phase: the immediate diagnosis condition (n = 15); the differential diagnosis condition (n = 16), or the structured reflection condition (n = 15). In the immediate diagnosis condition, participants were requested to read the case and write down the most likely diagnosis for the case. Both accuracy and speed

Table 1 Diagnoses of the cases used in the different phases of the study

Training	Immediate diagnostic task	Delayed diagnostic task
Chest pain	Chest pain	Chest pain
Acute myocardial infarction	Acute myocardial infarction	Acute myocardial infarction
Acute viral pericarditis	Acute viral pericarditis	Acute viral pericarditis
Aortic dissection		
Jaundice/abnormal liver enzymes	Jaundice/abnormal liver enzymes	Jaundice/abnormal liver enzymes
Acute viral hepatitis	Acute viral hepatitis	Acute viral hepatitis
Choledocholithiasis	Choledocholithiasis	Choledocholithiasis
Haemolytic anaemia	Filler case	Filler case
	Filler case	Filler case

Table 2 Example of a case used in the study (correct diagnosis: acute viral pericarditis)

A 78-year-old man is brought to the emergency department with retrosternal pain that started 6 hours earlier and has become more intense. The pain becomes worse on inspiration, less intense when the patient sits down and is not alleviated with nitroglycerin – sublingual. The pain is accompanied by perspiration. There has been no dyspnoea. The patient felt in good health until 4 days ago, when he began to have a runny nose, non-productive cough, myalgias and anorexia. He does not smoke and consumes alcohol moderately. He does not have diabetes or hypertension

Physical examination

Blood pressure: 105/60; pulse: 105/minute; temperature: 38.5 °C

The rest of the physical examination did not show further abnormalities

Laboratory tests (reference ranges in brackets)

Haemoglobin: 14.0 g/dL^{-1} (14.0–17.5 g/dL^{-1}); white blood cell count: 10 000 μL (4500–11 000 μL) with normal differential count; platelet count: 190 000 μL (150 000–350 000 μL); glucose: 110 mg/dL^{-1} (70–110 mg/dL^{-1}); urea: 35 mg/dL^{-1} (10–50 mg/dL^{-1}); creatinine: 0.8 mg/dL^{-1} (0.60–1.30 mg/dL^{-1}); CK: 195 U/L (38–174 U/L); CK-MB: 6% (4–6%); CK-MB 6 hours later: 5% (4–6%)

Electrocardiogram shows sinus tachycardia, ST elevations in all leads

Imaging tests

Chest X-ray: no abnormalities

were stressed, and participants were instructed to move on to the next page to solve a word puzzle after they had made a diagnosis. This procedure was used to minimise the likelihood that participants would engage in reflection in the other experimental conditions.[10,11] In the differential diagnosis condition, participants were requested to read the case and write down the most likely diagnosis for the case. After that, they were asked to think about alternative diagnoses if their initial hypothesis were to prove incorrect, write them down and, then, draw a conclusion about the final diagnosis. After finishing this task, they were also instructed to move on and solve a word puzzle. In the reflection condition, participants were also requested to read the case and write down the most likely diagnosis for the case. They were then asked to follow a structured procedure to reflect upon the case. They were requested to: (i) list the findings in the case description that support this diagnosis; (ii) list the findings that speak against this diagnosis, and (iii) list the findings expected to be present if the diagnosis were correct, but which were not described in the case. They were subsequently asked to list alternative diagnoses, assuming that their initial diagnosis would prove to be incorrect, and to follow the same procedure (steps 1–3) for each alternative diagnosis. Finally, they were asked to draw a conclusion by ranking the diagnoses in order of likelihood and selecting the most accurate diagnosis for the case. Note that whereas differential diagnosis will involve some degree of reflection, the procedure for structured reflection particularly focused students' attention on identifying contradictory and absent findings in the case. No feedback was provided in any of the conditions about the accuracy of the diagnoses.

The maximum amount of time allowed for each case was 7 minutes in all conditions. A pilot study with Year 4 students from a different medical school in the same city had shown this to be a sufficient amount of time. If participants finished sooner, they were able to work on the puzzle until their time was up. The researcher in attendance monitored the time and indicated when participants could move on to the next case.

Diagnostic performance tests

The two tests (immediate and delayed) each used new sets of six clinical cases, of which two involved chest pain, two featured jaundice and abnormal liver enzymes, and two were filler cases (Table 1). All cases differed from the cases participants had seen in the learning phase. Participants received the cases in a booklet, one per page, in randomised order. They were requested to read the case and write down the most likely diagnosis. Feedback about the correctness of their responses was provided to participants only after the study was completed.

Assessment of previous clinical experience

One week before the learning phase, participants assessed their level of clinical experience with each of the diseases they would encounter in this study (embedded in a longer list) using a 5-point scale in which 1 = I have never seen a clinical case of this disease and 5 = I have seen several clinical cases of this disease (indicating extensive experience).

Data analysis

The mean clinical experience ratings for the diseases included in this study were computed. Analysis of variance (ANOVA) was performed on these data to check that there were no *a priori* differences in (self-reported) experience among students in the three experimental conditions.

The initial diagnoses provided by the participants in the learning phase and on the diagnostic performance tests were evaluated as correct, partially correct or incorrect, receiving scores of 1, 0.5 or 0, respectively. The diagnosis was considered correct whenever the core correct diagnosis of the case was provided (e.g. 'pericarditis' in a case of acute viral pericarditis). When the core diagnosis was not given, but one component of the diagnosis was mentioned, the diagnosis was considered partially correct (e.g. 'myocardial ischaemia' in a case of acute myocardial infarction). When the participant's response did not fall into one of these categories, the diagnosis was considered incorrect. Two experts in internal medicine (ASM, JMP) independently evaluated participants' responses for each case. Responses had been previously transcribed from the booklets to sheets of paper so that evaluators were not aware of the experimental condition under which the diagnoses had been provided. Their evaluations corresponded for 92% of the diagnoses; discrepancies were resolved by discussion.

For each participant, the scores obtained on the four cases of diseases that reappeared in each phase (i.e. cases of diseases studied in the learning phase that were used in the diagnostic tests [Table 1]) were summed. Mean diagnostic scores on those cases were then computed.

Repeated-measures ANOVA (significance level: 0.05) with experimental condition (immediate diagnosis, differential diagnosis, reflection) as a between-subjects factor and performance moment (learning phase, immediate test, delayed test) as a within-subjects factor was conducted on the mean diagnostic performance scores obtained in the four cases of interest in each phase. This analysis tested the hypothesis that reflection while solving clinical cases would foster learning and would lead to better diagnostic performance on the tests than that attained by the other conditions. A significant interaction effect in this repeated-measures ANOVA was further analysed by looking at a depiction of the interaction and then conducting specific independent-samples t-tests (with Bonferroni correction) to test whether observed differences in performance among conditions at each performance moment were significant, and paired t-tests (with Bonferroni correction) to test whether observed differences in performance within conditions over time were significant.

RESULTS

Previous clinical experience with the diseases did not differ among conditions (immediate diagnosis: mean ± SD 2.06 ± 0.63; differential diagnosis: mean ± SD 2.02 ± 0.54; reflection: mean ± SD 2.24 ± 0.72; $F_{(2,43)} = 0.49$, $p = 0.62$).

Table 3 shows the means and SDs of diagnostic performance in the three phases for each condition. The repeated-measures ANOVA showed no significant main effect of performance moment ($F_{(2,42)} < 1$, $p = 0.781$). However, there was a significant interaction between performance moment and experimental condition ($F_{(2,86)} = 4.25$, $p = 0.003$; $\eta_p^2 = 0.17$), depicted in Fig. 1. We tested those differences between conditions within each phase, as well as differences within conditions across phases, suggested to be present in Fig. 1 for significance, using, respectively, independent-samples and paired one-tailed t-tests with Bonferroni correction for multiple tests at each phase (with two tests per phase for the independent-samples t-tests, the significance level was 0.025; with three tests for the paired t-tests, the significance level was 0.017).

Figure 1 suggests that students in the structured reflection condition performed worse than those in the immediate diagnosis and differential diagnosis conditions during the training phase and on the immediate test. This difference was not significant during the training phase (both comparisons: $p > 0.27$), but it was significant on the immediate test for reflection versus differential diagnosis ($t_{(29)} = 2.36$, $p = 0.012$, $d = 0.82$) and reflection versus immediate diagnosis ($t_{(28)} = 1.98$, $p = 0.028$, $d = 0.76$) (only borderline given the Bonferroni correction). However, Fig. 1 also suggests that students in the structured reflection condition outperformed those in the other conditions on the delayed test (1 week later), and this difference was significant

Table 3 *Mean diagnostic performance score (maximum score: 1.0) as a function of performance moment and approach to practice clinical problem solving during the learning phase and comparisons between the structured reflection and other conditions*

	Immediate diagnosis group (*n* = 15)		**Differential diagnosis group (*n* = 16)**		**Structured reflection group (*n* = 15)**		**Reflection versus immediate diagnosis**		**Reflection versus differential diagnosis**	
	Mean score	**95% CI**	**Mean score**	**95% CI**	**Mean score**	**95% CI**	**p-values**	**Cohen's *d***	**p-values**	**Cohen's *d***
Learning phase	0.59	(0.44–0.74)	0.58	(0.47–0.70)	0.52	(0.35–0.70)	0.270	0.24	0.271	0.22
Immediate test	0.61	(0.52–0.70)	0.62	(0.54–0.70)	0.48	(0.38–0.58)	0.028	0.76	0.012	0.82
Delayed test	0.52	(0.43–0.60)	0.48	(0.37–0.58)	0.66	(0.56–0.76)	0.013	0.87	0.005	1.0

95% CI = 95% confidence interval

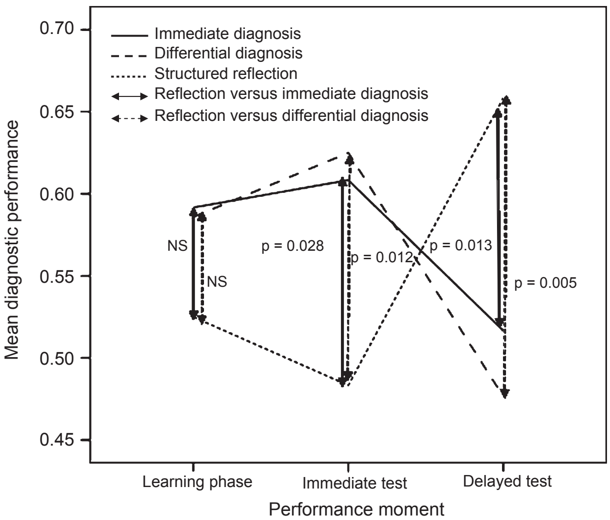

Figure 1 Interactions in mean diagnostic performance scores among conditions across phases. NS = not significant

for reflection versus differential diagnosis ($t_{(29)} = 2.74$, p = 0.005, $d = 1.0$) and reflection versus immediate diagnosis ($t_{(28)} = 2.35$, p = 0.013, $d = 0.87$). These comparisons are also displayed in Table 3, along with the magnitude of effects, which proved to be large, according to Cohen's guidelines,[13] especially for the delayed test.

Paired *t*-tests showed that performance did not differ significantly between the learning phase and the immediate test in the immediate diagnosis condition ($t_{(14)} < 1$, p = 0.411), the differential diagnosis condition ($t_{(15)} < 1$, p = 0.279) or the reflection condition ($t_{(14)} < 1$, p = 0.304). By contrast, between the immediate and the delayed tests, performance seemed to decrease (Fig. 1) in the immediate diagnosis condition ($t_{(14)} = 1.85$, p = 0.042) albeit not significantly (given the Bonferroni correction), decreased significantly in the differential diagnosis condition ($t_{(15} = 2.54$, p = 0.012), and *increased* significantly in the reflection condition ($t_{(14)} = 3.15$, p = 0.003). Performance from the learning phase to the delayed test improved in the reflection condition ($t_{(14)} = 2.17$, p = 0.024) albeit with borderline significance (given the Bonferroni correction), but did not differ significantly in the immediate diagnosis condition ($t_{(14)} < 1$, p = 0.190) or the differential diagnosis condition ($t_{(15)} = 1.70$, p = 0.055).

DISCUSSION

The findings demonstrate that structured reflection while practising the diagnosis of clinical cases fosters the learning of clinical presentations associated with those diseases more effectively than making immediate diagnoses or generating differential diagnoses, when measured after a delay. After 1 week, students who had been instructed to reflect during the learning phase outperformed students in the other conditions on new exemplars of the same diseases (i.e. diseases they had seen in the learning phase). The reflection condition was the only condition to show that learning had occurred when comparing performance in the learning phase with performance on the delayed test after 1 week. On the test that immediately followed the learning phase, however, students in the reflection condition performed worse than students who generated an immediate single diagnosis or a differential diagnosis during the learning phase. It should be noted, however, that their performance on the immediate test was no worse than their performance in the learning phase, which was also slightly lower, although not significantly so, than the performance of students in the other two conditions (Fig. 1).

Developing students' diagnostic competence is of primary importance in medical education, and how clinical teaching should be organised to better achieve this goal has been a subject of much discussion.[1,2] This discussion has explored both guiding principles for clinical teaching[14,15] and instructional approaches to help students benefit more from their practice with clinical cases.[1,9] Despite being theoretically sound, these proposals were often not accompanied by an empirical investigation of their effects on learning. There have been reports of empirical studies comparing different approaches adopted by psychology students[16,17] or Year 1 medical students[18] for solving clinical cases. These studies, however, investigated the effects on diagnostic performance on the cases at-hand, but not on *learning* by measuring performance on novel cases in the future. Taken together, our findings suggest that students can learn more from practising with clinical cases when they engage in structured reflection to diagnose these cases than when they simply give a diagnosis or even when they generate a differential diagnosis, as they are usually recommended to do during clinical practice. Why would reflection have this beneficial effect on learning to diagnose clinical cases?

Our explanation is that reflection helps by fostering the restructuring of existing knowledge about the diseases encountered during practice with clinical cases. Developing the ability to make accurate diagnoses depends on acquiring a large knowledge base of mental representations of diseases,[7,8] particularly well-developed illness scripts.[3–5] Illness scripts contain information on the set of findings expected in patients with a certain disease, as well as acceptable variations of the typical presentations of the disease. The set of signs and symptoms belonging to a certain script is usually not exclusive; some elements of a script may also be present in the scripts for other diseases. The differential diagnosis of a patient's problem, then, requires activating illness scripts relevant to the clinical presentation and verifying the extent to which their elements match findings in the patient history.[5,15] Therefore, the richer and better developed illness scripts of diseases related to a certain clinical presentation are, the more they enhance a practitioner's competence in differential diagnosis.

In this study, when students engaged in reflection while diagnosing a clinical case, they were required to think about how each finding in the case related to a diagnostic hypothesis, and to proceed with this process of matching findings and scripts for all alternative diagnoses. This process of elaboration on the case information and search for meaning helps to activate additional relevant knowledge and fosters the integration and reorganisation of pre-existing knowledge. Thus, even if no new information was provided to the students in this study and no *new* knowledge could be acquired, reflection enriched the illness scripts they had already stored, which, in turn, facilitated the diagnostic process when they encountered different exemplars of the same diseases at a later stage in the study. It is noteworthy that students in the differential diagnosis condition did not learn as much as those in the reflection condition. Generating alternative diagnoses is likely to have led the students in the differential diagnosis group to also engage in reflection to some degree. However, students in the structured reflection condition went through the step-by-step process of matching findings from the case to each diagnosis and identifying contradictory and absent findings. The latter does not necessarily happen during the generation of differential diagnoses, but it seems to be a crucial aspect of the reflection procedure for the enrichment of illness scripts.

This finding that the effects of reflection on learning only become apparent after a delay is in line with those of other studies demonstrating that the effects of elaboration or deeper processing of information on learning emerge only after some delay.[19,20] Interestingly, in the short-term, structured reflection seems to lead to poorer performance than the other conditions. However, it should be noted that none of the conditions resulted in significantly higher or lower performance on the immediate test than in the learning phase. Therefore, although students in the reflection condition performed worse than those in the other conditions, students in the other conditions showed no significant signs of learning on the immediate test. A potential reason why the beneficial effects of reflection show only after some delay may be that reflection imposes a high cognitive load during the learning phase and leads to exhaustion, or that reflection leads to initial confusion given that many different features and diagnoses have been considered in detail.[21] However, as a result of elaborate processing, the quality of illness scripts is improved, which leads to better longer-term performance. As an analogy, consider for example the robust effect shown by research on contextual interference: a random practice schedule requires much effort and initially slows down performance improvement compared with a blocked practice schedule, but proves to be more effective for learning and transfer in the long run.[21,22] However, as we did not take cognitive load measurements and did not ask learners for ratings of certainty or confusion, this explanation remains speculative and should be tested in future research.

The study took place in a single school with a relatively small number of students in the same year of training, which may restrict its generalisability. The 1-week interval between training and test is a regular methodological choice in studies on learning[19,20,23] (see also research on the testing effect[24,25]) and longer delays increase the chance that uncontrolled interference by other factors might influence students' learning. However, based on our data, we cannot exclude the possibility that the positive effects of reflection on learning will diminish after longer periods of time; this is a question to be addressed in future research.

To the best of our knowledge, the present study represents the first to investigate the effects of structured reflection as an instructional strategy to enhance students' acquisition of diagnostic competence. What is especially interesting about our findings for medical educators is that this procedure of structured reflection is relatively easy to apply while having students practise with clinical cases, either in classroom activities or in clinical settings. Future research should also investigate whether students can be taught to use the procedure themselves and use it to diagnose different problems in the future.

Contributors: SM had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of data analysis. SM and TvG contributed to the study conception and design, the acquisition, analysis and interpretation of data, statistical analysis, and the drafting of the manuscript. ASM contributed to the acquisition, analysis and interpretation of data, and to statistical analysis. RMDdF contributed to the acquisition, analysis and interpretation of data and provided administrative, technical or material support. JMP contributed to the acquisition, analysis and interpretation of data. RMJPR contributed to the analysis and interpretation of data and provided administrative, technical or material support, and supervised the study. HGS contributed to the study conception and design, and to the analysis and interpretation of data, provided administrative, technical or material support and supervised the study. All authors contributed to the critical revision of the paper and approved the final manuscript for publication.

Acknowledgements: the authors are grateful to the students who dedicated their time to participating in the study. During the realisation of this work, TvG was supported by a Veni grant from the Netherlands Organisation for Scientific Research (NWO; 451-08-003).
Funding: none.
Conflicts of interest: none.
Ethical approval: this study was approved by the Research Ethics Committee of José do Rosário Vellano University (UNIFENAS) Medical School (Belo Horizonte, MG, Brazil).

REFERENCES

1 Kassirer JP. Teaching clinical reasoning: case-based and coached. *Acad Med* 2010;**85**:1118–24.
2 Reilly BM. Inconvenient truths about effective clinical teaching. *Lancet* 2007;**370**:705–11.
3 Schmidt HG, Boshuizen HPA. On acquiring expertise in medicine. *Educ Psychol Rev* 1993;**5**:1–17.
4 Schmidt HG, Norman GR, Boshuizen HP. A cognitive perspective on medical expertise: theory and implication. *Acad Med* 1990;**65**:611–21.
5 Charlin B, Boshuizen HPA, Custers EJ, Feltovich PJ. Scripts and clinical reasoning. *Med Educ* 2007;**41**: 1178–84.
6 Schmidt HG, Rikers RMJP. How expertise develops in medicine: knowledge encapsulation and illness script formation. *Med Educ* 2007;**41**:1133–9.
7 Norman GR. Research in clinical reasoning: past history and current trends. *Med Educ* 2005;**39**:418–27.
8 Eva KW. What every teacher needs to know about clinical reasoning. *Med Educ* 2005;**39**:98–106.
9 Parsell G, Bligh J. Recent perspectives on clinical teaching. *Med Educ* 2001;**35**:409–14.
10 Mamede S, Schmidt HG, Penaforte JC. Effects of reflective practice on the accuracy of medical diagnoses. *Med Educ* 2008;**42**:468–75.
11 Mamede S, Schmidt HG, Rikers RMJP, Custers EJFM, Splinter TAW, van Saase JLCM. Conscious thought beats deliberation without attention in diagnostic decision making: at least when you are an expert. *Psychol Res* 2010;**74**:586–92.
12 Mamede S, van Gog T, van den Berge K, Rikers RMJP, van Saase JLCM, van Guldener C, Schmidt HG. Effect of availability bias and reflective reasoning on diagnostic accuracy among internal medicine residents. *JAMA* 2010;**304** (11):1198–203.
13 Cohen J. A power primer. *Psychol Bull* 1992;**112** (1):155–9.
14 Bowen JL. Educational strategies to promote clinical diagnostic reasoning. *N Engl J Med* 2006;**355**:2217–25.
15 Custers EJFM, Stuyt PMJ, De Vries Robbé PF. Clinical problem analysis (CPA): a systematic approach to teaching complex medical problem solving. *Acad Med* 2000;**75**:291–7.

16 Ark TK, Brooks LR, Eva KW. Giving learners the best of both worlds: do clinical teachers need to guard against teaching pattern recognition to novices? *Acad Med* 2006;**81**:405–9.
17 Eva KW, Hatala RM, Leblanc VR, Brooks LR. Teaching from the clinical reasoning literature: combined reasoning strategies help novice diagnosticians overcome misleading information. *Med Educ* 2007;**41**:1152–8.
18 Coderre S, Bruce Wright B, McLaughlin K. To think is good: querying an initial hypothesis reduces diagnostic error in medical students. *Acad Med* 2010;**85**:1125–9.
19 Woods NN, Brooks LR, Norman GR. It all makes sense: biomedical knowledge, causal connections and memory in the novice diagnostician. *Adv Health Sci Educ Theory Pract* 2007;**12**:405–15.
20 Woods NN, Brooks LR, Norman GR. The role of biomedical knowledge in diagnosis of difficult clinical cases. *Adv Health Sci Educ Theory Pract* 2007;**12** (4): 417–26.
21 van Merrienboer JJ, Sweller J. Cognitive load theory in health professional education: design principles and strategies. *Med Educ* 2010;**44**:85–93.
22 Helsdingen A, van Gog T, van Merriënboer JJG. The effects of practice schedule on learning a complex judgement task. *Learn Instr* 2011;**21**:126–36.
23 Schmidt RA, Bjork RA. New conceptualisations of practice: common principles in three paradigms suggest new concepts for training. *Psychol Sci* 1992;**3**:207–17.
24 Butler AC. Repeated testing produces superior transfer of learning relative to repeated studying. *J Exp Psychol Learn Mem Cogn* 2010;**36**:1118–33.
25 Roediger HL, Karpicke JD. The power of testing memory: basic research and implications for educational practice. *Perspect Psychol Sci* 2006;**1**:181–210.

Received 11 July 2011; editorial comments to authors 22 September 2011, 17 November 2011; accepted for publication 8 December 2011

