DELIBERATE REFLECTION UPON CLINICAL CASES AS AN INSTRUCTIONAL APPROACH TO FOSTER LEARNING OF SCIENTIFIC TEXTS

Ligia Maria Cayres Ribeiro

ISBN 978-94-6361-769-7

Deliberate Reflection Upon Clinical Cases as an Instructional Approach
to Foster Learning of Scientific Texts

Bewust redeneren over klinische casus als methode om het leren van wetenschappelijke teksten te bevorderen

Thesis

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

Prof.dr. A.L. Bredenoord

and in accordance with the decision of the Doctorate Board.
The public defence shall be held on

Wednesday 25 of January 2023 at 15:30 hrs
by

Ligia Maria Cayres Ribeiro
born in Belo Horizonte - Minas Gerais, Brazil.

Erasmus University Rotterdam

Doctoral Committee:

Promotors:	Prof.dr. H.G. Schmidt
	Dr. S. Mamede Studart Soares
Other members:	Prof.dr. T.A.J.M. van Gog
	Prof.dr. G.W.C. Paas
	Prof.dr. W.W. van den Broek

TABLE OF CONTENTS

Chapter 1

General Introduction

Reflection has been considered an essential characteristic of competent physicians and consequently a determinant of the quality of care provided to patients.[1] Reflective practice has even become a requirement for licensing by some medical boards. Reflection has also been much valued in medical education. Instructional approaches that are believed to support learning and professional development through reflection have been incorporated from undergraduate to postgraduate medical curricula.[1-3] Despite the great attention reflection has gained within medical education in the past years, there is a gap between what is expected from reflection and the evidence that support these expectations. The limited empirical research in the field does not provide sufficient understanding of which instructional approaches indeed foster reflection, in which contexts, what outcomes can be expected from them and what mechanisms underly these outcomes. Much remains to be known about how to make the best use of reflection to foster medical students' learning. This thesis was dedicated to examining whether deliberate reflection upon clinical cases could facilitate medical students' subsequent learning of scientific texts that are relevant to these cases. It also explored two possibly complementary mechanisms that could underly the potentially positive effect of reflection: motivation and cognitive processing of information.

Before presenting the studies conducted as part of the thesis, this introductory Chapter will synthesize the literature that provides the conceptual framework behind the studies. As the definition of reflection itself is not consensual, the general concepts of reflection as adopted by this thesis will be presented, followed by their applications into the field of clinical reasoning. Subsequently, the evidence from cognitive and educational psychology research that supports the hypotheses of this thesis on the effects of reflection on students' subsequent learning of scientific texts will be discussed. Finally, the thesis' questions and a brief description of the experiments designed to answer each one of them will be provided.

WHAT IS REFLECTION

In our daily lives, we can make sense of many things without much cognitive effort. If we find our living room floor wet in the morning and see that a window was left opened during the night, it will be inevitable to think that rain was the cause of the problem. If it indeed rained during the night, this initial conclusion is confirmed. However, if it did not rain and we cannot easily find an alternative explanation for the problem, a state of doubt will arise and demand a conscious effort to understand what is happening and how we should react to it. John Dewey,[4] in the early 1900's, called this deliberate reasoning process *reflective thinking*, which contrasts with the *unintended thinking* that "pops up" in our minds, the unconscious reasoning processes we use to solve a great deal of familiar problems. In Dewey's words, reflective thinking *is an active, persistent and careful consideration of any belief or supposed form of knowledge in the light of* *the grounds that support it, and the future conclusions to which it tends.*[4] A wet floor with no clear cause will ask for "testing" different possible explanations against any available evidence until a solid conclusion on its cause can be made. If we, however, assume the first hypothesis that emerges as the cause of the problem, "jumping" into premature conclusions, reflective thinking will not take place. Reflection demands that we suspend judgement until a careful analysis of a problem allows for the emergence of evidence that support our conclusions.

Reflection, therefore, is a reasoning process that helps us to solve problems to which we do not have, at a first glance, clear solutions. It is easy to understand the value of reflection for professional practice, particularly when uncertainty prevails, as in medicine. The diagnostic tasks that medical doctors face in their daily practice, for example, are often challenging. A single symptom can be caused by different diseases, a single disease can have different presentations and self-medication can blur things even more, to mention some of the many possible sources of uncertainty. In summary, clinical problems are frequently "messy" problems that do not allow for immediate conclusions.

However, medical schools have not always acknowledged the relevance of preparing future doctors to deal with uncertainty. In the 1980's, Donald Schön contributed to change this scenario. He raised attention to the fact that schools usually prepared future professionals to deal with organized, well-structured problems that are found in traditional academic classes and books, not for the ill-defined problems that happen in the real world. He suggested schools should teach students to manage uncertainty and learn from it.[5]

For Schön, one can respond to the unexpected by brushing it aside, ignoring the elements that challenge one's current knowledge. Alternatively, s/he can respond to it by engaging in reflection. Reflecting while solving a problem, which he named *reflection in action,* addressing uncertainty by restructuring one's own initial explanations and solutions as needed, certainly favors better outcomes. After uncertainty is initially managed, however, reflection can still go on. Reflecting "back" at the uncertainty experienced and how one responded to it or, according to Schön *reflecting on it,* could provide a better understanding of the situation, eventually changing previous conclusions.[5] For Schön, both reflection *in* and *on* action could also be of help to understand why a situation is/was challenging, to identify factors in the context and in ourselves that contributed to the difficulties experienced. Both can help one to become aware of gaps in his/her knowledge that would be useful to manage the situation and therefore be a motivational drive to engage in learning activities to fulfill these gaps. Reflection could, therefore, not only lead to better solution to the problem at hand, but also reshape knowledge and practice, supporting professional development.[5] Schön's theories contributed to the raise of reflection as an important ability to be developed by students and as a learning tool to be used by teachers. In all levels of medical education, reflection entered the agenda.[1-3]

Reflection could be triggered and can be helpful in any challenging medical situation, be it emotional or cognitive. Although all aspects are relevant for patient care, this thesis focuses on reflection in the context of diagnostic reasoning, or the process of interpreting patients' clinical data to find out what is their most likely diagnosis.

Reflection in the process of diagnostic reasoning

According to Dewey, reflection is not the random emergence of information to conscience: it is a voluntary effort to create a logical chain of knowledge or ideas that support each other in order to find a conclusion to a problem.[4] In the context of clinical reasoning we could have, for example, a patient who presents with cough. If this symptom is not accompanied by a set of other clinical findings that allow a prompt diagnosis, one will have to interpret clinical data in a systematic way to understand it. The random guessing of different hypothesis, with no connections to the patient at hand, is not the best way to solve the problem. Instead, it would be more reasonable to make a hypothesis based on some of the clinical data and test this initial tentative solution against additional information. If this hypothesis is supported by additional data instead of refuted by them, the initial solution may be confirmed. If it does not, other possibilities will probably have emerged at this point and could also be tested until a hypothesis finally fits the data. This would be the inductive/deductive process of reflection as defined by Dewey[4] and resembles the hypothetico-deductive approach used by medical doctors to diagnose patients' problems.[6,7]

Building on the theory of Dewey and Schön,[4,5] Mamede, Schmidt and Penaforte[8] developed a procedure to diagnose medical clinical cases that translates reflection to the domain of diagnostic reasoning. It consists of a deliberate act of comparing and contrasting different diagnostic hypotheses by matching the findings that would be expected if each were correct with the patient's clinical findings before deciding on the most likely diagnosis for the patient. This analysis consists of identifying clinical data that supports the first tentative hypothesis, followed by the identification of data that contradicts this hypothesis and data expected, were this hypothesis correct, but that is lacking in the case. After the analysis of the first hypothesis, two alternative hypothesis are generated, and the same procedure is run with each of them. Finally, hypotheses are ranked according to probability. One example of this procedure can be seen in Appendix 1. To highlight the deliberate effort that characterizes this process, it was named *deliberate reflection*, a term that will be used henceforth. This deliberate reflection procedure alone, without any additional support, has been shown to improve medical doctors' accuracy on challenging clinical cases,[8] adding empirical support to the assumption that reflection would be helpful in demanding professional tasks. But could it better inform future tasks as well?

There is some evidence that this deliberate reflection procedure can also improve performance in future diagnostic tasks, including two experiments run by Mamede *et al* with fourth-year medical students. In the first experiment, students diagnosed clinical cases either reflecting upon the cases, giving a single diagnose or giving differential diagnosis, i.e., to cite the most likely hypothesis and two alternative ones, without further actions (phase 1).[9] A week later, students diagnosed new clinical cases with the same diseases they had seen a week before (phase 2). The second experiment was similar except that in the second phase students also diagnosed clinical cases with similar clinical presentations of the diseases they diagnosed a week before but different diagnoses ("adjacent" diseases).[10] Reflection improved students' diagnostic accuracy both on the same and adjacent diseases. No learning tasks related to the clinical cases were offered to the students in between the two phases of the experiments, and the positive effect of reflection only emerged in the experiments' second phases. These findings suggest that deliberate reflection helped students to reorganize the knowledge that they already had about the clinical cases, improving their mental representations of the diseases and, thereby, better informing future diagnostic tasks.

Employing the deliberate reflection procedure in the diagnosis of the simulated clinical cases *per se* has been shown, therefore, to have positive outcomes for medical students' diagnostic performance when they encounter future cases. The positive effect of reflection came out even if the students were not exposed to new knowledge about the diseases encountered in these cases. However, the students certainly did not know everything about these diseases. In fact, they would certainly need to acquire additional knowledge by engaging in learning strategies, such as individual study or lectures about relevant contents. Would deliberate reflection while diagnosing clinical cases also foster students' learning from such subsequent activities? It is well-established that solving a problem before the study of materials relevant to that problem positively affects what students learn from these materials.[11,12] However, if different approaches to diagnosing clinical problems would have different outcomes in supporting other relevant learning activities is yet to be investigated. It may be that any way of diagnosing clinical cases before studying, for example, a book chapter on the case's disease would help. Investigating whether, and how, deliberate reflection upon clinical cases would foster subsequent learning of scientific text is the purpose of this thesis. This is relevant because helping students to *expand* their knowledge about diseases is as important as helping them to *reorganize* their current knowledge. Before introducing the research questions, however, two mechanisms that could support the assumption that deliberate reflection upon clinical cases could foster learning of new information, *motivation* and *cognitive processing,* will be discussed.

DELIBERATE REFLECTION AS A LEARNING TOOL: THE MOTIVATION PATH

Could reflection trigger interest?

Interest is both an individual's predisposition to engage in a certain task and a psychological state of heightened attention, increased cognitive function and persistence.[13,14] In order for one to engage, persist and master a task, like studying a scientific text, s/he must be interested in it.

We all have "personal" interests. Some medical students, for example, prefer surgery over family medicine and others rather study heart than study brain diseases. These *individual interests* are usually developed through a long period of time, are relatively stable and have long-term consequences.[14,15] It can influence career-choices, for example. There is another kind of interest, however, that is more in the "interestingness" of tasks or information than in individuals' (stable) preferences. It is a context-related interest and, therefore, usually short-living, which is known as *situational interest* (SI).[14] Initial understanding about SI came from reading research, from which we learned that it is triggered, for example, by text content features with which the reader can identify him/herself. A good example of SI interest was given by the Covid-19 pandemic, when many who were never remotely interested in epidemiology became interested in understanding some of its concepts, like herd immunity. Reading research has informed that SI can also be triggered by text structural characteristics, such as unexpectedness, incongruity and uncertainty.[13,14]

Becoming interested in a certain topic is a good first step to learn about it but, for learning to occur, SI, once triggered, must be sustained. If it fades before one engages with a learning activity long enough to learn from it, SI will be wasted. In the academic context teachers can, for example, manipulate tasks presented to students after triggering their interest: if the task is meaningful to students, they should persist with it. It could be that learning about the topic would, in itself, sustain interest: the more we know about a certain topic, the further we would want to study it, to the point that a SI can eventually become an individual interest.[15] There is no consensus about how SI evolves, however, and some authors believe that SI has an underlying component that would, in fact, trigger it and determine our behavior afterwards: the *knowledge-gap hypothesis.*

Awareness of knowledge gaps and situational interest

By building upon different theories on curiosity and findings from studies outside education, Loewenstein[16] hypothesized that the intrinsically motivated desire for specific information, which resembles the concept of SI, would not be a primary but rather a secondary drive mediated by the unpleasant feeling of knowledge deprivation. According to Loewenstein, when one who has a "background" knowledge about a topic realizes gaps in this knowledge, a feeling of deprivation would emerge and trigger a will to close the specific gap of this specific topic. Once the gap is closed, the unpleasant feeling of deprivation is gone, and curiosity would fade.

To explore Loewenstein's knowledge-gap hypothesis for curiosity in the educational context, Rotgans and Schmidt observed the triggering and evolution of situational interest along educational activities in a series of experiments. They developed a self-reported questionnaire to measure SI that assess both attention and affect dimensions of it. It consists of six questions such as "I am fully focused on today's topic; I am not distracted by other things" and "I will enjoy working on today's topic" (see Appendix 2). The questionnaire, scored on a 5-point Likert-type scale ranging from 1 (not true at all) to 5 (very true for me), has been shown to have good construct validity as a confirmatory factor analysis confirmed the hypothesis that the items measured one underlying factor.[17] The questionnaire reliability was determined by Hancock's coefficient *H* in different studies which ranged from 0.9-0.95, above the threshold of 0.7, the recommended cut-off.[17,18] With this questionnaire, they measured students' SI in different moments throughout real classroom problem-based learning (PBL) activities.

A PBL cycle of activities begins with the presentation of a problem that will support the discussions of a topic. Next, students and tutor engage in an initial discussion of the problem that aims at the generation of learning goals. Subsequently the students, individually, study the learning goals and, finally, the class elaborates on the problem.[11] With different measures along such activities, they were able to check if SI can be indeed triggered by a problem, which is supposed to induce a feeling of uncertainty, and how it develops as students learn more about this problem. In one observation, polytechnic students' SI raised after the problem presentation and decreased after individual study, suggesting that, as students learned more about the topic being studied, their interest in it faded.[17] A second observation, this time with secondary school students, measured both students' SI and knowledge on the topic being studied at different times. They observed that, as students' knowledge increased, their SI decreased.[18] These findings support the *knowledge-gap hypothesis*: the problem would reveal students' knowledge gaps about the topics and trigger their interest, or their "thirst", to fulfil those gaps. Once the "thirst" was satisfied, interest faded.[18] Two experiments run by the same authors and using the same SI questionnaire, in which they manipulated students' prior knowledge about the topics being studied, also support this assumption: studying the topic before discussing a problem about it inhibited students' SI triggered by the problem.[18,19] But is it necessary that one is *consciously* aware of knowledge gaps to become interested in a specific topic?

Rotgans and Schmidt measured secondary school students' SI, using the aforementioned questionnaire, about a geography topic and also asked students to answer the question "how much do you know about this topic?" in a 5-point scale: form 1 (very little) to 5 (more than enough), a measure of their awareness of knowledge gaps (AKG) on the topic.[18] They found SI to be negatively related to AKG, suggesting that students had to be aware of their knowledge deficits about the topic in order to become interested in it.

However, there is contradictory evidence regarding the need of AKG to trigger SI. Glogger-Frey *et al*[12] also studied this topic in a different context. They compared student teachers' AKG, epistemic curiosity[1], interest and learning outcomes when they had to *invent a solution* to a task or to study a canonic solution to the same task, before they were exposed to a learning activity on this topic. They compared, therefore, two "preparatory" strategies to a subsequent direct learning activity: *inventing-a-solution* and *studying-a-canonic solution*. To that end, they developed a questionnaire on AKG with nine questions such as "this task showed I don't know things yet" and "I was not able to find a proper solution to this task" (see Appendices 2 and 5), with a six-point rating scale, with 6 being *absolutely true*. This questionnaire showed good reliability measured by Cronbach's alpha: 0.89. Students in the inventing-a-solution group showed higher scores both on AKG and epistemic curiosity right after the preparatory phase of the experiment, relative to those who studied the canonic solution.[12] In an experiment with a similar design run by the same authors with 8th grade students working with a physics topic, AKG was also higher for the *inventing-a-solution* group relative to the control group, but epistemic curiosity was similar between groups. It could be that the effect of AKG on SI might depend on other factors, perhaps students' self-regulation skills, but if AKG is needed to trigger SI is a topic still to be explored.

Does interest influence engagement with learning activities and learning outcomes?

It is reasonable to assume that the more one is interested in a topic, the more s/he will engage with learning activities relevant to this topic and, ultimately, learn about it. The empirical evidence that this actually happens comes mostly from reading research, which has been shown that interest influences text-reading choices, engagement with texts and text information processing.[14,15,20,21] Using a computer software, Ainley *et al*[14] tracked 8th grade students' self-reported interest and behavior while reading short texts about topics like chameleons and body image. After correcting findings for the order in which texts of the different topics were presented on screen, a positive effect of topic interest emerged. Students that showed a positive interest in chameleons, for example, chose the text about chameleons ahead of others, and those who showed a negative interest in chameleons delayed the reading of this text. In this experiment, topic interest also positively influenced students' persistence with reading the different texts. The persistence with the texts also positively affected students' scores on tests about

1 Although epistemic curiosity and situational interest can be seen as different concepts, there is evidence that they share similar underlying psychological mechanisms.[19]

the texts, suggesting that topic interest, a context-related interest and, therefore, a situational interest, can foster students' learning by motivating them to persist with a learning task.

Schiefele and Krapp[20] also studied the influence of interest on students' learning of texts. They compared the recall of texts of college students that showed high and low interest about the content of those texts. Interest about the texts' contents was measured with a self-reported questionnaire. Learning about the texts was measured through a test with questions of different levels of complexity: from simple recall of information to application of information to novel situations. Interest positively influenced students' scores on more complex questions but not on simple questions. While text interest did not influence how much *verbatim* information students remembered about the text, it triggered deeper levels of information processing about it. These results provide empirical support that interest can positively affect learning, at least learning of texts. It should be noticed, however, that if interest is directed towards distracting features of a text, it can hinder learning of relevant information about it.[14]

Evidence of the influence of interest in learning outcomes outside the reading research field is, however, scarcer. Nevertheless, SI has been shown to be a good predictor of engagement in learning and learning in classroom observations.[17,22] Linnenbrink-Garcia *et al*[22] measured the SI of secondary students about the learning courses they were taking at a science camp. They also asked the courses' instructors to assess students' participation in class activities, group discussions and their contributions to discussions with thoughtful and provoking questions. These assessments were taken as measures of students' engagement with the learning activities. Students with higher levels of what the authors called *maintained-SI-feeling* had higher scores on their assessments of engagement relative to those who showed lower *maintained-SI-feeling*'s scores. Similar results were found by Rotgans and Schmidt[17] observing polytechnic students in a day of PBL activities: SI was positively correlated with students' achievement behavior as measured by their tutors. Moreover, and very important, Rotgans and Schmidt also measured these students' scores on a test about the topic of interest of this PBL activities, at the end of the day, and found that higher achievement behaviors translated into higher scores on this end-of-session test.[17] These findings support the expectations that interest would positively influence learning also in real classrooms. In the aforementioned experiment of Glogger-Frey *et al*[12] with student teachers, however, higher levels of epistemic curiosity did not translate into higher engagement with a learning task, measured as the time students spent studying the topic. Higher epistemic curiosity also resulted in no better learning about the topic, measured as students' scores on a task that required transfer of the studied knowledge. The conflicting results might be a consequence of the use of different instruments to measure interest, academic behavior and achievement used in the experiments, but it is also possible that personal and context factors, yet to be understood, could influence the mediating processes that happens between the emergence of SI and the learning consequences of this interest.

Deliberate reflection can foster some of the factors associated with situational interest, such as challenge and uncertainty. Therefore, it is reasonable to expect that reflection would be a better trigger of SI than more conventional clinical cases' diagnostic approaches, such as differential diagnosis. It is also reasonable to expected that this "booster" in students' SI would also translate into students' enhanced learning about information relevant to the cases that are subsequently presented to them. Experiments 1 and 2 of this thesis explored these assumptions.

DELIBERATE REFLECTION AS A LEARNING TOOL: THE COGNITIVE PATH

As discussed so far, problem-solving strategies have been shown to be, at least in some contexts, a good motivational drive to students. But problem-solving can also operate in favor of learning through cognitive mechanisms, such as the activation, elaboration and restructuring of knowledge.

The dual-store model of memory proposes that learning occurs when we pay enough attention to a certain information, such as a text fragment we are reading, to 'move' it to our working memory. Once in the working memory, it can be processed and, finally, stored in the long-term memory. Working memory would be the conscious '*locus*' of our memory system where 'thinking' occurs, i.e., where we organize new information and integrate it with what we already know about the particular subject.[23] The integration of new and prior knowledge in our working memory seems to be a critical point of learning because it would allow us to build organized knowledge structures that, once stored, are easier to retrieve in the future.[24] To be used while processing new information, however, our prior knowledge on a certain subject has to be mobilized from long-term to working memory or, in other words, to be 'activated'.[25] Activation of prior knowledge is, therefore, important to the learning of new knowledge. "Bringing" new and prior knowledge into working memory at the same time offers the opportunity to elaborate on them and to create organized knowledge structures. This can be as simple as relating a phone number to one's own birthdate or as complex as linking distinctive features to specific diseases with similar presentation that would allow to differentiate them. The latter example is clearly more challenging: it would be necessary to make sense of new knowledge and to organize it coherently with prior knowledge to finally have a renewed and more sophisticated knowledge structure on the subject that can be accurately retrieved from long-term memory in the future.[25] Generating explanations is an instructional strategy that can facilitate such elaboration processes, as research on PBL and self-explanation demonstrates.[26-31] For example, Van Blankenstein *et al*[26] compared two groups of undergraduate students who, individually, watched a recorded group discussion of a problem on physics either passively or interacting with it, by giving their own explanations for the problem along the video. Subsequently, students studied a text about the subject and, finally, did a recall-test about the problem. Students who interacted with the discussion scored around 30% higher on the immediate test relative to those who watched the discussion passively. An advantage of the interaction was also observed one month after the experiment. Similar results were found by Larsen and Roediger[29] with first-year medical students working with neurology topics in an experiment with self-explanation. These studies suggest that the attempts to explain problems facilitates the elaboration and the retention of new knowledge relevant to the problems.

The aforementioned deliberate reflection procedure developed by Mamede, Schmidt and Penaforte[8] has been shown to improve medical students' diagnostic accuracy without any attempt to expand students' knowledge, probably helping them to reorganize their current knowledge into better knowledge structures. Could it also facilitate learning of new knowledge as other problem-solving instructions, like PBL, have been shown to do?

The initial analysis of a problem in a PBL session can motivate students to look for learning resources, such as texts, and engage with them.[11,27] By activating students' prior knowledge, it can also facilitate learning from these resources.[26] It is reasonable to expect that deliberate reflection on clinical cases prior to exposure to new knowledge about the diseases present in the cases could have similar effects on learning. By compelling students to consider alternative diagnoses for a clinical problem in a structured and conscious way, deliberate reflection would potentially confront students with knowledge deficits. Such reflection would also only be possible by bringing to mind prior knowledge about the plausible diagnoses under consideration, which would be elaborated upon while attempting to weighing alternative explanations.

The second and third experiments of this series explores if deliberate reflection upon clinical cases could foster medical students' subsequent learning of scientific texts that contain relevant information to the diagnosis of these clinical cases, presented to them just after the diagnosis of these cases. The third and fourth experiments explores the mechanisms that could underly such positive effect of reflection.

RESEARCH QUESTIONS AND OVERVIEW OF THE STUDIES IN THIS THESIS

The research questions of this thesis are:

1. Would deliberate reflection upon clinical cases foster medical students' motivation to engage in subsequent study of scientific texts relevant to these cases?
2. Would deliberate reflection improve the learning outcomes of scientific text study?

3. Which mechanisms would underly a positive effect of deliberate reflection on learning of scientific texts: motivation and/or cognition?

Four experiments were run to answer these questions and each one is described in separate chapters of this thesis. The series of experiments started by building upon previous evidence from fields other than medical education showing that challenging tasks that raise uncertainty can trigger students' situational interest.[13-19] Diagnosing clinical cases, whatever the strategy used, is likely to be a challenging task to medical students. Deliberate reflection upon the cases, which demands scrutinizing different diagnostic hypothesis before making a conclusion, is expected to be more challenging and to raise more uncertainty than, for example, to cite alternative diagnoses alone. The first experiment, reported in Chapter 2, compared medical students' situational interest, measured through a self-reported questionnaire,[17] triggered by the diagnosis of clinical cases through deliberate reflection and differential diagnosis.

In the first experiment, students who deliberately reflected upon the cases showed higher scores on SI, relative to those who gave a differential diagnosis. The second experiment moved forward with a twofold purpose. First, obtaining additional evidence of the effect of deliberate reflection on students' motivation, now through a behavioral measure: study-time. Second, examining whether this enhanced motivation to study would translate into higher learning. In the experiment described in Chapter 3, the time medical students spent reading a scientific text relevant to clinical cases just after they diagnosed the cases, either through deliberate reflection or differential diagnosis, was compared. The experiment also compared students' scores on a subsequent knowledge test about the text they had just studied. These scores were taken as a measure of learning of the text contents.

In the second experiment, relative to differential diagnosis, students who reflected upon clinical cases engaged longer in studying a text relevant to these cases and showed higher scores on a test about this text. We were unable, however, to understand if the higher scores on the test were a consequence of the longer engagement of students with the text, i.e., simply a consequence of higher motivation to study it, or if any cognitive mechanism also played a role. Considering the evidence from other problem-solving strategies,[26-31] we hypothesized that, relative to differential diagnosis, deliberate reflection would facilitate students' cognitive processing of the text. For example, reflection could facilitate the incorporation of new information from the text to students' current knowledge. The third experiment resembled the second, but besides manipulating diagnostic strategy (either deliberate reflection or differential diagnosis), the time students could engage with the text relevant to the cases was also manipulated: students could read it with no restriction of time, i.e., as long as they wanted, or with restriction of time. Study-time manipulation intended to "isolate" the motivation role in students' learning from the text and to allow the emergence of an effect that could not be attributed to motivation alone, revealing therefore a cognitive processing effect. Chapter 4 reports on this experiment.

The results of study three suggested that the positive effect of deliberate reflection that was observed on students' learning was predominantly a consequence of cognitive mechanisms, rather than motivation. However, this study did not explore which such mechanisms were at play. Chapter 5 describes an experiment that explored the potential of deliberate reflection to positively influence medical students' activation of and elaboration on prior knowledge about the diseases presented in the cases. The experiment compared the amount of information that medical students could recall about a set of diseases when they had previously diagnosed a set of clinical cases of these diseases (either through deliberate reflection or differential diagnosis) or with no prior exposure to clinical cases.

Finally, Chapter 6 closes this thesis with a summary of the main findings of the experiments and a discussion on how they can contribute to medical teachers in their daily practice and inform future research.

REFERENCES

1. Sandars J. The use of reflection in medical education: AMEE Guide No. 44. *Med Teach.* 2009;31(8):685-95.
2. Mann, K, Gordon, J, MacLeod, A. Reflection and reflective practice in health professions education: a systematic review. *Adv Health Sci Educ*. 2009; 14: 595–621
3. Uygur J SE, De Paor M, Wallace E, Duffy S, O'Shea M, Smith S, Pawlikowska T. A Best Evidence in Medical Education systematic review to determine the most effective teaching methods that develop reflection in medical students: BEME Guide No. 51. *Med Teach.* 2019; 41(1):3-16.
4. Dewey J. *How we Think*. Boston: D. C. Haeth and Co.; 1913.
5. Schön, D. A. Understanding The Need for Artistry in Professional Education *in* Schön, D. A. (1987). *Education The Reflective Practitioner: Toward a New Design for Teaching and Learning in Professions.* San Francisco Jossey-Bass Higher Education Series, pp 1-20.
6. Elstein AS, Schwartz A. Clinical problem solving and diagnostic decision making: selective review of the cognitive literature. *BMJ*. 2002; 324: 729
7. Norman, G. Research in clinical reasoning: past history and current trends. *Med Educ.* 2005; 39: 418–427
8. Mamede S, Schmidt HG, Penaforte JC. Effects of reflective practice on the accuracy of medical diagnosis. *Med Educ*. 2008; 42:468-75.
9. Mamede S, van Gog T, Moura AS, de Faria RMD, Peixoto J, Rikers RMJP, *et al*. Reflection as a strategy to foster medical student's acquisition of diagnostic competence. *Med Educ*. 2012; 46: 464-72.
10. Mamede S, van Gog T, Moura AS, de Faria RMD, Peixoto JM, Schmidt HG. How can students' diagnostic competence benefit most from practice with clinical cases? The effects of structured reflection on future diagnosis of the same and novel diseases. *Acad Med*. 2014; 89(1):1-7.
11. Schmidt, H.G. Problem-based learning: rationale and description. *Med Educ*. 1983; 17 (1): 11-16.
12. Glogger-Frey I, Fleischer C, Gruny L, Kappich J, Renkl A. Inventing a solution and studying a worked solution prepare differently for learning from direct instruction. *Learn Instr*. 2015; 39:72-87.
13. Hidi S. Interest and its contribution as a mental resource for learning. *Rev Educ Res*. 1990; 60(4): 549-571.
14. Ainley M, Hidi S, Berndorff D. Interest, Learning and the Psychological Processes that Mediate their Relationship. *J Educ Psychol*. 2002; 94(3):545-61.
15. Hidi, S, Renninger A. The Four-Phase Model of Interest Development. *Educ Psychol*. 2006;41(2), 111-127.
16. Loewenstein G. The psychology of curiosity: a review and reinterpretation. *Psychol Bull*. 1994;116(1):75-98.
17. Rotgans JI, Schmidt HG. Situational interest and academic achievement in the active-learning classroom. *Learn Instr*. 2011; 21: 58-67.
18. Rotgans JI, Schmidt HG. Situational interest and learning: Thirst for knowledge. *Learn Instr*. 2014; 32: 37-50.
19. Schmidt HG, Rotgans JI. Epistemic Curiosity and Situational Interest: Distant Cousins or Identical Twins? 2021. *Educ Psychol Rev.* 2021; 33: 325–352.
20. Schiefele U, Krapp A. Topic interest and free recall of expository text. *Learn Indiv Diff*. 1996; 8(2):141-160.
21. Schraw G, Lehman S. Situational Interest: A Review of the Literature and Directions for Future Research. *Educ Psychol Rev.* 2001; 13: 23–52.

22. Linnenbrink-Garcia L, Pattal EA, Messersmith EE. Antecedents and consequences of situational interest. *Br J Educ Psychol*. 2013; 83(4): 591-614.
23. Ormrod JE. Introduction to memory and cognition. In: Ormrod JE, editor. *Human Learning*. 5th ed ed. Boston: Pearson Education; 2008. p. 150-92
24. Mayer RE. Information Processing. In: Harris KR, Graham S, Urdan T, eds. *APA Educational Psychology Handbook. Vol. 1. Theories, Constructs and Critical Issues,* 1st ed. Washington DC: American Psychology Association 2012; 85-99.
25. Ormrod JE. Long-term memory storage and retrieval processes. In: Ormrod JE, editor. *Human Learning.* 7th ed ed. Boston: Pearson Education. 2016. p. 206-48.
26. van Blankenstein FM, Dolmans DHJM, van der Vleuten APM, Schmidt HG. Which cognitive processes support learning during small-group discussion? The role of providing explanations and listening to others. *Instruc Sci*. 2011; 39: 189-204.
27. Schmidt HG, Rotgans J, Yew EH. The process of problem-based learning: what works and why. *Med Educ*. 2011; 45(8): 792-806.
28. Chi MTH, de Leeuw N, Chiu M-H, LaVancher C. Eliciting self-explanations improves understanding. *Cognitive Sci*. 1994; 18: 439-77.
29. Larsen DP BA, Roediger III HL. Comparative effects of test-enhanced learning and self-explanation on long term retention. *Med Educ*. 2013; 47: 674-82.
30. Schmidt HG, De Volder M, De Grave WS, Moust JHC, Patel VL. Explanatory model in the processing of science text: the role of prior knowledge activation through small-group discussion. *J Educ Psychol*. 1989; 81(4): 610-9.
31. VanLehn K CM. A model of the self-explanation effect. *J Learn Sci*. 1992; 2(1): 1-59.

Chapter 2

Effect of reflection on medical students' situational interest: an experimental study.

Ligia Maria Cayres Ribeiro, Silvia Mamede, Alexandre Sampaio Moura, Eliza Maria de Brito,
Rosa Malena Delbone de Faria, Henk G. Schmidt

Published in: Medical Education 2018 May; 52(5): 488-496. DOI 10.1111/medu.13491

ABSTRACT

Context: Reflection has been considered crucial to learning. Engaging in reflection while solving problems is expected to foster identification of knowledge gaps and interest in learning more about them, the latter a major motivational force in learning. Although theoretically sound, this assumption still lacks empirical evidence.

Objective: This experiment investigated whether reflection while diagnosing clinical cases of different levels of difficulty influences medical students' awareness of knowledge gaps and situational interest.

Methods: Forty-two 4th-year students from a Brazilian medical school were randomly allocated to diagnose 6 clinical cases (3 difficult; 3 easy), either by following a structured-reflection procedure (reflection group) or by giving alternative diagnoses (control group). Subsequently, for each case, all students rated their situational interest and awareness of knowledge gaps.

Results: Situational interest was significantly higher in the reflection than in the control group (respectively, mean=4.10, standard deviation=0.50 vs mean=3.66, standard deviation=0.48; p=.005; range:1-5). The effect size was large (d=0,92). Awareness of knowledge gaps was higher in the reflection than in the control group, but the difference was not significant. Case difficulty influenced both situational interest, which was significantly higher on easy than difficult cases (respectively, mean=3.96, standard deviation=0.56 vs mean=3.80, standard deviation=0.55; p=.004) and awareness of knowledge gaps, with higher scores observed on difficult than easy cases (respectively, mean=3.99, standard deviation=0.46 vs mean=3.66, standard deviation=0.53; p<.001). No interaction between experimental condition and case difficulty emerged.

Conclusion: Relative to providing alternative diagnoses while solving cases, structured reflection increased medical students' interest on them and may therefore be a useful tool for teachers concerned with enhancing students' motivation for learning. Surprisingly, easy cases promoted higher situational interest despite the higher awareness of knowledge gaps on difficult cases. This suggests an inhibitory potential of case difficulty on students' interest to learn, a possibility that demands further investigation.

INTRODUCTION

Reflection has been defined as the attentive thought one gives to a particular event or problem, when faced with uncertainty or the unexpected, in order to understand and appropriately respond to that given situation.[1,2] When engaged in reflection, one searches for possible explanations or solutions for the problem at hand, explores their implications and check whether they are valid. The results would then be a renewed, usually better, understanding of the problem which may be beneficial not only in the present situation but also in similar problems confronted in the future.[2] However, this is not the only way through which reflection upon one's experiences may improve performance. Reflection has been seen as an important learning strategy because it may also reveal gaps in knowledge, skills or attitudes,[3] which would trigger interest in learning. This would be an important "side effect" of reflection because interest is recognized as a crucial determinant of engagement in learning activities and, consequently, of learning outcomes.[4,5,6] These assumptions, however, should not be taken for granted as they rely mostly on a theoretical framework, and empirical evidence that they actually happen is still missing, especially in medical education.[7] This article reports on a study on the effects of reflection upon clinical cases on medical students' interest to learn more about them.

In medical practice, physicians may engage in reflection when in a state of uncertainty or perplexity triggered, for example, by diagnostic challenges or unexpected treatment outcomes.[8] Failures and novel situations that require creative approaches have also been pointed out as reflection triggers.[9] While reflection-on-practice involves thoughtful analysis of a past experience, practitioners also engage in reflection in the moment of problem-solving to critically review their understanding of and proposed solution for it (reflection in action).[2,9] Reflection while diagnosing clinical problems may therefore involve generating a tentative diagnosis, confronting this hypothesis with the patient's clinical findings, seeking alternative explanations for them and verifying these alternatives until setting a solution for the patient.[8] Reflection is also said to give the professionals the opportunity to seek for what is "behind" their challenge: which elements in the problem or in themselves have led to uncertainty. Specially by looking at themselves and at their own reasoning and responses to the situation, the professionals are expected to be aware of the knowledge that could have been useful to deal with the situation.[3] Therefore, they can recognize learning needs and engage in further activities to fulfill them, expanding their knowledge.[2] It is noteworthy that this research tradition, as well as the more recent research discussed in following, has directed little attention to other factors that may play a role in becoming aware of knowledge gaps, such as the accuracy of one's self-assessment. It is to keep in mind that reflection is not a judgmental perception of one's level of competence, and, as some authors have stressed, should not be confounded with self-assessment, an ability not always reliable.[10] Reflection on one's own experiences has shown to be in itself an important mechanism to achieve and maintain professional expertise throughout life as the studies on the role of deliberate practice in expertise have demonstrated.[11] If reflection upon one's experiences is considered to trigger learning among practicing physicians, it is likely to play a similar role among medical students.

Between the awareness of knowledge gaps (AKG) and the engagement in a learning activity, however, there must be a link: the interest in the topic and its consequent motivation to seek for learning. Interest, a psychological state of heightened attention, concentration and affect, is crucial in determining if one will engage, and persist, in a learning task.[4] There is substantial evidence that interest enhances learning by focusing students' attention on a task, increasing their persistence with the task, and developing more positive affection on it. [12,13,14] A student can be generally interested in a certain subject, such as a medical specialty. This has been named "individual interest" and is a reasonably stable, internal condition.[4] However, another, more short-lived sort of interest may also be generated by environmental stimuli, usually associated with conflict and uncertainty, like a diagnostic challenge. This "situational interest" (SI) focuses learners' attention on a learning activity and compels them to persist on it.[5,12,15] "Situational interest" has then been described as "thirst" for knowledge. It seems the type of interest that could arouse in a medical student who reflects upon a to-be-diagnosed clinical case and realizes that he would need to know more about a particular topic to be able to differentiate between the seemly possible diagnoses.

The role of reflection on clinical cases as a trigger of SI, though plausible, remains as a conjecture as most of the empirical evidence in SI research comes from studies within different settings, such as primary and secondary education. In line with what would be expected in reflective practice, these studies have found novelty and unexpectedness of the information and confrontation with a challenging problem to be potential triggers of interest. For example, Rotgans and Schmidt observed real-life, challenging problems to increase students' SI in classroom activities, though a subsequent study showed that problems fostered SI only in the presence of students' self-perception of knowledge gaps in the subject.[6,12] This latter finding supports the account that AKG is a key factor in triggering-maintaining curiosity[1] and suggests that it may be an underlying mechanism of SI.[12] Nevertheless, it is not entirely clear whether being aware of one's own gaps in knowledge is a requisite for, or is associated with, increase in SI. Indeed, there is some contradictory evidence. Glogger-Frey *et al*, for example, did not find correlation between AKG and SI in two different studies with psychology students.[16] Whether situational interest is necessarily related with AKG remains, therefore, unclear. What seems clear, from the existing research, is that engaging in solving problems, especially challenging problems, leads to increased SI, perhaps associated with perception of knowledge gaps.

The aforementioned studies have been conducted in learning environments and with materials very different from medical education. Nevertheless, they suggest that engaging students in solving problems that are challenging enough to require them to mobilize their prior knowledge fosters SI, especially when the problems facilitate recognition of the limitations of this prior knowledge. Structured reflection upon to-be-diagnosed cases confronts medical students with the challenging of comparing alternative diagnoses for the case, which requires activation of existing knowledge to weigh the evidence in favor of each of diagnosis and leads to scrutinizing the grounds of one's reasoning. Diagnosing through reflection can be seen therefore as a more demanding task than simply providing a differential diagnosis, and it would be reasonable to expect that reflection upon to-be-diagnosed cases may foster high levels of SI among medical students, especially when the problems raise uncertainty. Whether this depends on AKG or not is difficult to preview, considering the findings of the aforementioned studies.

To test this idea, we conducted an experiment with 4^{th}-year medical students comparing their AKG and SI when they worked on solving clinical cases from different levels of difficulty, either deliberately reflecting upon them or making differential diagnosis. We hypothesized that: (1) structured reflection would foster SI relative to giving differential diagnosis; (2) SI would vary according to case difficulty, being higher when students worked with difficult cases; and (3) there would be an interaction between experimental condition and case difficulty, possibly with higher SI when students reflected on difficult than on easy cases. Regarding the relationship between SI and AKG, we had no prior hypothesis due to the conflicting findings of previous research.

METHODS

Design

The study was a one-phase experiment conducted in October 2015, with random allocation of participants to either an experimental or a control condition. In both conditions, participants diagnosed both difficult and easy cases by following a procedure to reflect upon the cases (experimental) or to make differential diagnosis (control). After diagnosing the cases, participants rated their situational interest and awareness of knowledge gaps.

Setting and participants

All 54 4th-year medical students at José do Rosário Vellano (UNIFENAS), in Belo Horizonte, Brazil, were invited to participate in the study. UNIFENAS has a six-year problem-based curriculum, with the two final years dedicated to clerkships. We selected 4-th year students because at this point in their training they have been exposed to knowledge about the diseases used in the study during tutorial groups and lectures but have limited clinical experience with them.

All volunteers were recruited as participants and gave written consent. The experiment was run as an extra-class activity. Subsequently, a teacher discussed the cases with the participants, who received an extra-curricular activity certificate as acknowledgement for their participation.

Materials and procedure

Six written clinical cases were employed in the study. The cases consisted of a description of clinical symptoms, physical examination and laboratory tests findings. Each case had a most likely diagnosis that had been validated in previous studies.[17,18] The diagnoses of the cases were: pyelonephritis, community acquired pneumonia, liver cirrhosis, haemolytic anaemia, aortic dissection and nephrotic syndrome. Based on diagnostic accuracy scores obtained in these previous studies, the first three cases were considered easy and the last three, difficult.

A questionnaire was developed to measure the students' SI and AKG by using two previously validated instruments. To evaluate SI, we used a four-item instrument developed by Rotgans and Schmidt,[6] which has shown good construct validity, established by means of confirmatory factor analysis, and high reliability (Hancocks's coeficient 0,90). It consists of items such as "I enjoyed working on this topic" and "I was fully focused on this topic". AKG was measured by using a nine-item instrument developed by Glogger-Frey et al[16] which demonstrated high construct validity and reliability (Cronbachs's alfa 0,89) and consists of items such as "My knowledge was insufficient to complete this task" and "I am not sure if I have found a proper solution for this task". Both questionnaires use a 5-point Likert scale, from not true at all to very true for me. These instruments were adapted for the present study by translating them to Brazilian Portuguese and making minimal changes in the statements of the items (e.g., referring to a "clinical case" rather than a "topic" or "task"). The adapted items were reviewed by three of the co-authors (L.M.C.R., S.M., A.S.M.), and a pilot was conducted with five 5th-year medical students to check for face validity and readability. The pilot showed no further adjustments in the instruments to be necessary.

Figure 1 shows the experimental procedure. In both conditions, the clinical cases and the questionnaires were presented to students in booklets in the following sequence: 1) a clinical case, 2) a crossword (for the control condition) or the same clinical case plus instructions for reflection (for the reflection condition), 3) the questionnaire for SI and AKG. This sequence of pages/tasks was repeated for each clinical case. Each booklet contained the same 6 clinical cases (3 difficult; 3 easy), alternating easy and difficult cases. Two versions of the booklets were created, inverting the order in which the cases were presented to control for order effect.

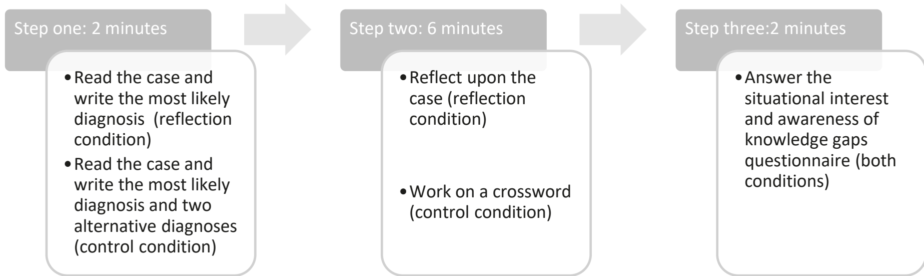


Figure One: The intervention procedure for each one of the 6 clinical cases.

For each case, the students were asked to follow a set of different steps depending on the experimental condition under which they performed. In the reflection condition, students were requested to: 1) read the case and write down the most likely diagnosis for the case, 2) reflect upon the case by following a structured procedure[17,18] which, briefly, consists of writing the clinical findings that match their first diagnostic hypothesis, those that contradict it, and those that are expected were this first hypothesis true but are not described in the case, write two alternative diagnoses and run the same analysis before making a conclusion, 3) answer the questionnaire about SI and AKG. Students from the control condition were asked to carry out the following sequence of steps: 1) read the case, write down the most likely diagnosis for the case and two alternative diagnoses, 2) work on a crossword containing medical terms not related to the cases, which was used to ensure that the two conditions spent the same amount of time in the task, 3) answer the questionnaire about SI and AKG. The instructions to each experimental condition were presented exclusively in writing in the respective booklet, thereby preventing crossing instructions between students who performed under one or the other condition.

The experiment was conducted in a single room, which was large enough to keep students working individually. After introducing the study, the first author (L.M.C.R.) distributed the booklets, which had been previously organized in random order, to the students, thereby randomly assigning participants to one of the two conditions (experimental and control). The students worked with the same booklet throughout the experiment, with each student performing therefore only under the condition to which he/she was initially assigned. This procedure for randomization was chosen since it was not possible to anticipate which students would volunteer to the study and actually attend the activity.

Time and progress throughout the sequence of steps were controlled. The researcher asked students to only open their booklets after being allowed to do so. After having time to read the instructions, time was allocated as follows: 2 minutes to write the most likely diagnosis (reflection condition) or to write the most likely diagnosis and two alternative diagnosis (control condition); 6 minutes to reflect upon the case (reflection condition) or to solve the crossword (control condition); 2 minutes to answer the questionnaire on SI and AKG, for both conditions (Figure 1). The participants did not receive any information about their diagnostic performance during the experiment.

Data analysis

The diagnoses provided by the students were firstly independently assessed by two board certified internists (A.S.M.; E.M.B.), who were not aware of the experimental condition under which they had been made. They classified each response as correct (scored 1), whenever the core diagnosis was present (e.g. "pneumonia", in the "community acquired pneumonia" case); partially correct (scored 0.5), if the core diagnosis was not present, but a component of it was (e.g. "urinary tract infection" in the "pyelonephritis" case); and incorrect (scored 0), when the response did not fall into any of these categories. The raters agreed in the score attributed in 92% of the responses and resolved discrepancies by discussing them and reaching consensus in a subsequent meeting.

We checked the reliability of the instruments used in the study by computing Cronbach's alfa for each scale.

The main outcome measurements of the study were the mean scores of SI and AKG. These scores were computed by averaging the responses provided for the items in the two scales for each participant and, subsequently, for each experimental condition. Mean diagnostic accuracy scores were computed through a similar procedure. Two separate mixed ANOVAs with experimental condition (reflection or control) as between-subjects factor and case difficulty as within-subjects factor (difficult and easy) were performed on the mean scores of SI and AKG.

RESULTS

Forty-three out of the 54 invited students accepted to participate in the study. One outlier (z-scores above 3.29 in the main outcome measurements) was removed from the experimental group after the exploratory data analysis. This led to 42 participants, 21 in each group condition, of which 30 (70%) were female (14 and 16 students, respectively, in experimental and control group), reflecting the predominance of female students in the school (60%). Thirty-five students were between 20 and 25 years-old (17 and 18 students, respectively, in experimental and control group).

The adapted instruments used to measure SI and AKG showed acceptable level of reliability, with a Cronbach's alfa, respectively, of 0,75 and 0,90.

The mean scores of SI reported by students from the two experimental conditions are displayed in Table 1. There was a significant main effect of experimental condition, $F(1, 40) = 10.27$, $p = .003$, $d=0.92$), with students who reflected upon the cases reporting higher SI than those who made differential diagnosis both in easy and difficult cases. The main effect of case difficulty, $F(1, 40) = 9.52$, $p = .004$, $d=0.28$ was also significant, with SI showing to be higher in easy than in difficult cases regardless of the experimental condition. The effect sizes, as measured by Cohen's *d* were large and small, respectively, for experimental condition and case difficulty (considering *d* values of 0.2 for small, 0.5 for medium and 0.8 for large effect sizes).[19] No significant interaction emerged between experimental condition and case difficulty, $F(1, 40) = 0.7$, $p = .79$.

Table 1. Mean scores (range 1 – 5; standard deviation into brackets) of situational interest as a function of experimental condition and case difficulty.

	Reflection group	Control group	Overall
Easy cases	4.19 (0.55)	3.73 (0.48)	3.96 (0.56) [b]
Difficult cases	4.05 (0.51)	3.56 (0.48)	3.80 (0.55) [b]
Overall	4.10 (0.50) [a]	3.65 (0.48) [a]	

[a] *Significant main effect of experimental condition (p = .003, d=0.92);* [b] *Significant main effect of difficulty (p = .004, d=0.28).*

Table 2 presents the mean scores of AKG reported by students from the two experimental conditions. Students who reflected upon the cases reported higher scores of AKG than those who made differential diagnosis, but this difference was not significant, $F(1, 40) = 1{,}17$ $p = .286$. There was a significant main effect of case difficulty, $F(1, 40) = 22{,}75$, $p < .001$, $d = 0.67$, with AKG showing to be higher in difficult than in easy cases regardless of the experimental condition. No significant interaction emerged between experimental condition and case difficulty, $F(1, 40) = 1{,}17$, $p = .367$.

Table 2. Mean scores (range 1 – 5; standard deviation into brackets) on awareness of knowledge gaps as a function of experimental condition and case difficulty.

	Reflection group	Control group	Overall
Easy cases	3.70 (0.55)	3.61 (0.40)	3.66 (0.53) [a]
Difficult cases	4.10 (0.44)	3.89 (0.48)	3.99 (0.46) [a]
Overall	3.82 (0.61)	3.75 (0.40)	

[a] *Significant main effect of case difficulty (p < .001, d =0.67)*

The average diagnostic accuracy showed by students from the two experimental conditions are displayed in Table 3. The groups did not significantly differ in diagnostic accuracy, $F(1, 40) = 1.14$, $p = .291$. There was a highly significant difference between easy and difficult cases, $F(1, 40) = 477.17$, $p < .001$, $d = 4.52$, validating our a priori categorization of cases difficulty level.

No significant interaction effect between experimental condition and case difficulty emerged, showing that both groups performed equally well, both in difficult and easy cases, $F(1, 40) = 1.51$, $p = .23$.

Table 3. Diagnostic accuracy (range 0 – 1; standard deviation into brackets) as a function of experimental condition and case difficulty.

	Reflection group	**Control group**	**Overall**
Easy cases	0,74 (0,17)	0,82 (0,19)	0,78 (0,18) [a]
Difficult cases	0,07 (0,14)	0,07 (0,14)	0,07 (0,13) [a]

[a] *Significant difference between easy and difficult cases (p < .001, d =4.52).*

DISCUSSION

The purpose of this study was to investigate whether structured reflection on to-be-diagnosed clinical cases would influence medical students' SI and AKG, and whether an eventual effect of reflection depends on case difficulty. To that end, we compared SI and AKG scores of students who engaged on structured reflection while diagnosing easy and difficult to diagnose clinical cases with those of students who only gave differential diagnosis. Regardless of case difficulty, structured reflection led to higher scores of SI and AKG, although the difference observed on the latter was not significant. Case difficulty interfered with both SI and AKG in an opposite manner: while the AKG was higher for difficult than easy cases, the SI was higher for easy than difficult cases.

The positive effect of reflection on SI was substantial both in easy and difficult cases. From experiments with reading and learning activities with secondary, psychology and economy students, we have learned that "thirst" for knowledge is triggered, among others, by unexpectedness, suspense and challenging real life problems.[5,6,12,16] The strong positive effect of reflection on students' SI when working with clinical cases that we observed is in agreement with these findings and expands this evidence to the medical field. Justifying the initial impression about the patient's problem, arguing against it and looking at the problem from different perspectives is a more challenging task than giving alternative diagnoses. It is not surprising, therefore, that reflection led to so much more interest. It is surprising, however, that, overall, the SI scores were significantly higher on easy than difficult cases. We expected the opposite effect, since the "thirst" for knowledge is supposed to be triggered by doubt.[12] Two possible explanations for this finding may be raised. First, the difficult cases that we used turned out to be, unexpectedly, extremely difficult to the participants. The average diagnostic accuracy score (0.07 on a 0-1 range) was much lower than previous scores observed with similar audience.[17,18] We raise, therefore, the possibility of an inhibitory potential of extreme difficulty on students' interest to learn. If they perceive the knowledge necessary to solve a problem as much beyond their reach, they may not become interested in studying it further. An *optimal level of incongruity*, above which people would face an unpleasant and disruptive effect on their curiosity, has been mentioned in the literature.[20] Second, working with too difficult cases faced students with many unfamiliar clinical data, which may have led to confusion. Research in reading has found ease of comprehension important to SI.[14] These assumptions are, however, only conjectures and need further investigation.

The findings regarding awareness of knowledge gaps are also only partially in line with our expectations. Structured reflection did not generate significantly higher AKG than just giving differential diagnoses. To further explore this finding, we focused our attention on the questionnaire that we used to measure this variable. It was originally developed for secondary and teacher students[16] and was slightly adapted for this study. We observed that most of the items in the questionnaire seemed to assess not only AKG but also students' frustration while working on the cases ('*this case was too hard to be solved completely*'; '*at certain times, I was stuck with diagnosing this case*'). Responses to this type of question may have also expressed the student's (sometimes negative) feelings about his/her performance. Two of the questions have statements focused on knowledge without referring to diagnostic performance itself and seemed therefore to check awareness of knowledge gaps per se ('*working on this case clarified that I don't know certain things yet*'; '*my knowledge was insufficient to complete this case*'). In order to explore whether different results would come out on these more knowledge-focused questions, we performed a *post hoc* analysis by computing a score of awareness of knowledge gaps including only these two questions. A significant main effect of experimental condition emerged, $F(1, 40) = 4{,}51$ $p = .04$, d=0.34 with students who reflected upon the cases showing higher scores on these two questions than those from the control group, regardless of case difficulty.

Regarding the controversy on the role of AKG as a condition to triggering SI,[12,16] this study suggests that SI in a topic may emerge even if individuals are not consciously aware of their knowledge gaps on it. In the context of this study, it may be a consequence of the difficulty to measure AKG among medical students, as we just discussed. Our findings suggest, however, that SI and AKG may, at least in certain contexts, emerge independent of each other, in line with what has been found previously.[16] Medical education may be one of such contexts, but this issue certainly requires further clarification.

Educational implications

Our findings have implications for medical education. Reflection has been considered a key competence for medical students and practitioners. If reflection upon to-be-solved clinical problems indeed fosters students' interest, it would be therefore a helpful tool in teachers' hands. There is some evidence that structured reflection while practicing with clinical cases fosters learning through restructuring prior knowledge.[17,18] Nevertheless, expertise development also depends on expansion of one's knowledge base. "Thirst" for knowledge about a certain topic must emerge, triggering engagement in learning activities that can quench it. We have shown that a procedure to foster structured reflection upon cases makes students much more interested in the topic than conventionally diagnosing the case. The procedure is relatively easy to apply during practice with clinical cases and may be therefore a useful addition to the toolbox of clinical teachers who are concerned with motivating their students to engage in further study.

Limitations

Our experiment was conducted in a single medical school, with participants engaged in the same year of training. The generalisability of its findings is, therefore, limited. The unexpectedly high difficulty of the clinical cases may have influenced the relationship between difficulty and SI results. This finding raises an interesting question on the *inhibitory potential* of extreme difficulty on students' interest to learn, for which we could not find any available evidence in the literature. When AKG is concerned, it cannot be excluded that the sample size has not provided sufficient power to detect an effect that could have been found with larger sample. Notice, however, that studies analyzing the effects of interventions on AKG have used similar sample sizes with positive results.[12,16] The limited number of items included in our post hoc analysis of the AKG may have compromised its validity. Other experiments, however, have used similar strategies o assess AKG.[12,16] It is also important to highlight that we found reflection to foster SI in clinical cases, but we did not assess if it would foster engagement in learning or contribute to expand students' knowledge. Although there is substantial empirical evidence to support the association between SI and engagement in learning,[4,5,6] such studies were conducted outside the medical education field.

CONCLUSION

In summary, we studied the influence of structured reflection while working with clinical cases and case difficulty on SI and AKG among 4th year medical students. We found reflection to improve SI, an important finding for medical educators, who can engage their students in this relatively simple, short-time consuming process during practice with simulated and real cases. We also found easy cases to trigger more SI than difficult cases, maybe because the difficult cases were beyond an *optimal level of incongruity.* The discrepancies that we observed on AKG and SI scores, also found in a previous study,[16] suggest that much remains to be empirically investigated about the theoretic framework on the role of reflection in learning from experiences of problem-solving. Assessing the outcomes of reflection on perceptions about one's own knowledge and needs and how they are influenced by different factors, such as task difficulty, can contribute to better understand how reflecting upon clinical problems fosters learning.

REFERENCES

1. Dewey J. What is thought? In: Dewey J. *How we think*. Boston: D.C. Heath & Co; 1910. p. 2-22.
2. Schön DA. Understanding the need for artistry in professional education In: Schön DA. *Educating the reflective practitioner: toward a new design for teaching and learning in professions*. San Francisco**:** JBHE Series; 1987. p. 1-20.
3. Sandars J. The use of reflection in medical education: AMEE Guide No. 44. *Med Teach*. 2009; 31(8): 685-695.
4. Hidi S. Interest and its contribution as a mental resource for learning. *Rev Educ Res*. 1990; 60(4): 549-571.
5. Ainley M, Hidi S, Berndorff D. Interest, learning, and the psychological processes that mediate their relationship. *J Educ Psychol*. 2002; 94(3):545-561.
6. Rotgans JI, Schmidt HG. Situational interest and academic achievement in the active-learning classroom. *Learn Instruction*. 2011; 21:58-67.
7. Lockyer J, Gondocz T, Thivierge RL. Knowledge translation: the role and place of practice reflection. *J Contin Educ Health Prof*. 2004; 24:50-56.
8. Mamede S, Schmidt HG. The structure of reflective practice in medicine. *Med Educ*. 2004; 38:1302–1308.
9. Mann K, Gordon J, MacLeod A. Reflection and reflective practice in health professions education: a systematic review. *Adv in Health Sci Educ*. 2009;14:595–621.
10. Eva KW, Regehr G. "I'll never play professional football" and other fallacies of self-assessment. *J Contin Educ Health Prof*. 2008; 28(1): 14-19.
11. Ericsson KA. Acquisition and maintenance of medical expertise: a perspective from the expert-performance approach with deliberate practice. *Acad Med*. 2015; 90(11):1471-86.
12. Rotgans JI, Schmidt HG. Situational interest and learning: Thirst for knowledge. *Learn Instruction*. 2014; 32:37-50.
13. Schraw G, Lehman S. Situational interest: a review of the literature and directions for future research. *Educ Psychol Rev*. 2001; 13(1), 23-52.
14. Schraw G, Flowerday T, Lehman S. Increasing situational interest in the classroom. *Educ Psychol Rev*. 2001; 13(3):211-224.
15. Heidi S, Renninger KA. The four-phase model of Interest Development. *Educ Psychol*. 2010; 41(2):111-127.
16. Glogger-Frey I, Fleischer C, Grüny L, Kappich J, Renkl A. Inventing a solution and studying a worked solution prepare differently for learning from direct instruction. *Learn Instruction*. 2015; 39:72-87.
17. Mamede S, van Gog T, Moura AS, de Faria RMD, Peixoto JM, Rikers RMJP, Schmidt HG. Reflection as a strategy to foster medical students' acquisition of diagnostic competence. *Med Educ*. 2012; 46: 464-472.
18. Mamede S, van Gog T, Moura AS, de Faria RMD, Peixoto JM, Schmidt HG. How can students' diagnostic competence benefit most from practice with clinical cases? The effects of structured reflection on future diagnosis of the same and novel diseases. *Acad Med*. 2014; 89(1): 1-7.
19. Sullivan GM, Feinn R. Using effect size- or why the *p* value is not enough. *J Grad Med Educ*. 2012; 4(3): 279-282.
20. Loewenstein G. The psychology of curiosity: a review and reinterpretation. *Psychol Bull*. 1994;116(1):75-98.

Chapter 3

Effects of deliberate reflection on students' engagement in learning and learning outcomes

Ligia Maria Cayres Ribeiro, Silvia Mamede, Eliza Maria de Brito, Alexandre Sampaio Moura,
Rosa Malena Delbone de Faria, Henk G. Schmidt

Published in: Medical Education 2019;53(4):390-7. DOI: 10.1111/medu.13798

ABSTRACT

Context: Reflection in practice is assumed to enhance interest in knowing more about a topic, increasing engagement in learning and learning outcomes. However, this claim lacks empirical evidence, particularly in medical education. The authors investigated the effects of deliberate reflection upon clinical cases on medical students' engagement in a learning activity and learning outcomes.

Methods: Three-task (diagnostic task; learning activity; test) experiment conducted in August 2017. Seventy-two 4^{th}-year students from UNIFENAS-BH Medical School, Brazil, diagnosed 2 clinical cases with jaundice as the chief complaint either by following a deliberate reflection procedure or making differential diagnosis. Subsequently, all participants received the same study-material about the diagnosis of jaundice. Finally, they took a recall test on the study-material. Outcome measurements: study-time; test scores.

Results: There was a significant effect of experimental condition on students' engagement in the learning activity and on learning outcomes. Students who deliberately reflected upon the cases invested more time in studying the material than those who made differential diagnosis (respectively, mean=254.97, standard deviation=115.45 versus mean=194.96, standard deviation=111.68; p=.02; d=0.53). Deliberate reflection was also related to higher scores in the test relative to differential diagnosis (respectively, mean=22.08, standard deviation=14.94 versus mean=15.75, standard deviation=9.24; p=.03; d=0.51). Medium effect sizes (Cohen's *d)* were observed in both measurements.

Conclusions: Relative to making differential diagnosis, deliberate reflection while diagnosing cases fostered medical students' engagement in learning and increased learning outcomes. Teachers can employ this relatively easy procedure, possibly both with simulated and real scenarios, to motivate their students and help them expand their knowledge, an important requirement for their professional development.

INTRODUCTION

A hiker, in an unfamiliar field, faces a branching. Unsure of what direction to follow, he/she scrutinizes the options and tries to find evidence in favor of one or another direction searching references in memory, exploring sights and sounds, using a compass or even climbing a tree. John Dewey[1] uses this simple example to illustrate reflection: the attentive thought we engage in when facing hesitation, doubt; the deliberate reasoning process of exploring a challenge, from different perspectives, until finding a proper solution for it.[1, 2] The hiker has better chances to find the right way through reflection than by randomly choosing a direction, but only if he/she has relevant knowledge of the context. A compass, for example, will be useless if he/she cannot use it. If this is the case, once (hopefully!) home, he/she might, then, feel motivated to engage in learning activities to acquire this skill, actually engage in learning and, ultimately, master it. This example can be generalized to other contexts, such as medical education. Because of its potential to trigger engagement in learning, reflection, therefore, can be - and indeed has been - considered an important educational strategy.[3, 4] It should be noted, however, that although based on solid theory, there is little empirical evidence that this actually happens, especially in medical education. This article reports on a study that investigated the effects of deliberate reflection while solving clinical cases on medical students' engagement in a follow-up learning activity and learning outcomes.

"Forked-road" situations are common in medical practice because cases are often ambiguous, with patients frequently presenting clinical findings that raise diagnostic uncertainty. In these situations, clinicians may engage in reflection in the midst of the encounter with the patient, while it is still possible to respond to the situation at hand, what Donald Schön called "reflection-in-action".[2] For a medical doctor facing a diagnostic challenge this would involve looking at the patient's problem with different potential explanations in mind, confronting different hypotheses with available clinical data and testing them until setting a diagnostic solution for the patient.[5,6] It is likely that, in the course of this reflective reasoning, the clinicians, especially when they are novice ones, recognize gaps in their knowledge or just realize that they are not so certain about their choices. This might raise interest in knowing more about the problem, which would consequently trigger engagement in learning activities, resulting in increased knowledge about the topic. Indeed, in a previous experiment, deliberate reflection while solving clinical cases by arguing for and against one's diagnostic hypotheses, listing patient's findings that match and refute them, fostered medical students' situational interest on the cases compared to the more conventional approach of giving differential diagnosis.[7] It is reasonable to assume that this deliberate reflection procedure would foster medical students' learning, because situational interest, a transient and context-related interest also described as "thirst" for knowledge, has proved to be a good predictor of engagement in learning and learning outcomes in experiments with audiences outside medical education.[8,9,10,11]

This would be a promising educational effect of deliberate reflection because engagement in learning activities, whether measured as years of schooling, days of instruction or hours of classes, has repeatedly shown to improve learning outcomes.[12] Rather than in long-term activities, a medical student or a physician facing a diagnostic challenge is likely to engage, possibly individually, in short-term learning activities such as reading a text on a selected medical topic. However short, this time may make a difference. Research has shown a positive effect on learning outcomes of engagement even in short learning activities measured in minutes of study-time. For example, an experiment on teaching students' learning of journal's quality assessment either by inventing or studying a worked-solution found study-time to be positively correlated to students' outcomes, even after controlling for prior knowledge and independently of learning strategy.[13] Similar results have been found with secondary school students,[14] and the positive relationship between study-time and learning outcomes is reinforced by research showing that reducing students' time to master a new topic decreased their scores on immediate and late tests.[15] Therefore, if deliberate reflection while practicing the diagnosis of clinical cases indeed fosters engagement in learning activities and learning outcomes, it would be a helpful tool for the development of medical students' clinical knowledge. To our knoweldge, however, there was no empirical evidence that this actually happens.

To address this question, we conducted an experiment with 4th-year medical students, who solved clinical cases either deliberately reflecting upon them or giving differential diagnosis. Although some degree of reflection is expected while students think of alternative diagnoses for a clinical case, the deliberate reflection procedure used in this experiment demanded comparing and contrasting clinical data with different diagnoses, searching for evidence supporting and refuting each one in a systematic way. This structured process is expected to raise more uncertainty than the more conventional approach of generating alternative diagnoses. Subsequently, an appropriate learning task and a test were administered. Engagement in the learning activity, measured as study time, and learning outcomes, measured as scores on a test, were obtained. We hypothesized that: (1) deliberate reflection would foster engagement in learning and (2) deliberate reflection would foster learning, relative to providing a differential diagnosis.

METHODS

Design

The study was an experiment with random allocation of participants to either an experimental or a control condition. The experiment consisted of a single-session divided in three parts - a diagnostic task, a study task, and a recall task. In the diagnostic task, participants diagnosed two clinical cases by following either a structured procedure to reflect upon the cases (experimental) or to make differential diagnosis (control). After diagnosing the cases, participants from both conditions were presented to the same study-material about the differential diagnosis of jaundice. Subsequently, they performed a cued-recall task about the material that they had just studied.

Setting and participants

All 123 4th-year medical students at José do Rosário Vellano University (UNIFENAS), in Belo Horizonte, Brazil, were invited to voluntarily participate in the study. UNIFENAS has a six-year problem-based curriculum, with the two final years dedicated to clerkships. We selected 4th year students because at this point in their training they have been exposed to knowledge about patients with jaundice during tutorial groups and lectures but have limited clinical experience with them.

The students who volunteered for the study were recruited as participants and gave written consent. The experiment was run as an extra-class activity in six sessions offered in different days to accommodate to the students' timetable. A lottery with an electronic tablet as prize was offered to the students as acknowledgement for their participation.

Materials and procedure

A computer-based exercise, consisting of two to-be-diagnosed clinical cases, a study-material and a cued-recall task, was created to this study using Qualtrics.

The cases had jaundice as the main clinical finding and consisted of a written description of clinical symptoms, physical examination and laboratory tests findings. Each case had a most likely diagnosis that had been validated in previous studies.[16, 17] The diagnoses of the cases were acute viral hepatitis and choledocholithiasis.

The study-material consisted of an illustrated text presenting a brief review of bilirubin physiology and physiopathology followed by the presentation of the clinical cases participants had diagnosed, with the key clinical findings valuable to differentiate between the causes of jaundice highlighted and linked to boxes with their interpretation and explanation (e.g. "*Coluria is an indicator of cholestasis because only free direct bilirubin is excreted in urine").*

To measure students' learning of the study-material, a cued-recall task with eight open-ended questions, addressing topics on clinical history, physical examination and laboratory tests concerning the differential diagnosis of jaundice, was created. Each item cued recall of a specific part of the material and, for that part, requested the participants to write down all the information that they remembered from the text that they had just read. An example of a cued question is "*List all the relevant physical examination findings to the evaluation of patients with jaundice and explain how they help on the differential diagnosis".* A pilot of the whole exercise was previously conducted with 15 4th-year medical students, non-participants in this study, to check for understandability and resulted in few slight adjustments.

At the end of the exercise participants were asked to answer questions on gender, age and an estimation of their previous experience with jaundice, measured by how many real patients with jaundice they recalled having assisted.

In both conditions, the exercise was presented to students on a computer screen in the following sequence: 1) a clinical case followed by different instructions to work on the case, depending on the experimental condition to which they were assigned (see below); 2) the study-material, 3) the cued-recall test, and 4) the personal data and jaundice experience questionnaire.

For each case, the students were asked to follow a set of different steps depending on the experimental condition under which they performed. In the deliberate reflection condition, students were requested to: 1) read the case and type down the most likely diagnosis for the case, 2) reflect upon the case by following a structured procedure[6] which, briefly, consists of typing the clinical findings that are in line with their first diagnostic hypothesis, those that contradict it, and those that are expected were this first hypothesis true but are not described in the case, type two alternative diagnoses and run the same analysis for each diagnosis before making a conclusion. Students from the control condition were asked to carry out the following sequence of steps: 1) read the case, type down the most likely diagnosis for the case and two alternative diagnoses, 2) work on a crossword containing medical terms not related to the cases, which was used to ensure similar time on task in the two conditions while minimizing the control condition engagement in reflection, 3) type down their conclusion on the most likely diagnosis for the case. As working on a crossword after diagnosing a case is unusual to students, which could hinder compliance, we stated in the instructions to the control condition that, although it could seem irrelevant, a task like it might help thinking process and should, therefore, be taken seriously. The instructions to each experimental condition were presented exclusively on screen, thereby preventing crossing instructions between students who performed under one or the other condition.

The experiment was conducted in a computer lab with partitioned seats keeping students working individually. Two different electronic addresses were created in Qualtrics, each one forwarding to the experimental or the control condition exercise. After a brief introduction to the study, printed electronic addresses, which had been previously organized in random order, were distributed to the students, thereby randomly assigning participants to one of the two conditions (reflection or control). The students worked on the same exercise throughout the experiment, with each student performing therefore only under the condition to which he/she was initially assigned. This procedure for randomization was chosen since it was not possible to anticipate which students would volunteer to the study and actually attend the activity. A teacher was present in all sessions both to support the participants and to inhibit students from consulting resources other than the exercise.

Time and progress throughout the sequence of steps were computer-controlled. After the instructions, time was allocated as follows: 2 minutes to type the most likely diagnosis (reflection condition) or to type the most likely diagnosis and two alternative diagnosis (control condition); at least 4 up to 6 minutes to reflect upon the case (reflection condition) or to solve the crossword (control condition). There was no fixed time to work on the study-material or on the cued-recall task. The software automatically recorded time spent on each task and responses for each participant. The participants could neither skip steps nor move backward on the exercise but were free to progress in their own pace while studying the material and carrying out the recall task. They did not receive any information about their diagnostic performance while working on the cases. The correct diagnoses were explained in the study-material.

Data analysis

The diagnoses provided by the students in the diagnostic task were firstly independently assessed by two board certified internists (L.M.C.R.; E.M.B.), who were not aware of the experimental condition under which they had been made. They classified each response as correct (scored 1), whenever the core diagnosis was present (e.g. "hepatitis", in the "acute viral hepatitis" case); partially correct (scored 0.5), if the core diagnosis was not present, but a component of it was (e.g. "gallstones" in the "choledocholithiasis" case); and incorrect (scored 0), when the response did not fall into any of these categories. The raters agreed in the score attributed in 87% of the responses and resolved discrepancies by discussing them and reaching consensus in a subsequent meeting. The accuracy of initial diagnoses was measured to check whether the two experimental groups were similarly acquainted with the problem under study.

The amount of information reported in the participants' responses (recall task) was evaluated by counting the number of idea units present in each response.[18,19] We considered as idea units text fragments, such as a word or a short sentence, meaningful to the task. Since we were interested in measuring learning of the information presented in the study-material and used a cued-recall task, the idea units counted on participants' responses had to meet three criteria: to consist of correct information, to be actually present in the study-material and to be consistent with the question cue. Each idea unit that met these criteria was scored with one point. Two authors (L.M.C.R., E.M.B.) firstly discussed and reached a consensus on the idea units present in the study-material. One example of idea unit, consistent with the cue "*Explain how urine and feces' color help differentiating the causes of jaundice*" is "*coluria indicates cholestasis*". Subsequently, the same authors independently assessed 10% of participants' tasks, reaching an initial agreement of 86%. After discrepancies were resolved in further discussions, the first author assessed all participants responses. Only manifest content was considered: no possible underlying meanings were counted.[20]

The main outcome measurements of the study were the time students spent on the study-material, which has been taken as a measure of learning process, and their scores on the cued-recall task, assumed as a measure of learning outcomes. The mean study-time was computed for each experimental condition. The scores on the cued-recall task were computed by summing the scores of each participant and, subsequently, averaging them for each experimental condition. Mean diagnostic accuracy scores were computed through a similar procedure. Two separate ANOVAs with experimental condition (reflection or control) as between-subjects factor were performed on the mean study-time and on the mean cued-recall task scores. T-tests were performed to compare age and diagnostic accuracy and Pearson Chi-square tests to compare gender and previous experience with real patients with jaundice between experimental and control groups.

RESULTS

Seventy-four (60%) out of the 123 invited students participated in the study. One outlier was removed from the control group after the exploratory data analysis. One participant was removed from the reflection group for accessing text reference other than the experiment material during the exercise, which was not allowed. This led to 72 participants, 36 in each group condition.

Table 1 presents age, gender, previous experience with patients with jaundice and initial diagnostic accuracy on the clinical cases as a function of experimental condition. No significant differences emerged in age, $t(54.11) = 1.35$, $p = .18$ or gender, $X^2(1) = .53$, $p = .63$. Previous experience with patients with jaundice, $X^2(1) = 0$, $p > .99$, and the initial diagnostic accuracy scores on the clinical cases, $t(70) = .50$, $p = .61$, also did not differ between groups, indicating that they were comparable.

The mean study-time and cued-recall task scores for the two experimental conditions are displayed in Table 2. There was a significant main effect of experimental condition on study-time, $F(1, 70) = 5.03$, $p = .02$, $d = .53$, with students who deliberately reflected upon the cases engaging longer in the subsequent learning activity than those who had given differential diagnoses. There was also a significant main effect of experimental condition on learning outcome, $F(1, 70) = 4.68$, $p = .03$, $d = .51$, with students who deliberately reflected upon the cases showing higher scores on the cued-recall task than those who gave differential diagnoses. The effect sizes, as measured by Cohen's *d*, were medium for both outcomes (considering *d* values of 0.2 for small, 0.5 for medium and 0.8 for large effect sizes).[21]

Table 1: Characteristics of the participants in the experimental and control groups (standard deviations into brackets).

	Reflection group	Control group	Overall
Age (mean)	22.81 (1.83)	23.67 (3.36)	23.24(2.72)
Gender			
Male	12 (33%)	15 (42%)	27
Female	24 (77%)	21 (68%)	45
Previous experience with patients with jaundice			
Yes	21 (58%)	21 (58%)	42
No	15 (42%)	15 (42%)	30
Initial diagnostic accuracy scores on clinical cases (range 1-2)	1.29 (0.74)	1.37 (0.66)	1.33 (0.69)

Table 2: Mean study-time (seconds) in the learning activity and cued-recall scores in the test (total number of information units) as a function of experimental condition (standard deviation into brackets)

	Reflection group	Control group	Overall
Study-time	254.97 (115.45)	194.96 (111.68)	224.96 (116.76)[a]
Cued-recall score	22.08 (14.94)	15.75 (9.24)	18.92 (12.74)[b]

[a] *Significant main effect of experimental condition (p=0.02, d=0.53)*
[b] *Significant main effect of experimental condition (p=0.03, d=0.51)*

DISCUSSION

In this study we investigated the effect of deliberate reflection while solving clinical cases on medical students' engagement in a learning activity and learning outcomes. To that end, we compared study-time and test scores of students who worked with two clinical cases through deliberate reflection with those who made differential diagnosis. Students who worked with reflection engaged 31% longer in the learning activity and attained a 40% higher score on the cued-recall test. The difference in favor of the reflection condition was significant in both measurements, with a medium effect-size.

These results are in line with our expectations: students who deliberately reflected upon to-be-diagnosed clinical cases had a more challenging task than those who worked with differential diagnosis. It is reasonable to expect that students who are requested to provide alternative diagnoses for a case engage in some degree of reflection to perform the task. The deliberate reflection, however, requires students to compare and contrast the alternative diagnoses, searching for evidence supporting and refuting each one in a systematic way. This process might raise uncertainty and facilitate recognition of knowledge gaps, which has already shown to foster interest and engagement in learning within students outside medical education.[8,9] It is not surprising, therefore, that deliberate reflection led to longer engagement in the learning activity. This positive effect of deliberate reflection on actual engagement in studying the learning material is also consistent with previous findings showing that deliberate reflection fostered medical students' interest in knowing more about the cases.[7]

As deliberate reflection increased both study-time and test scores, the latter may have been a consequence of the former. By leading to investment of more time in studying the learning material, deliberate reflection increased learning outcomes. This assumption is consistent with the positive relationship between the amount of study-time and learning results observed in experiments in which students worked individually,[13,14] as our participants did, such as in research on teaching journal's quality assessment. Engagement in learning, after all, has shown to be a consistent and important mediator of learning, even if the time-engagement is short, a matter of minutes,[12] such as what is expected to happen when physicians or medical students are confronted with an uncertain diagnosis. However, another possible explanation for our findings has to be considered: deliberate reflection may have mobilized students' prior knowledge, and, once activated, it would make it easier to understand the new information presented in the study material and to integrate it into existing knowledge structures.[22,23] It could be also an interaction between these two effects, but exploring the mechanisms underlying the effects of reflection was not within the scope of our study and requires further research.

Educational implications

Our findings add empirical support to the importance of deliberate reflection as a key competence for medical students and practitioners. It demonstrates the potential of deliberate reflection upon clinical cases, a simple, short-time consuming process of confronting diagnostic hypotheses with patients' clinical data, as a motivational force for students' engagement in learning activities. Clinical teachers could use it as a strategy to motivate their students to study medical topics related to clinical cases. It also expands the evidence of engagement in learning as an important mediator of learning outcome to the medical domain. To the best of our knowledge, our findings provide the first empirical evidence of the effects of deliberate reflection on study engagement in learning and learning outcomes in medical education.

Limitations

Our study was run in a single medical school, with participants of the same year of training and that worked with a single clinical topic, jaundice, which limits the generalizability of its findings. We measured students' engagement by time on task because it is a variable known to influence learning outcomes,[12] but there could be other dimensions, such as intellectual engagement, that we did not explore. Also, as we used participants' immediate post-test scores as a measure of learning outcome, we cannot foresee if the positive effect of deliberate reflection that we observed on learning would last longer. Finally, the task of solving crosswords in the control condition may be seen as potentially harmful for the learners, as the crosswords could have distracted them from the diagnostic task. However, because students in the control group provided alternative diagnoses for the case *before* rather than *after* solving the crossword, they would not be affected by the crossword task *while* generating the alternative diagnoses. The crossword may eventually have influenced the accuracy of their final diagnosis, but this was not an outcome measurement of the study. Our intention was to compare the effect of deliberate reflection and providing alternative diagnoses on engagement in a subsequent study task. Nevertheless, it cannot be excluded that solving the crossword generated excessive cognitive load which somehow carried over to the study task though it is worth noticing that the deliberate reflection procedure has also been shown to involve high cognitive load.[24]

CONCLUSION

In summary, we studied the influence of deliberate reflection while working with to-be-diagnosed clinical cases on engagement in learning and learning outcomes among 4th year medical students. We found positive effects of deliberate reflection on both, which adds evidence to sustain it as an important competence for medical students. It also expands the evidence on the potential benefits of a deliberate refection procedure that can be easily used by clinical teachers to motivate their students, possibly both on real and simulated clinical environments. Nevertheless, there are relevant questions still opened. How deliberate reflection fostered learning: through expansion or reorganization of knowledge? Would the deliberate refection's effect observed be similar in different context, for example, with more/less experienced students? The answers can contribute to better practices in medical schools. Paraphrasing Dewey, it is necessary to climb the field's tree, survey additional facts and see how these things relate to one another.[1]

REFERENCES

1. Dewey J. What is thought? In: Dewey J, editor. *How We Think*. Boston: D. C. Heath and Co; 1910. P2-22.
2. Schön DA. Educating the Reflective Practitioner: *Toward a New Design for Teaching and Learning in Professions*. San Francisco: Jossey-Bass Publishers; 1987.
3. Sandars J. The use of reflection in medical education: AMEE Guide No. 44. *Med Teach*. 2009 Aug; 31(8): 685-695.
4. Ericsson KA. Acquisition and maintenance of medical expertise: a perspective from the expert-performance approach with deliberate practice. *Acad Med*. 2015 Nov; 90(11): 1471-1486.
5. Mamede S, Schmidt HG. Correlates of reflective practice in medicine. *Adv Health Sci Educ*. 2005 Nov; 10(4): 327-337.
6. Mamede S, Schmidt HG, Penaforte JC. Effects of reflective practice on the accuracy of medical diagnoses. *Med Educ*. 2008 May; 42 (5): 468-475.
7. Ribeiro LMC, Mamede S, Moura AS, de Brito EM, de Faria RMD, Schmidt HG. Effect of reflection on medical students' situational interest: an experimental study. *Med Educ*. 2018 May; 52(5): 488-496.
8. Rotgans JI, Schmidt HG. Situational interest and learning: Thirst for knowledge. *Learn Instruc*. 2014 Aug; 32: 37-50.
9. Rotgans JI, Schmidt HG. Situational interest and academic achievement in the active-learning classroom. *Learn Instruc*. 2011 Feb; 21(1): 58-67.
10. Linnenbrink-Garcia LP, Patall EA, Messersmith EE. Antecedents and consequences of situational interest. *B J Educ Psychol*. 2013 Dec; 83 p. 591-614.
11. Hidi S. Interest and its contribution as a mental resource for learning. *Rev Educ Res*. 1990;60(4): 549-571.
12. Fredrick WC, Walberg HJ. Learning as a function of time. *J Educ Res*. 1980; 73(4): 183-194.
13. Glogger-Frey I, Fleischer C, Gruny L, Kappich J, Renkl A. Inventing a solution and studying a worked solution prepare differently for learning from direct instruction. *Learn Instruc*. 2015; 39: 72-87.
14. Ainley M, Hidi S, Berndorff, D. Interest, learning and the psychological processes that mediate their relationship. *J Educ Psych*. 2002 Sep; 94(3): 545-561.
15. Gettinger M. Time allocated and time spent relative to time needed for learning as determinants of achievement. *J Educ Psych*. 1985 Feb; 77(1): 3-11.
16. Mamede S, van Gog T, Moura AS, de Faria RMD, Peixoto JM, Rikers RM, Schmidt HG. Reflection as a strategy to foster medical student's acquisition of diagnostic competence. *Med Educ*. 2012 May; 46(5): 464-472.
17. Mamede S, van Gog T, Moura AS, de Faria RMD, Peixoto JM, Schmidt HG. How can students' diagnostic competence benefit most form practice with clinical cases? The effects of structured reflection on future diagnosis of the same and novel deseases. *Acad Med*. 2014 Jan; 89(1): 1-7.
18. Meyer BJF. The organization of prose and its effects on memory. In: van Dijk TA, Hendricks WO. *North-Holland studies in theoretical poetics vol 1*. Amsterdam/Oxford: North-Holland Publishing Co. New York: American Elsevier Publishing Co.; 1975. p 249.
19. Schiefele U, Krapp A. Topic interest and free recall of expository text. *Learn Indiv Diff*. 1996; 8(2):141-160.
20. Fraenkel JR, Wallen NE, Hyun H. Content analysis. In: . Fraenkel JR, Wallen NE, Hyun H, editors. *How to design and evaluate research in education*. 8^{th} ed. New York: McGraw-Hill; 2012. p. 477-504.
21. Sullivan GM, Feinn R. Using effect size-or why the p value is not enough. *J Grad Med Educ*. 2012 Sep; 4(3): 279-282.

22. Mayer RE. Information Processing. In: Harris KR, Graham S, Urdan T, editors. APA *Educational Psychology Handbook: Vol. 1. Theories, Constructs, and Critical Issues.* 1st ed. Washington, DC: American Psychology Association; 2012. p 85-99.
23. Schmidt HG, Devolder ML, Degrave WS, Moust JHC, Patel VL. Explanatory models in the processing of science text - the role of prior knowledge activation through small-group discussion. *J Educ Psych*. 1989; 81(4): 610-619.
24. Ibiapina C, Mamede S, Moura A, Elói-Santos S, van Gog T. Effects of free, cued and modelled reflection on medical students' diagnostic competence. *Med Educ*. 2014 Aug; 48: 796-805.

Chapter 4

Exploring mechanisms underlying learning from deliberate reflection: An experimental study

Ligia Maria Cayres Ribeiro, Silvia Mamede, Eliza Maria de Brito, Alexandre Sampaio Moura,
Rosa Malena Delbone de Faria, Henk G. Schmidt

Medical Education 2021; 55 (3): 404-12. DOI: 10.1111/medu.14410

ABSTRACT

Introduction: Previous research suggests that, relative to generating a differential diagnosis, deliberate reflection during practice with clinical cases fosters learning from a subsequently studied scientific text and promotes interest in the subject matter. The present experiment aimed to replicate these findings and to examine whether motivational or cognitive mechanisms, or both, underlie the positive effects of reflection.

Methods: 101 5^{th}-year medical students participated in an experiment containing four phases: Students (1) diagnosed two clinical cases of jaundice-related diseases either through deliberate reflection or differential diagnosis; (2) reported their situational interest and awareness of knowledge gaps; (3) studied a text about jaundice, either under free- or restricted-time; (4) and recalled the text. Outcome measures were text-recall, situational interest and awareness of knowledge gaps.

Results: A main effect of diagnostic approach on recall of the text was found, with the reflection group recalling more studied material than the differential diagnosis group (means: 72.56 *vs*. 58.80; $p = .01$). No interaction between diagnostic approach and study-time (free or restricted) emerged, nor was there a main effect of the latter. Relative to the differential diagnosis group, students who reflected upon the cases scored significantly higher on both situational interest (means: 4.45 *vs*. 3.99, $p < .001$) and awareness of knowledge gaps (means: 4.13 *vs*. 3.85, $p < .01$).

Discussion: Relative to generating differential diagnoses, reflection upon clinical cases increased learning outcomes on a subsequent study-task, an effect that was independent of study-time, suggesting that cognitive mechanisms underlie this effect, rather than increases in motivation to study. However, higher scores on situational interest and awareness of knowledge gaps and a tendency towards larger gains when time was free suggest that higher motivation may also contribute to learning from reflection.

INTRODUCTION

Reflection can be defined as the processes of questioning one's own assumptions and their consequences for decision-making. It is usually triggered, and is particularly useful, when we face challenging situations that raise uncertainty.[1] Medical education places much value in developing medical students' ability to engage in reflection because as professionals they will manage ill-defined problems, such as clinical cases for which a clear diagnosis is not available. Indeed, a series of studies have shown that engaging in deliberate reflective processes can improve the accuracy of diagnostic reasoning, as detailed below. Besides being helpful in professional practice, however, reflection is also considered important as a learning tool. It is expected to foster students' engagement in learning activities,[2,3] and there is some empirical evidence that reflecting upon to-be-diagnosed clinical cases can foster medical students' learning from subsequent study activities relevant to those cases.[4,5] However, this evidence is still preliminary and the mechanisms that underlie such potential positive effects of reflection are not yet understood. The experiment reported in this article examined whether the previously observed positive effect of reflection on medical students' learning replicates and explored two possible mechanisms that may underlie this effect: motivating students to engage in learning activities and fostering cognitive processes that facilitate learning.

The aforementioned studies on the effects of reflection on students' learning[4,5] employed the "deliberate reflection procedure", originally developed as a tool to improve physicians' performance.[6] The procedure involves reasoning through clinical problems by generating a tentative diagnostic hypothesis, confronting this hypothesis with the patient's clinical findings, considering alternative explanations for the problem, and arguing for and against each emerging explanation before making a final decision. Besides helping physicians solve difficult clinical cases,[6] the deliberate reflection procedure has been used as a learning tool. Mamede *et al.*[7,8] observed that students who engaged in deliberate reflection during practice with clinical cases provided better diagnoses of new cases of the same (or related) diseases in the future than their peers who had used more conventional approaches. The results were attributed to a refinement of students' mental representation of the diseases, since no additional knowledge was offered to the students in these experiments.

Ribeiro *et al.*[4,5] used the same deliberate reflection procedure in two experiments with medical students, exploring the effects of reflection on learning of *new* material (as opposed to refinement of previously learned knowledge). Specifically, they studied the potential of reflection to foster students' engagement with, and learning from, a text that provided detailed explanations about the cases. In the first experiment, fourth-year medical students diagnosed clinical cases either through the deliberate reflection procedure or through generating differential diagnoses. Scores on students' situational interest (SI), as measured by a self-reported questionnaire,[9] were obtained. SI is a form of transitory interest also labeled "thirst" for knowledge.[10] SI is expected to possibly derive from the awareness of knowledge gaps (AKG) that emerges when we try to solve challenging problems; the feeling of being unable to make sense of things can trigger a need to close those gaps. Indeed, SI has been shown to be a good predictor of engagement in learning activities.[9-12] The experiment showed deliberate reflection to foster students' SI relative to giving differential diagnosis, but did not evaluate whether students actually engaged in more learning activities related to the cases. A follow-up experiment showed that, relative to differential diagnosis, deliberate reflection fostered engagement with a subsequently studied relevant text (as measured by the amount of time spent processing that text) and higher cued recall from this text.[5] These findings suggest that deliberate reflection is an effective learning tool able to foster engagement in learning activities and, ultimately, learning. The experiments, however, did not investigate the psychological mechanisms underlying the positive effect of reflection on students' learning. Since time spent in learning activities can influence learning outcomes,[13] the observed benefits to learning could have arisen from greater motivation leading to longer engagement with the text. That means, deliberate reflection would foster learning simply by increasing time invested in studying the material. Reflection could, however, also have affected the nature of the information processing thereby directly influencing students' cognition (i.e., the knowledge they took away from studying the text). It would be reasonable to expect that reflection would foster cognitive processes such as activation of prior knowledge and/or elaboration on the new information, as it has been observed in other learning strategies involving problem-solving.[14-19]

We conducted an experiment to examine whether increased *motivation,* facilitation of *cognitive processing*, or a combination of both mechanisms underlie the effect of deliberate reflection on learning. Fifth-year medical students diagnosed two clinical cases either through deliberate reflection or by giving differential diagnosis. Subsequently, they performed a study task, reading a relevant text, either under free- or restricted time conditions, and took a cued recall test. Assume for a moment that *only* higher levels of motivation (as expressed in longer engagement with the text) would cause an effect of diagnostic approach on cued recall. Then a main effect of diagnostic approach and an interaction effect between diagnostic approach and time (free vs. restricted) would emerge because under restricted time conditions no difference in cued recall would be observed. In addition, a main effect of time (free vs. restricted) would be expected. However, if the positive effect of deliberate reflection was the result of better cognitive processing *only*, a main effect of diagnostic approach on cued recall would be expected, and, more importantly, *no* interaction with study-time (free vs. restricted) should be seen. The latter would signify that the effect of the treatment is the same, even if study-time is restricted (i.e., even when motivation is obstructed from expressing itself through longer engagement). Third, presence of a main effect of both time and diagnostic approach, and absence of an interaction effect, would suggest that both processes, motivational and cognitive, have been involved. To further explore the possible motivational mechanism underlying reflection, participants' SI and AKG were also measured. We hypothesized that, relative to students who gave differential diagnoses, students who reflected upon the cases would have higher scores on SI and AKG.

METHODS

Design

The experiment consisted of an exercise with four phases that took place sequentially in a single session; (1) a diagnostic phase, (2) measurement of SI and AKG, (3) study of a relevant text and (4) a cued recall test. In the diagnostic phase, participants diagnosed two clinical cases with jaundice as the chief complaint either by following the deliberate reflection procedure or giving differential diagnosis. After diagnosing each case, participants answered questionnaires on SI and AKG. In the subsequent study phase, participants studied the same text about the differential diagnosis of jaundice either with free time or within a fixed maximum amount of time allocated for the task. Finally, all participants performed a cued recall test about the material they had just studied. There were, therefore, four different experimental conditions, with the approach used for the diagnostic phase (deliberate reflection *vs* differential diagnosis) crossed with the time conditions for the study phase (free study-time *vs* restricted study-time). Participants were randomly allocated to one of the four conditions (see Figure 1).

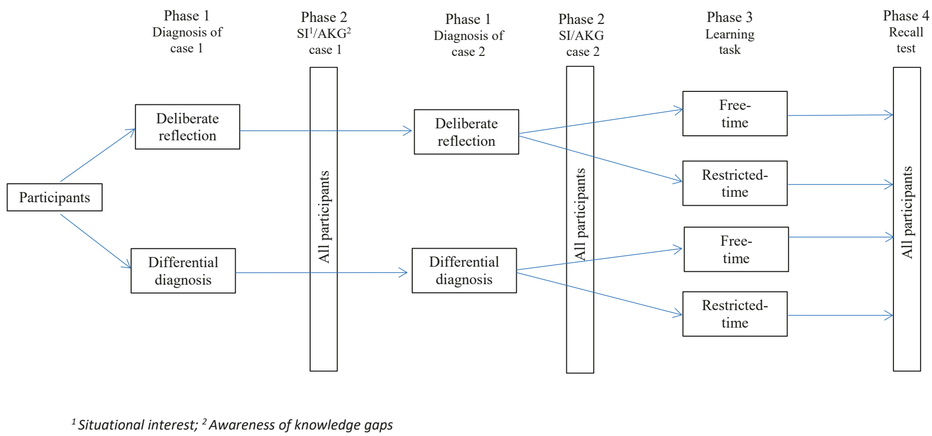


Figure 1: Experiment design and tasks flow

Setting and participants

Three-hundred and fifty fifth-year medical students at the Federal University of Minas Gerais (UFMG), in Belo Horizonte, Brazil, were invited to volunteer for the experiment between July

2018 and July 2019. This school has a six-year curriculum with the two final years dedicated to clerkships. We recruited these students because at this time of their training they have been exposed to knowledge about jaundice but have limited clinical experience with it. The experiment was an extracurricular activity carried out in different sessions to accommodate students' timetables. Participants gave their written consent and were offered the possibility to discuss the experiment with the authors after completing it.

Ethical approval for the study was provided by UFMG Research Ethics Committee.

Materials

All materials were presented to participants via computer using the *Qualtrics* suite. To begin, participants were asked to rate their confidence in diagnosing patients with jaundice using a five-point scale (*very low confidence* to *very high confidence*) and to estimate how many real patients with jaundice they recalled having encountered (either 0, 1-3, >3).

Phase 1 used two clinical cases that were employed in previous studies[4,5]. Both had jaundice as the main clinical finding and consisted of written descriptions of clinical symptoms, physical examination and laboratory tests. Each had a confirmed most likely diagnosis: acute viral hepatitis or choledocholithiasis. [7,8] We selected jaundice because it allowed for an exercise that covered relevant clinical information in a condensed time, which usually fosters participation and compliance with instructions.

Phase 2 used questionnaires on SI and AKG that had similarly been used in previous research.[4] Rotgans and Schmidt[9,10,20, 21] developed the SI questionnaire, which was shown to have acceptable reliability when administered to medical students at a similar level of training[4]. The SI questionnaire contains six questions like *I was totally focused while working on this task.* The AKG questionnaire was developed by Glogger-Frey *et al*[22] for research on instructional design and has similarly shown good reliability in previous study with medical students.[4] The questionnaire has nine questions such as *working on this task revealed I don´t know certain things yet.* The questionnaires requested participants to answer each question by using five-point Likert scales, ranging from *not true at all* to *very true to me,* and are available in Appendix 2.

In Phase 3, the study material consisted of an illustrated text presenting a review of bilirubin physiopathology followed by the presentation of the clinical cases participants had just diagnosed, with the key clinical findings valuable to differentiate between the causes of jaundice highlighted and linked to boxes with their interpretation and explanation (e.g. "*Coluria indicates cholestasis because free direct bilirubin is excreted in urine").*

Finally, Phase 4 used a cued recall test with eight open-ended questions to measure students' learning of the study-material. The test, previously used in an experiment with similar participants,[5] addressed topics on clinical history, physical examination and laboratory tests concerning the differential diagnosis of jaundice. Each item cued recall of a specific part of the material and requested the participants to write down information from the text they had just read. For example, "*List all the relevant physical examination findings to the evaluation of patients with jaundice and explain how they help on the differential diagnosis*". The cued recall test is available in Appendix 3.

Procedures

The experimental interventions were presented during Phase 1 (manipulation of the diagnostic instructions) and Phase 3 (time allowed for study).

For Phase 1, students in the deliberate reflection condition were requested to read the case and follow a procedure to reflect upon the case[6] which, briefly requires 1) typing the most likely diagnosis for the case, 2) typing the clinical findings that are in line with their first diagnostic hypothesis; findings that contradict it; and those that are expected (if this first hypothesis were true) that are not described in the case; 3) typing two alternative diagnoses and the same findings for each diagnosis; 4) drawing a conclusion on the most likely diagnosis for the case. Students in the differential diagnosis condition were asked to carry out the following sequence of steps: 1) read the case, type the most likely diagnosis and two alternative diagnoses, 2) work on a crossword containing medical terms not related to the cases (to ensure similar time spent in the two conditions while minimizing this group's engagement in reflection), 3) typing their conclusion on the most likely diagnosis for the case. The instructions stated that, although it could seem irrelevant to complete a crossword, a task like it might help decision-making and should, therefore, be taken seriously. For all groups, time to progress throughout the sequence of steps in the diagnostic phase was controlled by Qualtrics as follows: 2 minutes to type the most likely diagnosis/diagnoses; 7 minutes to reflect upon the case (reflection condition) or to solve the crossword and type the final diagnosis (control condition).

After diagnosing the first clinical case, the computer moved participants to Phase 2 (the SI and AKG questionnaires). Participants answered the questionnaires at their own pace. Next, the suite presented the second clinical case followed by a second set of SI and AKG questionnaires.

Subsequently, in Phase 3 (the learning task), the study material was presented with a request that students read it. Time was controlled by the software with participants being told whether study-time would be restricted or not. Those for whom study-time was restricted had up to 180 seconds to read the text. This was the median time a similar group of students spent on the same material in a previous experiment in which no restriction of time was imposed.[5]

Participants could move onto Phase 4 in less than 180 seconds if they wanted to. Those for whom time to study was free could read it as long as they wanted.

Upon completion of the learning task, the next screen presented participants with Phase 4, the cued recall test. They were requested to answer the questions by making an effort to write all information they could remember. For the test phase there was no restriction of time.

The experiment was conducted in a computer lab with students working individually. Each student received a copy of a link that would give access to one of the four versions of the Qualtrics experiment. The links had been previously organized in random order to randomly assign participants to one of the four conditions, since it was not possible to anticipate which students would volunteer for the experiment.

Participants could neither skip steps nor move backward on the exercise. The software automatically controlled and recorded time spent on each phase and recorded each participant's responses. No information about their diagnostic performance was provided to participants while they worked on the cases, but the study material of Phase 3 explained the correct diagnoses. One of the authors was present in all sessions to help in case of computer problems and to inhibit students from consulting resources other than those presented as part of the exercise.

Data analysis

The diagnoses provided by the students in the diagnostic task were independently assessed by two board-certified internists (LMCR and EMB), who were blinded to experimental condition. They classified each response as correct (scored 1) whenever the core diagnosis was present (e.g. "hepatitis", in the "acute viral hepatitis" case); partially correct (scored 0.5), if the core diagnosis was not present, but a component of it was (e.g. "biliary colic" in the "choledocholithiasis" case); and incorrect (scored 0) when the response did not fall into any of these categories. The raters agreed in the score attributed in 87% of the responses and resolved discrepancies by reaching consensus in a subsequent meeting. The accuracy of initial diagnoses was measured to check whether the two experimental groups were similarly acquainted with the problem under study by summing, for each participant, scores achieved on the two cases and subsequently averaging within each experimental condition.

The amount of information reported in participants' responses (cued recall task) was evaluated by counting the number of idea units present in each response.[23,24] We considered idea units to be text fragments, such as a word or a short sentence, meaningful to the task. Each idea unit was scored with one point as long as it met three criteria: consisted of correct information, was consistent with the question cue and was actually present in the study-material. For example, for the cue "*Explain how prothrombin activity helps differentiating the causes of jaundice*" one idea unit was "*prothrombin activity depends on liver function*". Two authors (LMCR and EMB), independently assessed 10% of participants' responses, reaching an initial agreement of 84%. After discrepancies were resolved, the first author coded all participant responses. Only manifest content was considered (i.e., no possible underlying meanings were counted).[25] Appendix 4 presents an example of the test scoring procedure.

The main outcome measurements of the study were: Cued recall (as measured by the total number of accurate idea units produced in cued recall), and SI and AKG, taken as measures of engagement in learning. To compute cued recall, we first summed the number of idea units recorded by each participant and then averaged across participants for each experimental condition. To obtain the SI and AKG scores, for each participant, we averaged the ratings provided to the items of the questionnaires, and subsequently computed the mean for each experimental condition. Actual processing times were registered by the software for each participant and we computed means for each experimental condition.

Data were analyzed using SPSS for Macintosh, version 25 (IBM Corp, Armonk, NY). Significance level was set at $p < .05$. To check whether groups were similar in terms of the extraneous variables age, initial diagnostic accuracy, and confidence in diagnosing jaundice, two-way ANOVAs were conducted with diagnostic approach and study-time (free vs. restricted). Pearson *Chi*-square tests were carried out comparing gender and previous experience with patients with jaundice between the four experimental groups. All other variables were analyzed applying ANOVAs: separate one-way tests to compare SI and AKG according to diagnostic approach and a two-way test with diagnostic approach and study-time (free vs. restricted) on cued recall.

RESULTS

Inclusion of participants and outliers

One-hundred and five (30%) of the 350 invited students agreed to participate. Four outliers (z-scores > 2.58 in study-time and/or test-score) were removed after exploratory data analysis; this is a commonly applied threshold as only 1% of values can be expected to be outside this range.[26] This led to 101 participants being included in the analyses.

Analysis of success of randomization

Table 1 presents participants' characteristics as a function of experimental condition. No significant differences between groups emerged in age, $F(1,97) = .88$, $p = .17$, gender, $X^2(3) = 2.74$, $p = .43$, previous experience with patients with jaundice, $X^2(3) = .93$, $p = .82$, confidence in diagnosing jaundice $F(1,97) = .81$, $p = .37$, or initial diagnostic accuracy, $F(1,97) = .18$, $p = .67$.

Table 1: Characteristics of the participants according to the diagnostic approach followed to diagnose the clinical cases in Phase 1 and study-time condition for the learning task in Phase 3 (standard deviation into brackets).

		Free study time		**Restricted study time**		
		Deliberate reflection	Differential diagnosis	Deliberate reflection	Differential diagnosis	Total
N		25	24	27	25	101
Age		23.96 (2.35)	23.5 (2.04)	23.33 (1.8)	24 (2.3)	23.7 (2.12)
Gender	Male	15	14	15	19	63
	Female	10	10	12	6	38
Previous experience	No experience	8	4	5	4	21
	1-3 patients	14	17	20	17	68
	> 3 patients	3	3	2	4	12
Confidence in diagnosis (range 1-5)		2.88 (.73)	2.79 (.78)	2.89 (.80)	3.08 (.81)	2.91 (.78)
Initial diagnostic accuracy (range: 0-2)		1.32 (.78)	1.41 (.71)	1.61 (.56)	1.60 (.47)	1.49 (.64)

Analysis of outcomes

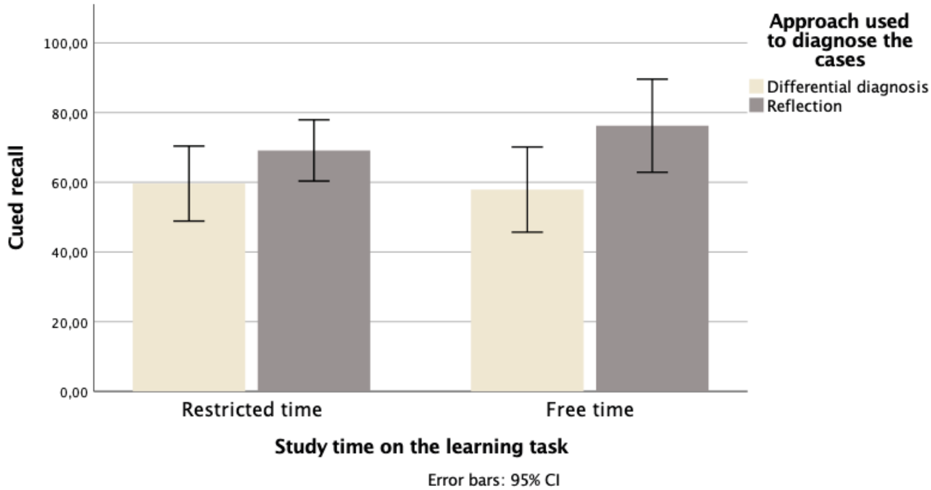
Table 2 presents the main outcome measurements. There was a significant main effect of diagnostic approach on cued recall [means: 72.56 *vs*. 58.80, $F(1, 97) = 6.43$, $p = .01$, $d = .54$], with students who deliberately reflected upon the cases showing higher scores on the test than those who gave differential diagnoses. Importantly, there was no effect of study-time (free *vs*. restricted) on cued recall [$F(1, 97) = .24$, $p = .63$, *ns*]; nor was an interaction effect demonstrated between diagnostic approach and study-time [$F(1, 97) = .65$, $p = .42$, *ns*]. These findings support a cognitive explanation for the effect of deliberate reflection on learning and cued recall, at the expense of a motivational, or a combined cognitive/motivational explanation (see also figure 2).

Table 2: Mean situational interest (SI), awareness of knowledge gaps (AKG) and study-time (seconds) as a function of the diagnostic approach followed to diagnose the clinical cases in Phase 1 and study-time condition for the learning task in Phase 3 (standard deviation into brackets).

	Free study time		**Restricted study time**		
	Deliberate reflection	Differential diagnosis	Deliberate reflection	Differential diagnosis	Total
N	25	24	27	25	101
Cued recall	76.24 (32.39)	57.91 (28.94)	69.14 (22.21)	59.64 (26.07)	65.88 (28.11)
Actual time*	405.09 (222.89)	347.17 (183.60)	168.80 (23.39)	170.02 (20.04)	269.97 (176.55)
SI	4.47 (.44)	4.01 (.49)	4.44 (.44)	3.97 (.52)	4.23 (.51)
AKG	4.24 (.41)	3.86 (.55)	4.03 (.41)	3.84 (.42)	3.99 (.49)

**Actual time needed to process the text, although represented here, was not further analyzed because half of the groups worked under time restrictions, rendering the resulting data of limited value. However, actual study time acquired under the free study time condition was further post hoc analyzed.*

Figure 2: Mean recall-test scores as a function of diagnostic approach and study-time (free vs restricted).



Significant effect of diagnostic approach: p = 0.013, *d*= 0.54

As in a previous study,[4] a significant main effect of diagnostic approach on SI was found, with participants who diagnosed the cases through deliberate reflection showing higher scores on SI [means: 4.45 vs. 3.99, $F(1,99) = 26.27$, $p <.001$, d=1.01] than those who gave differential diagnoses. A main effect in favor of reflection also emerged for AKG [means: 4.13 *vs.* 3.85, $F(1,99)= 8.42$, $p < .01$, d=.58]. These findings suggest that deliberate reflection fosters an interest in the topic-at-hand and makes students aware of knowledge deficits. This enhanced interest does not seem to translate into increased engagement with the subject matter as measured by study-time. We will return to this issue in the Discussion section.

As a post-hoc test to further explore the influence of time, we used structural equation modeling to assess whether actual study-time served as a moderator variable between diagnostic approach and cued recall for participants in the free-study time condition. Figure 3 presents a model of this relationship that fits the data, but the beta-weight of the relationship between diagnostic approach and study-time is not significantly different from zero.

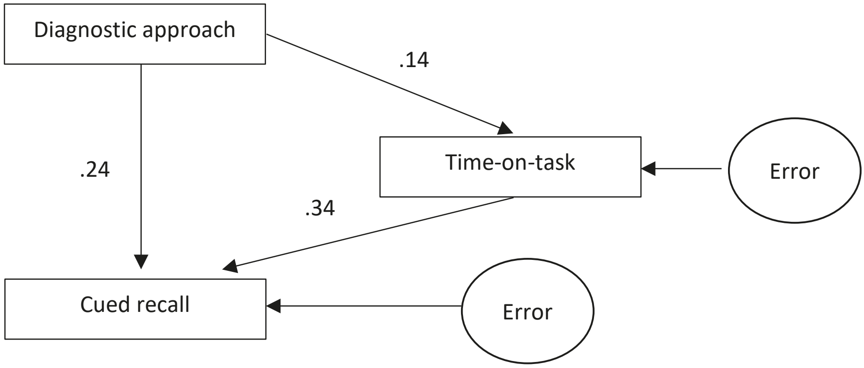


Figure 3: Structural equations model of the relationship between diagnostic approach, time-on-task, and cued recall for the free-time condition only (N = 49).

DISCUSSION

In this experiment we investigated whether deliberate reflection while practicing with clinical cases fosters medical students' learning and examined two potential mechanisms through which deliberate reflection could affect learning: by fostering motivation and by facilitating cognitive processes involved in learning. To do so, we ran an experiment in which fifth-year medical students diagnosed clinical cases presenting patients with jaundice either through deliberate reflection or differential diagnosis, reported their SI and AKG about the topic, studied a text referring to the topic either with free or restricted-time and, finally, answered a test about this text. Students who deliberately reflected upon the cases reported higher scores on SI and AKG and obtained higher scores on the cued recall test relative to those who gave differential diagnosis. These findings are in line with our hypotheses that working on cases through deliberate reflection would be more beneficial to students' learning than the more conventional approach of generating a differential diagnosis, replicating what was observed in previous experiments.[4,5] The effect size of deliberate reflection was large for SI and medium for AKG and cued recall.[27]

This effect of diagnostic approach on cued recall was independent of study-time (free *vs*. restricted), given that there was neither a significant effect of study-time on cued recall nor an interaction between diagnostic approach and study-time. These findings suggest a cognitive-processing-only explanation for the findings, rejecting the possibility of motivational influences. This argument is outlined in the Introduction: if a motivational mechanism-only would underlie the effect, keeping time restricted would wipe-out any differences between both diagnostic approaches under that condition, causing an interaction effect in the data. We did not observe such interaction, making an explanation in terms of motivation-only unlikely. A third possibility that we considered was that *both* cognition and motivation play a role in the effect of deliberate reflection on learning. However, this possibility was falsified by the fact that there was no main effect of time (free *vs*. restricted) on cued recall. Were a combination of both operating, performance on the cued-recall test should be significantly higher under free time conditions than under restricted time. We will return to the issue of motivation below.

Thus, deliberate reflection upon the cases seems to have facilitated cognitive processes that are known to influence learning outcomes. One such cognitive process is the *activation of prior knowledge*, as it has been observed in the context of PBL.[14-17] Although we did not have measures of information processing while students diagnosed the cases, it is reasonable to think that confronting diagnostic hypotheses with patients' data in a systematic way mobilized students' knowledge from long-term to working-memory[28] more extensively than having them listing differential diagnoses. Besides the activation of prior knowledge, by requiring learners to connect clinical findings with each other and with the hypotheses, reflection might also have facilitated *elaboration* of knowledge. These processes fostered by reflection possibly facilitated processing and incorporation of the new information encountered in the learning task. This result is consistent with studies on PBL and self-explanation that have found students who attempt to explain, in their own words, the problems with which they are working to perform better on tests on knowledge relevant to those problems than those who do not explain.[15,29] In summary, deliberate reflection is suggested to activate prior knowledge and elaboration to a larger extent than simply listing differential diagnoses, thereby facilitating processing of new information and subsequent recall to a larger extent.

A tantalizing finding, however, is the unequivocal positive effect of deliberate reflection on two measures of engagement, situational interest (SI) and awareness of knowledge gaps (AKG), often considered measures of motivation. [9-12, 20,21] Deliberate reflection compels students to scrutinize different diagnostic hypotheses for the cases, arguing for and against them before making any conclusions, which may create additional challenge and bring knowledge gaps to surface, thereby stimulating interest in the topic-at-hand.[10] We expected increased interest to translate into longer engagement with the learning task for those whose study-time was not restricted. Students in the deliberate reflection condition indeed tended to invest more time than their colleagues from the differential diagnosis condition, but the expected positive interaction between diagnostic approach and study-time did not emerge. Reading research, however, shows that interest is not associated only with study-time, but also with choices, and can translate into what people choose to focus attention on.[11,24] This is a motivational mechanism that might have influenced results particularly in the restricted-time groups, who were aware their time would be restricted and might, therefore, have focused attention on information that could fulfil their perceived knowledge gaps, optimizing study-time. As our experiment does not allow us to identify specific motivational mechanisms, these assumptions require further investigation.

A second finding that does not seem to entirely fit with our conclusion that only cognitive processes can be held responsible for the effects of deliberate reflection and that study-time, as an indicator of engagement, does not play a role, is that, in the free-time condition, students who reflected spent 17% more time on reading the text. Although this difference was not significant, we decided to relate, for the free-time condition only, actual study-time as a moderator variable between diagnostic approach and cued recall, in a post-hoc analysis using structural equations modeling. Figure 3 presents a model of this relationship that fits the data.

However, the beta-weight of the relationship between diagnostic approach and study-time is non-significantly different from zero. This suggests that study-time plays a role in determining performance on the cued-recall test, but this effect is *not* driven by the diagnostic approach. Rather, it must be driven by extraneous factors, possibly related to individual differences among students.[30]

Limitations

This experiment was run in a single medical school, with all participants in the same training year, and a single medical subject, jaundice, was explored. Though reasons that would make the findings specific to these particular students or topic are not clear, the generalizability of the findings may have been compromised as a result. In addition, we cannot exclude the possibility that participants in the differential diagnosis condition who had to solve crosswords puzzles may have experienced cognitive load that somehow carried over to the test phase and could have negatively influenced their performance. However, deliberate reflection has been shown to involve substantially higher cognitive load than other problem-solving approaches with which it has been compared.[31] Therefore, it is unlikely that cognitive load has affected only participants in the differential diagnosis condition. Finally, since we used immediate post-test scores as a measure of learning outcome, we are unable to know if the positive effect of deliberate reflection on learning would last longer.

CONCLUSION

Our experiment provides additional evidence of the positive effect of deliberate reflection on medical students' learning of subsequent relevant material observed in a previous experiment,[5] and suggests that this effect is mediated by cognitive mechanisms. Future research should study how these mechanisms are facilitated by reflection, thereby opening the door for an optimal use of deliberate reflection by clinical teachers.

REFERENCES

1. Dewey J. *How we Think*. Boston: D. C. Haeth and Co.; 1913.
2. Sandars J. The use of reflection in medical education: AMEE Guide No. 44. *Med Teach*. 2009;31(8):685-95.
3. Uygur J, Stuart E, De Paor M, et al. A Best Evidence in Medical Education systematic review to determine the most effective teaching methods that develop reflection in medical students: BEME Guide No. 51. *Med Teach*. 2019;41(1):3-16.
4. Ribeiro LMC, Mamede S, Moura AS, de Brito EM, de Faria RMD, Schmidt HG. Effect of reflection on medical students' situational interest: an experimental study. *Med Educ*. 2018;52(5):488-96.
5. Ribeiro LMC, Mamede S, de Brito EM, Moura AS, de Faria RMD, Schmidt HG. Effects of deliberate reflection on students' engagement in learning and learning outcomes. *Med Educ*. 2019;53(4):390-7.
6. Mamede S, Schmidt HG , Penaforte JC. Effects of reflective practice on the accuracy of medical diagnosis. *Med Educ*. 2008;42(5):468-75.
7. Mamede S, van Gog T, Moura AS, et al. Reflection as a strategy to foster medical student's acquisition of diagnostic competence. *Med Educ*. 2012;46(5):464-72.
8. Mamede S, van Gog T, Moura AS, de Faria RMD, Peixoto JM, Schmidt HG. How can students' diagnostic competence benefit most from practice with clinical cases? The effects of structured reflection on future diagnosis of the same and novel diseases. *Acad Med*. 2014;89(1):121-7.
9. Rotgans JI, Schmidt HG. Situational interest and academic achievement in the active-learning classroom. *Learn Instr*. 2011;21(1):58-67.
10. Rotgans JI, Schmidt HG. Situational interest and learning: Thirst for knowledge. *Learn Instr*. 2014;32:37-50.
11. Ainley M, Hide S, Berndorff D. Interest, learning and the psychological processes that mediate their relationship. *J Educ Psychol*. 2002;94(3):545-61.
12. Linnenbrink-Garcia L, Pattal EA, Messersmith EE. Antecedents and consequences of situational interest. *Br J Educ Psychol.* 2013;83(4):591-614.
13. Fredrick WC, Walberg HJ. Learning as a function of time. *The J Educ Res.* 1980;73(4):183-94.
14. Schmidt HG, Rotgans JI, Yew EHJ. The process of problem-based learning: what works and why. *Med Educ*. 2011;45(8):792-806.
15. van Blankenstein FM, Dolmans DHJM, van der Vleuten CPM, Schmidt HG. Which cognitive processess support learning during small-group discussion? The role of providing explanations and listening to others. *Instru Sci*. 2011;39:189-204.
16. Schmidt HG, De Volder M, De Grave WS, Moust JHC, Patel VL. Explanatory models in the processing of science text: the role of prior knowledge activation through small-group discussion. *J Educ Psychol*. 1989;81(4):610-9.
17. Loyens SMM, Jones SH, Mikkers J, van Gog T. Problem-based learning as a facilitator of conceptual change. *Learn Instr.* 2015; 38: 34-42.
18. VanLehn K, Jones RM, Chi MTH. A model of the self-explanation effect. *J Learn Sci*. 1992; 2(1):1-59.
19. Chi MTH, de Leeuw N, Chiu M-H, LaVancher C. Eliciting self-explanations improves understanding. *Cognitive Sci.* 1994;18(3):439-77.
20. Rotgans JI, Schmidt HG. The role of teachers in facilitating situational interest in an active-learning classroom. *Teaching and Teach Educ*. 2011;27(1):37-42.
21. Rotgans JI, Schmidt HG. Interest development: Arousing situational interest affects the growth trajectory of individual interest. *Contemporary Educ Psychol.* 2017;49:175-84.

22. Glogger-Frey I, Fleischer C, Gruny L, Kappich J, Renkl A. Inventing a solution and studying a worked solution prepare differently for learning from direct instruction. *Learn Instr*. 2015;39:72-87.
23. Meyer BJF. The organization of prose and its effects on memory. In: van Dijk TA, Hendricks WO, eds. *North-Holland studies in theoretical poetics.* Vol 1. Amsterdam: North-Holland Publishing Co; 1975.
24. Schiefele U, Krapp A. Topic interest and free recall of expository text. *Learn Indiv Diff.* 1996;8(2):141-60.
25. Fraenkel JR, Wallen NE. Content analysis. In: Fraenkel JR, Wallen NE, eds. *How to design and evaluate research in education*. 7th ed. New York: McGrraw-Hill; 2009: 472-98.
26. Field A. Exploring data. In: Field A, ed. *Discovering statistics using IBM SPSS*. 3rd ed. Los Angeles: Sage; 2009:85-124.
27. Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Front Psychol*. 2013;4:863.
28. Ormrod JE. Memory and its components. In: Ormrod JE, ed. *Human Laearning*. 7th ed. Boston: Pearson; 2016: 178-205.
29. Larsen DP, Buttler AC, Roediger HL 3rd. Comparative effects of test-enhanced learning and self-explanation on long-term retention. *Med Educ.* 2013;47(7):674-82.
30. Byrne B M. *Structural equation modeling with AMOS: basic concepts, applications, and programming*. New York: Taylor & Francis Group; 2010.
31. Ibiapina C, Mamede S, Moura A, Elói-Santos S, van Gog T. Effects of free, cued and modelled reflection on medical students' diagnostic competence. *Med Educ* 2014;48(8):796-805.

Chapter 5

Does deliberate reflection foster medical students' activation of and elaboration on prior knowledge?

ABSTRACT

Context. Deliberate reflection upon clinical cases has been shown to foster medical students' learning from scientific texts relevant to the cases, most likely through cognitive mechanisms. Which mechanisms, however, are yet to be understood. This experiment tested the hypothesis that deliberate reflection, compared to other approach to clinical reasoning, activates students' prior knowledge and helps them to elaborate on it to a larger extent.

Methods. Fourth-year medical students recalled from memory the "typical clinical picture" of three diseases, acute viral hepatitis, choledocholithiasis, and hemolytic anemia, after analyzing cases of these diseases either through deliberate reflection (DR) or differential diagnosis (DD). A third group recalled the typical clinical pictures without any previous exposure to clinical cases (control). The recall of the diseases was taken as a measure of students' knowledge about them, both prior knowledge and elaboration on this knowledge, and was compared between groups.

Results. Eighty-four students were included in the analysis. Scores on knowledge test were 14.82 (7.47), 16.71 (8.03) and 20.11 (7.49) for control, DD and DR, respectively. There was a significant difference between groups [F (2.81=3.42, p= 0.038], with a linear trend. Students who diagnosed clinical cases either through DD or DR performed better than those in the control group [t (81)= 2.022, p= 0.023 (one-sided), d= 0.94]. The comparison between DD and DR showed marginal results [t (81)= 1.66, p= 0.051 (one-sided), d= 0.44].

Conclusions: Diagnosing clinical cases, especially DR, favored students' activation of and elaboration on their prior knowledge about diseases. The results help to understand the mechanisms through which case problem-solving and DR can improve learning outcomes. Medical teachers can use DR to activate students' knowledge about diseases, facilitating their learning of new information.

INTRODUCTION

Deliberate reflection upon clinical cases is a diagnostic procedure that requires a systematic analysis of clinical data to find evidence that supports and contradicts different diagnostic hypotheses in a systematic way before making conclusions.[1] The procedure, initially tested with physicians, has been shown to help medical students to improve their diagnostic accuracy.[2-5] This positive effect of reflection was observed even when no additional information was offered to students while or after diagnosing the cases.[2,3] This deliberate reflection procedure has been shown to be also a promising strategy to foster learning of new information relevant to the clinical cases. Ribeiro *et al*[6,7] ran two experiments in which they analyzed the effects of deliberate reflection on students' engagement with scientific texts relevant to clinical cases and learning from them. In one experiment, after diagnosing clinical cases either through deliberate reflection or by providing differential diagnosis, students studied a text relevant to the cases and took a recall test about this text. Students who deliberately reflected upon the cases spent more time in studying the text and obtained higher scores on the recall test than those who provided differential diagnosis.[6]

It is reasonable to assume that this effect could be due to motivation to learn and its consequent longer engagement with the text.[8] That means, deliberate reflection would foster learning simply by increasing time spent studying the text. Nevertheless, as it happens with other problem-solving tasks,[9-16] deliberate reflection may also have affected students' cognitive processes thereby facilitating studying the text. In a subsequent and similar experiment, Ribeiro *et al*[7] manipulated study-time to test if an effect that could not be attributed to study-time alone would emerge. In this experiment, 5th-year medical students analyzed clinical cases either reflecting upon the cases or providing differential diagnosis. Subsequently, half participants in each diagnostic procedure group could read a text relevant to the cases as long as they wanted, and half could read the same text within a restricted amount of time. There were, therefore, four experimental groups: deliberate reflection/free study-time or restricted study-time; differential diagnosis/free study-time or restricted study-time. Finally, all participants took a recall test about the text. Students who reflected upon the cases showed higher scores on the recall test relative to those who provided differential diagnosis regardless of their study-time condition. This effect on test scores, independent of study-time, cannot to be attributed to students' motivation to read the text alone and indicates that reflection favored students' learning from the text through cognitive mechanisms.

What cognitive mechanisms could be involved? Reflection could have helped students to activate their prior knowledge about the diseases of the cases they diagnosed, which would facilitate the cognitive processing of new information about these diseases that was presented to them in the text. Learning of new information is facilitated by the activation of knowledge stored in long-term memory. Bringing this knowledge to working memory facilitate understanding of the new information and allows for integrating this new information with what we already know about a certain topic.[17,18] In educational contexts, problem-solving prior to study of textual material is an instructional approach that has been largely adopted because, among other positive effects, it fosters the activation of prior knowledge. When someone is asked to solve a problem, the knowledge s/he has about it is likely to be brought to working memory, even if the person is unable to completely solve this problem.[19] This is one of the mainstays of problem-based learning (PBL), supported by substantial experimental evidence. [9-14] For example, De Grave, Schmidt and Boshuizen[12] compared learning from a text about blood pressure regulation between a group of 1st-year medical students who had previously discussed a problem about blood pressure and a group who had discussed a problem about vision in PBL sessions. Students who discussed the blood pressure problem remembered more about the text than those who discussed the vision problem. There is evidence that this effect also happens with individual problem-solving.[14]

It is reasonable to expect that a clinical problem-solving task could have effects on activation of prior knowledge similar to the aforementioned one. It is also reasonable to expect that deliberate reflection, asking students to find distinctive features of specific diseases with similar presentation that would allow to differentiate them, could be a stronger facilitator of this activation relative to providing differential diagnosis. Besides facilitating the activation of students' prior knowledge, deliberate reflection could also foster elaborations on this knowledge. Intuitively, a student might be able to think of plausible diagnostic hypotheses to a clinical case without making sense of all available clinical data. Further analyzing the case might allow this student to make inferences about information that was, initially, unfamiliar or uncertain to him/her, "linking" it to a specific diagnostic hypothesis. This process of constructing knowledge while retrieving information from long-term to working memory, even though may eventually contain errors, help fill gaps in knowledge.[17] As deliberate reflection asks students to argue for and against each diagnostic hypothesis in a systematic way, it is reasonable to expect that it would foster elaborations on prior knowledge to a larger extent than providing differential diagnosis.

To test these hypotheses, we run an experiment in which medical students were asked to describe the typical clinical pictures of three diseases, which were taken as a measure of their knowledge about these diseases, just after diagnosing clinical cases of these diseases either through deliberate reflection or by providing differential diagnosis. A third group of students was asked to describe the typical clinical pictures of the diseases without any previous exposure to clinical cases. We expected knowledge recall to be higher for students who diagnosed the cases relative to those who did not, with an advantage for those who diagnosed the cases through deliberate reflection, relative to those who provided differential diagnosis.

METHODS

Design

The study was an experiment with participants randomly assigned to one of three conditions. In all conditions, participants were requested to describe the typical clinical pictures of three diseases. Depending on their experimental condition, they did this 1) after having diagnosed cases of the three diseases through deliberate reflection (DR), 2) after diagnosing the same cases by providing a differential diagnosis (DD), or 3) without previously diagnosing the cases. The latter acted as a control group. The participants in DR and DD groups received no feedback about their analysis or their diagnostic accuracy before describing the typical clinical pictures of the diseases.

Settings and participants

The participants of the study were 4th-year students of the UNIFENAS-Belo Horizonte Medical School. The school has a six-year PBL curriculum with the last two years dedicated to clerkships. We chose 4th year students because at this point of their medical training they are familiar with diagnosis of jaundice, a topic they engaged with in tutorial groups and seminars, especially in the 3rd-year, but have limited clinical experience with it. The experiment was approved by UNIFENAS' Ethics Committee (number 25037819.1.3001.5149) and participants gave their consent.

Twenty-two 4th-year students at UNIFENAS-BH Medical School were invited to a pilot test of the experiment in 2020. Nineteen students volunteered and completed the pilot. It tested the materials and procedures, which showed appropriate and received no adjustments. Based on the results of the pilot, a priori power analysis was performed for the main study, which assumed a medium effect size (Cohen's f = 0,34), the standard alpha level of 0.05, and lead to an estimation of 87 participants to be sufficient to achieve a power of 0,80.

Two-hundred and forty 4th year medical students at the same school were invited to volunteer to the experiment in May 2021 and April 2022. In 2021, the experiment was run during remote academic activities due to the Covid-19 pandemic, and invitations were made during synchronic online academic meetings and through message's *app*. Students were asked to work on their own at the moment that best fitted their timetable throughout one week. In 2022, the experiment was in person, and took place in a computer lab with participants working individually, under the supervision of a teacher, in five different sessions.

Materials and procedures

All materials were presented to participants exclusively via computer using *Qualtrics* software, which automatically randomized volunteers to one of the three experimental conditions.

Initially, participants rated their confidence in diagnosing patients with jaundice on a five-point Likert scale (*very low confidence* to *very high confidence*) and estimated their previous experience with jaundice, measured by how many real patients with jaundice they recalled having encountered (0, 1-3, 4-6, 7-9, >9). Subsequently, the Qualtrics program presented the students with instructions and examples of the tasks for the experimental group to which the participant had been assigned. The students could review the instructions if they wished, but once the exercise started, participants could neither skip steps nor move backward. The software automatically recorded each participant's responses and experiment's time and flow.

Participants in the DR condition were requested to: 1) for each case, read it and type down the most likely diagnosis for the case, 2) reflect upon the case by following a structured procedure[1] which, briefly, consists of typing the clinical findings that are in line with their first diagnostic hypothesis, those that contradict it, and those that are expected were this first hypothesis true but are not described in the case, type two alternative diagnoses and run the same analysis for each diagnosis before making a conclusion on the most likely diagnosis for the case. Subsequently, they were requested to 3) describe the typical clinical picture of the diseases. Students in DD condition were asked to 1) for each case, read it, type down the most likely diagnosis for the case and 2) type down two alternative diagnoses and, subsequently, 3) describe the typical clinical picture of diseases. The cases consisted of written descriptions of clinical history, physical examination and laboratory findings. Each case had a most likely diagnosis validated in previous experiments: acute viral hepatitis, choledocholithiasis, or falciform anemia.[2,3] For the control group the first task was to describe the typical clinical picture of diseases.

Participants in the differential diagnosis group had up to 2,5 minutes to give the initial diagnosis and two alternative diagnoses for each case. Participants in the deliberate reflection group had up to 6 minutes to reflect upon each case. Limiting the time that participants in the DD group spent diagnosing the cases was necessary to avoid their engagement in extensive reflection, which would blur the distinction between the DD and the DR experimental conditions.

The task of describing the typical clinical picture of the diseases consisted of three open-ended questions, each for one of the diseases represented in the diagnostic task. For example, *describe the typical clinical picture of a patient with choledocholithiasis as completely as you can.* Instructions to include history, physical examination and laboratory data and an example of a typical clinical picture of a different disease (herpes zoster) were presented to students just before the task to help them understand it. No restriction of time was set to the description of the typical clinical pictures of the diseases.

After describing the typical clinical pictures of the diseases, the control group diagnosed the same clinical cases DR and DD groups had diagnosed as their first task. This diagnostic task was included to measure the control group's initial diagnostic accuracy on the clinical cases.

To ensure that all participants spent similar time on the experiment, which was important for operational reasons, those on the differential diagnosis and control groups were presented to a time-filling task before diagnosing the cases and/or describing the typical clinical pictures of the diseases (see Figure 1). This task consisted of reading a text not related to the diagnosis of jaundice and solving two crosswords with medical terms unrelated to jaundice.

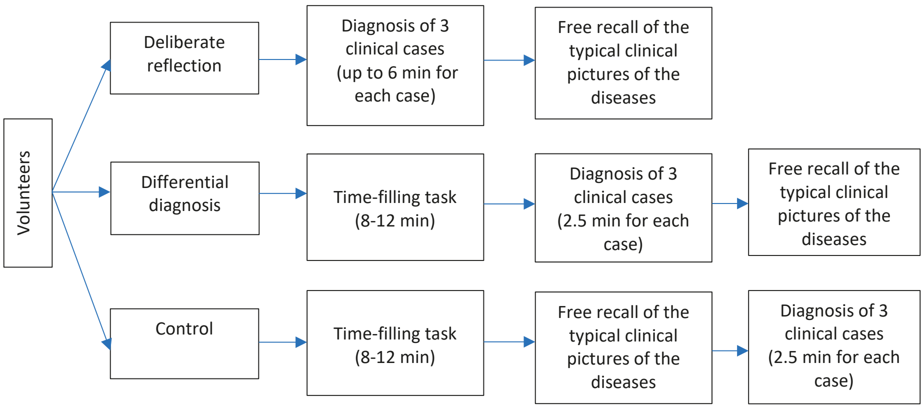


Figure 1: Experimental tasks and flow

No information about their diagnostic performance or about the diseases represented in the clinical cases was provided to the participants during the experiment. At the end of the experiment students were asked to disclose if, for any reason, they had consulted sources of information other than the experiment itself while doing it.

After completing the experiment, students were able to access an illustrated text about the differential diagnosis of jaundice based on the cases they had just diagnosed. This text was provided to them as a token that acknowledged their collaboration and was not part of the experiment.

Data analysis

The accuracy of initial diagnoses was measured to check whether the experimental groups were similarly acquainted with the problem under study. The initial diagnoses provided by the students in the diagnostic task were independently assessed by two board certified internists (LMCR and ASM), who were blinded to experimental condition. They classified each response as correct (scored 1), whenever the core diagnosis was present (e.g. "viral hepatitis", in the "acute viral hepatitis" case); partially correct (scored 0.5), if the core diagnosis was not present, but a component of it was (e.g. "cholelithiasis" in the "choledocholithiasis" case); and incorrect (scored 0), when the response did not fall into any of these categories. The raters agreed in the score attributed in 86% of the responses and resolved discrepancies by reaching consensus in a subsequent meeting. For each participant, scores achieved on each of the three cases were summed. Subsequently, they were averaged across participants for each experimental condition.

Participants' knowledge about the diseases was measured by counting the number of correct idea units present in their descriptions of their typical clinical pictures.[20,21] Notice that these descriptions represent the prior knowledge that participants had about the disease (they were not exposed to any additional knowledge), but as reconstructed by the engagement in problem-solving during the diagnostic task. For example, a participant, while reading one of the cases, thinks of cholestasis because there is an elevation of conjugated bilirubin, an information familiar to him/her (prior knowledge). However, s/he does not remember how to interpret the levels of alkaline phosphatase. As the conjugated bilirubin information indicates cholestasis, s/he "links" the levels of alkaline phosphatase described in this case with this syndrome, elaborating, therefore, on his/her prior knowledge (new knowledge). Afterwards, when asked to describe the typical clinical picture of a disease that causes cholestasis, both prior and new knowledge might have been cited.

We considered idea units to be text fragments, such as a word or a short sentence, that corresponded to a correct information about the diseases. Examples of idea units regarding the task to *describe the typical clinical picture of a patient with choledocholithiasis as completely as you can* are "*patient has jaundice* (1IU) *and upper right* (1IU) *abdominal pain* (1IU)", totaling, in this excerpt, 3 IU. Two authors (LMCR and EMB), independently assessed 10% of participants' responses, reaching an initial agreement of 88,6%. After discrepancies were resolved, the first author coded all participant responses. Only manifest content was considered (i.e., no possible underlying meanings were counted).[22]

The main outcome of the experiment was participants' scores on the descriptions of the typical clinical pictures of the diseases, which were taken as a measure of participants' knowledge about the diseases. To obtain these scores, we first summed the number of idea units recorded by each participant in all three descriptions and then averaged across participants for each experimental condition.

Separate one-way ANOVAs were run to compare age, initial diagnostic accuracy, confidence diagnosing jaundice and experiment-time across experimental conditions. *Chi*-square tests compared gender, experiment remotely done *vs* in person and experience with patients with jaundice between groups. A one-way ANOVA with experimental condition as fixed factor was performed on the number of idea units with the following contrasts planned to test our hypotheses: 1) diagnosis of cases by any approach before describing the typical clinical pictures of the diseases *vs* no exposure to clinical cases before this task, and 2) diagnosis of cases through deliberate reflection *vs* differential diagnosis. The data were analyzed using SPSS for Macintosh, version 25 (IBM Corp, Armonk, NY). Significance level was set at $p < 0.05$.

RESULTS

Thirty-nine students completed the experiment remotely in 2021 and forty students did it in person in 2022. Remote answers were carefully analyzed in order to find evidence of compromised data. Two participants did the experiment twice, and their duplicated answers were removed. Six answers were removed because time to complete the experiment was too long (> 4 times the average time), indicating that these participants might not have taken the experiment seriously. Two answers were removed because participants disclosed consulting external information while doing the experiment. Since nothing in the experiment was modified after the pilot, and the pattern of responses in the pilot sample was similar to that observed in the two other cohorts, we aggregated the participants of the pilot, which led to 88 participants included in the exploratory analysis. In this phase, one outlier regarding initial diagnostic accuracy and three regarding IU scores (*z*-scores >2.5) were removed. The total number of participants considered in the subsequent analysis was 84, 28 in each experimental condition.

Table 1 presents participants' characteristics and main results as a function of experimental condition. No significant differences between groups emerged in age, $F(2,80) = 0.71$, $p = .138$, gender, $X^2(2) = .716$, $p = .70$, proportion of participants who did the exercise remotely or in person, $X^2(2) = .11$, $p = .95$, confidence in diagnosing jaundice, $F(2,81) = 0.77$, $p = .467$, previous experience with patients with jaundice, $X^2(6) = 4.82$, $p = .567$, experiment time $F(2,81) = .2.01$, $p =.140$ and initial diagnostic accuracy, $F(1,81)=2.03$, $p = .138$.

Table 1 shows participants' scores on the descriptions of the typical clinical pictures of the diseases according to experimental condition. Levene's test of homogeneity of variances for these scores was not significant, allowing the ANOVA to be performed. There was a significant difference between groups [$F(2,81)= 3.42$, $p= 0.038$], with a significant linear trend [$F(1,81)= 6.653$, $p= 0.012$]. The first contrast showed that students who diagnosed clinical cases (either through DD or DR) prior to recall their knowledge of the diseases performed better than those in the control group [$t(81)= 2.022$, $p= 0.023$, (one-sided), $d= 0.94$]. The second contrast, which compared diagnosing cases through DD and DR, showed marginal results [$t(81)= 1.66$, $p= 0.051$ (one-sided), $d= 0.44$]. Effect sizes for contrasts 1 and 2 were, respectively, large and small, considering Cohen's *d* values of 0.2 for small, 0.5 for medium and 0.8 for large effect size.[23]

Table 1: Characteristics of participants and scores on the description of the typical clinical pictures of the diseases (knowledge test) according to experimental condition. Standard deviations (SD) and percentages into brackets.

	Deliberate reflection (n=28)	Differential diagnosis (n=28)	Control (n=28)	Overall (n=84)
Age	24.21 (3.12)	23.33 (2.70)	24.39 (4.46)	23.99 (3.50)
Females/males	16/12	10/18	9/19	53/31
Did the exercise remotely	9 (32%)	9 (32%)	10 (36%)	28 (33%)
Confidence diagnosing jaundice (confident or very confident)	6 (21.4%)	12 (42.9%)	5 (17.9%)	23 (27.4%)
Experience with patients with jaundice (0-3 patients).	25 (89.2%)	28 (100%)	27 (96.4%)	80 (95.2%)
Initial diagnostic accuracy	1.18 (0.67)	0.79 (0.63)	1.09 (0.95)	1.02 (0.77)
Duration in seconds.	2486 (792)	2599 (996)	2166 (690)	2417 (846)
Knowledge test scores				
Mean (SD)	20.11 (7.49)	16.71 (8.03)	14.82 (7.47)	17.21 (7.89)

DISCUSSION

In this experiment, we explored the potential of DR upon clinical cases to foster medical students' activation of and elaboration on their prior knowledge about diseases. To this end, we asked students to describe the "typical clinical picture" of three diseases, a measure of students' knowledge about them, after they diagnosed clinical cases of the same diseases through DR, DD, or without any exposure to clinical cases. We compared participants' knowledge scores between groups. Participants who diagnosed clinical cases scored higher in the knowledge test than those who did not, with a large effect size. Considering only participants who diagnosed clinical cases before the knowledge test, those who reflected upon the cases scored 20% higher than those who provided a differential diagnosis, a difference that was marginally significant with a small effect size.

These results are in line with our hypotheses. We expected students who diagnosed cases, regardless of diagnostic procedure, to score higher on the knowledge test than the students who did not diagnose any cases. As participants in the DD and DR groups received no information about the cases or the correct diagnoses before doing the test, the higher scores observed must be a consequence of the activation of their prior knowledge about the diseases, and possibly of elaborations on it. This result replicates what has already been observed in other problem-solving strategies, such as PBL and self-explanation[9-16] to the clinical reasoning field.

It is relevant because activating prior knowledge means bringing knowledge from long-term memory into working memory, where processing of the information at hand happens. Once in the working memory, one's prior knowledge can "meet" with new information, such as information presented in a text one is reading and that, through the sensory register, is also moved into working memory. The simultaneous presence of prior and new knowledge into working memory facilitates encoding of information and, therefore, learning.[17]

As we expected, students who deliberately reflected upon the cases showed higher scores in the knowledge test relative to those who provided a differential diagnosis, although the difference was marginally significant. The average initial diagnostic accuracy on the clinical cases was similar for all experimental groups and, overall, 34%. The cases were, therefore, challenging to participants, as expected, since they were previously exposed to information about jaundice in tutorial groups and seminars but had very limited experience with patients with jaundice. Participants had prior knowledge about the diagnosis of jaundice, but they did not "use" it frequently, which makes retrieval of information from long-term to working memory more difficult.[17] DR demands participants to argument not only for but also against different hypotheses before making a conclusion. This conscious effort of comparing and contrasting different diagnostic hypotheses possibly allowed a larger mobilization of students' knowledge about the diseases related to jaundice, relative to the more intuitive procedure of providing a differential diagnosis. This would facilitate the retrieval of the typical clinical pictures of diseases that cause jaundice subsequently. It might also have allowed participants to elaborate on their prior knowledge, while diagnosing the cases, to a larger extent than providing a differential diagnosis. Comparing and contrasting diseases might have facilitated making correct inferences based on prior knowledge, "linking" familiar and unfamiliar information. Recognizing the elevation of conjugated bilirubin as an indicator of cholestasis in one case, for example, might have allowed a participant to subsequently recognize that the elevation of phosphatase alkaline, presented in the same case, suggests the same diagnosis. Both activation of and elaboration on prior knowledge probably allowed DR participants to score higher than the DD participants in the knowledge test. This result is consistent with previous research that have been shown DR upon clinical cases to foster subsequent learning of scientific text relevant to the cases to a larger extent than DD,[6,7]and that suggest that this positive effect of DR is likely due to cognitive processing.[7]

Educational implications

Deliberate reflection is a relatively simple and short time-consuming instructional procedure that seems to facilitate the activation of and elaboration on students' prior knowledge to a larger extent than the more conventional approach of providing a differential diagnosis. The activation of and the elaboration on prior knowledge can facilitate learning. Medical teachers could, therefore, use it to foster individual learning from clinical experience both in real and simulated scenarios. It could also be used before activities with large groups, like seminars, to foster learning of the topic at hand.

Limitations

The experiment was run in a single medical school and volunteers were all 4th-year medical students. Although there are no reasons to believe that results would be different if clinical cases had a main clinical finding other than jaundice, generalizability of the findings should be cautious. It is not possible, for the groups who diagnosed clinical cases, to distinguish between prior and new knowledge (elaborations on prior knowledge while diagnosing the cases) in students' answers. The experiment also did not explore the effects of activation of and elaboration on prior knowledge of DR in cases of different levels of difficulty or students at different levels of expertise. Moreover, it was also not the purpose of the study to clarify the relative contribution of activation of and elaboration on prior knowledge for the effect of DR. These investigations were beyond the scope of this experiment and should be addressed in future research.

CONCLUSIONS

Deliberate reflection upon clinical cases has been shown to foster medical students' learning of scientific texts relevant to those cases, possibly through cognitive mechanisms.[7] The results of this experiment suggest that among such mechanisms are the activation of and elaboration on students' prior knowledge. Other cognitive mechanisms, as the processing of clinical cases' information, are yet to be studied.

REFERENCES

1. Mamede S, Schmidt HG, Penaforte JC. Effects of reflective practice on the accuracy of medical diagnoses. *Med Educ.* 2008;42:468-75.
2. Mamede S, van Gog T, Moura AS, et al. Reflection as a strategy to foster medical student's acquisition of diagnostic competence. *Med Educ.* 2012;46:464-72.
3. Mamede S, van Gog T, Sampaio AM, de Faria RMD, Maria JP, Schmidt HG. How can students' diagnostic competence benefit most from practice with clinical cases? The effects of structured reflection on future diagnosis of the same and novel diseases. *Acad Med.* 2014;89(1):121-7.
4. Ibiapina C, Mamede S, Moura A, Eloi-Santos S, van Gog T. Effects of free, cued and modelled reflection on medical students' diagnostic competence. *Med Educ*. 2014;48(8):796-805.
5. Mamede S, Figueiredo-Soares T, Eloi-Santos S, Faria RMD, Schmidt HG, van Gog T. Fostering novice students' diagnostic ability: The value of guiding deliberate reflection. *Med Educ*. 2019;53(6):628-637.
6. Ribeiro LMC, Mamede S, de Brito EM, Moura AS, de Faria RMD, Schmidt HG. Effects of deliberate reflection on students' engagement in learning and learning outcomes. *Med Educ*. 2019;53(4):390-7.
7. Ribeiro LMC, Mamede S, de Brito EM, Moura AS, de Faria RMD, Schmidt HG. Exploring mechanisms underlying learning from deliberate reflection: An experimental study. *Med Educ*. 2021;55:404-12
8. Fredrick WC, Walberg HJ. Learning as a function of time. *J Educ Res*. 1980;73(4):183-194.
9. Van Blankenstein FM, Dolmans DHJM, van der Vleuten CP, Schmidt HG. Which cognitive processes support learning during small-group discussion? The role of providing explanations and listening to others. *Instruc Sci.* 2011;39(2):189-204.
10. Schmidt HG, Rotgans JI, Yew EH. The process of problem-based learning: what works and why. *Med Educ*. 2011;45(8):792-806.
11. Dolmans DHJM, Schmidt, HG. What Do We Know About Cognitive and Motivational Effects of Small Group Tutorials in Problem-Based Learning? *Adv Health Sci Educ*. 2006; 11:321–336
12. De Grave WS, Schmidt HG, Boshuizen HPA. Effects of Problem-based Discussion on Studying a Subsequent Text: A Randomized Trial among First Year Medical Students. *Instruc Sci.* 2001; 29(1), 33-44.
13. Loyens SMM, Jones SH, Mikkers J, van Gog T. Problem-based learning as a facilitator of conceptual change. *Learn Instr*. 2015;38:34-42.
14. De Grave WS, Schimidt HG, Belien JJ, Moust JHC, De Volder ML, Kerkhofs LMM. Effecten van verschillende typen van activatie van voorkennis op recall, gemeten met een aanvultoets (Effects of different types of activation of prior knowledge on recall, measured with a completion procedure.). Paper presented at the Onderwijs Research Dagen, Tilburg, The Netherlands,1985.
15. VanLehn K, Jones RM, Chi MTH. A model of the self-explanation effect. *J Learn Sci*. 1992;2(1):1-59.
16. Chi MTH, de Leeuw N, Chiu M-H, LaVancher C. Eliciting self-explanations improves understanding. *Cogn Sci*. 1994;18(3):439-77.
17. Ormrod JE. Long-term memory storage and retrieval processes. In: Ormrod JE, ed. *Human Learning.* 7th ed. Boston: Pearson Education; 2016. p. 206-48.
18. Mayer RE. Information Processing. In: Harris KR, Graham S, Urdan T, McCormick CB, Sinatra GM, Sweller J, eds. *APA Educational Psychology Handbook. Vol. 1. Theories, constructs and Critical Issues,* 1st ed. Washington DC: American Psychology Association 2012; 85-99.
19. Loibl K, Rummel N. Knowing what you don't know makes failure productive. *Learn Instruc*. 2014;34:74-85.
20. Meyer BJF. The organization of prose and its effects on memory. In: van Dijk TA, Hendricks WO, eds. *North-Holland studies in theoretical poetics, vol 1.* Amsterdam/Oxford: North-Holland Publisching Co. New York: Elsevier Publishing Co, 1975. 1-249

21. Schiefele U, Krapp A. Topic interest and free recall of expository text. *Learn Indiv Diff.* 1996;8(2):141-60.
22. Fraenkel JR, Wallen, NE, Hyun H. Content analysis. In Fraenkel JR, Wallen, NE, Hyun H, eds. *How to design and evaluate research in education.* 8th ed. New York: McGrraw-Hill 2012; 477-504.
23. Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for *t*-tests and ANOVAs. *Front Psychol*. 2013;4:863.

Chapter 6

Summary and General discussion

Reflection has been considered an important instructional approach in medical education. Medical schools, teachers and boards have relied on reflection to support students' learning from undergraduate to postgraduate scenarios.[1,2] Although a strong theoretical framework supports the assumption that reflection fosters learning,[3,4] empirical evidence of the outcomes of reflection is presently limited, as are the mechanisms that could underly the potential positive effects of reflection in the context of medical education.

In this thesis, a series of experiments examined if deliberate reflection upon clinical cases could facilitate medical students' subsequent learning of scientific texts relevant to the cases. It also examined which mechanisms, motivation and/or cognition, could underly this potential positive effect of reflection.

This final Chapter presents a summary of the four experiments of this thesis, analyzes its results altogether and discusses how they can support medical teachers in their daily practice and inform future research.

SUMMARY OF THE MAIN FINDINGS

Chapter 2 presents an experiment that compares 4th-year medical students' self-reported situational interest triggered by the analysis of clinical cases through two different approaches: deliberate reflection, which requires students to argue for and against different possible diagnoses before reaching a conclusion, and differential diagnosis, a more conventional approach of citing the most likely and two alternative diagnoses without further considerations. Situational interest (SI) is a short-living and context-related kind of interest that has been shown to be triggered by challenge and uncertainty.[6-8] Fourth-year medical students have limited clinical experience and we expected, therefore, that diagnosing clinical cases, however the used approach, would be, to some extent, challenging to them. However, we believed that deliberate reflection would be more challenging than differential diagnosis and, therefore, that it would trigger students' SI to a larger extent. Results were in line with our hypothesis that deliberate reflection would foster higher levels of SI among the students relative to giving differential diagnosis, a promising effect of reflection, considering that SI has been shown to be a good predictor of engagement in learning and learning outcomes.[6,7]

Chapter 2 also explored the "knowledge deprivation" theory of interest,[9,10] i.e., the assumption that SI is triggered by the awareness of knowledge gaps about a certain topic, in the context of clinical reasoning. Students' self-reported awareness of knowledge gaps (AKG) was also measured through a 9 item questionnaire.[11] Students who reflected upon the cases showed higher scores on AKG, but the difference between groups was not significant, except for a *pos-hoc* analysis run with two questions selected from the questionnaire that seemed to be measuring students' perception of knowledge gaps (for example, *working on this case clarified that I don't know certain things yet*) in opposition to some questions that perhaps captured students' frustration while analyzing the cases (as *I was stuck with diagnosing this case*).

Besides comparing SI and AKG scores between groups according to the diagnostic approach used to diagnose the cases, Chapter 2 also compared these scores according to cases' difficulty. Three of the six clinical cases students diagnosed were considered easy cases, with participants showing an average initial diagnosis score of 0.78, and three were considered difficult cases, with an average initial diagnosis score of 0.07 (range 0-1). The most unexpected and perhaps interesting finding of this experiment was the fact that, despite AKG was higher for the more difficult cases, SI was not: students, regardless of the diagnostic approach under which they performed, showed higher levels of SI for the easier cases. Considering the *knowledge-deprivation* theory of interest,[9] we expected the opposite result. This "disconnection" of situational interest and awareness of knowledge gaps contrasts with the findings of Rotgans and Schmidt in the context of PBL,[10,12] but is similar to what Glogger-Frey *et all*[11] observed with 8th grade students in an experiment that compared the outcomes of different instructional approaches, inventing a solution and studying a canonic solution, to a problem. As the Glogger-Frey *et al* experiment suggest, awareness of knowledge gaps might not underly situational interest for some audiences, or under some circumstances. Medical students diagnosing clinical cases could be one of such conditions. We cannot exclude, however, the possibility that the unexpected level of difficulty of the difficult cases in this experiment, much higher than we anticipated, might have hindered students' interest for these cases.

Chapter 3 describes an experiment designed to check if the positive results of deliberate reflection on students' situational interest, measured as a self-reported questionnaire, would replicate when interest was measured by an observable behavior: time students spent studying a scientific text relevant to the cases. The hypothesis that higher interest would translate into more learning was also tested. In an electronic environment, students diagnosed clinical cases with jaundice as the main clinical finding either through deliberate reflection or differential diagnosis. The cases chosen to this experiment were estimated to be of moderate difficulty to the students, to avoid the negative effect of higher levels of difficulty on students' interest observed in the previous experiment. Subsequently, all students read a text about the differential diagnosis of jaundice, based on the cases they had just diagnosed, as long as they wanted. Students' engagement with this text, measured as the time they spent reading it, was recorded. Finally, students did a cued-recall test about the information presented to them in the text. Students' scores on this test were taken as a measure of students' learning about the text. As we expected and in line with the findings of experiments on reading and problem-solving, which have been shown challenge and unexpectedness to be triggers of interest,[6,7,12] students who deliberately reflected upon the cases engaged longer with the text and showed higher scores on the test when compared to those who gave differential diagnosis. These findings provide empirical support that deliberate reflection can, indeed, foster learning from scientific texts.

Two mechanisms could explain the positive effect of deliberate reflection on students' learning observed in this experiment: motivation and/or cognition. If the underlying mechanism of the positive effect of deliberate reflection on students' learning was purely motivational, students who reflected upon the cases would have learned more about the text than those who gave differential diagnosis solely because they spent more time reading the text. On the other hand, deliberate reflection could have facilitated students' processing of the information presented in the text, for example, activating their prior knowledge and, therefore, facilitating students' incorporation of new information.[13-15] A third hypothesis is that both mechanisms, motivational and cognition, contributed to the better results on learning fostered by reflection.

Chapter 4 explores which mechanisms would underly the positive effect of reflection o students' learning observed in Chapter 3: motivation and/or cognition. To that end, medical students were randomly allocated into one of four experimental groups. Groups 1 and 2 diagnosed clinical cases with jaundice as the main clinical finding through deliberate refection. Subsequently, group 1 read a text about the diagnosis of jaundice with no restrictions of time. Group 2 read the same text, but they had a limited amount of time to do so. Groups 3 and 4 diagnosed the same clinical cases giving differential diagnosis. Subsequently, group 3 read a text about the diagnosis of jaundice without restriction of time while group 4 had a limited amount of time to read it. We manipulated, therefore, two variables: diagnostic approach (deliberate reflection *vs* differential diagnosis) and study time (free time *vs* restricted time). Finally, all participants did a recall test about the text they studied. Students' scores in this test were taken as a measure of students' learning about the text. The results showed a main effect of diagnostic approach, no effect of study-time and no interaction between diagnostic approach and study-time: students who reflected upon the cases performed similarly and better than those who gave differential diagnosis, regardless of the study-time condition under which they performed. These results suggest that the underlying mechanism of the positive effect of deliberate reflection on students' test scores was cognitive, not motivational. It seems that deliberate reflection operates through mechanisms such as the processing of information of the clinical cases, the activation of prior knowledge about the diseases and/or the process of integrating new information presented in the texts into previously existing knowledge structures.

The role of the motivational mechanism, however, could not be excluded. Measurements of students' self-reported situational interest (SI) and awareness of knowledge gaps (AKG) were also obtained. As we expected and similarly to what was observed in Chapter 2, both SI and AKG scores were higher for students who reflected upon the cases relative to those who gave differential diagnosis, suggesting that motivation might have also played a role on students' performance on the test. This time the difference between groups was significant for both measures, including the AKG scores considering all its questions, even those that might have captured feelings of frustration rather than a more emotionally-free perception of knowledge gaps.

Chapter 5 builds on the results of Chapter 4, which suggested that the main mechanism underlying the positive effect of reflection on students' learning is cognitive. To test if the activation of and elaboration on prior knowledge could be among such mechanisms, an experiment compared students' description of the "typical clinical picture" of three diseases, which was taken as a measure of their knowledge about the diseases. Before doing this knowledge test, students diagnosed clinical cases of the same three diseases either through deliberate reflection or by providing differential diagnosis, receiving no information about their diagnostic accuracy. A third group, which worked as a control, did not diagnose any cases before the test. As other problem-solving strategies has been shown to activate prior knowledge, [13-17] we expected that students who diagnosed clinical cases to perform better on the knowledge test than the group that did not diagnose any cases. If deliberate reflection indeed does facilitate the activation of and elaboration on students' prior knowledge to a further extent than providing a differential diagnosis, there would also be an advantage for the first group. Results were in line with our expectations, with students who diagnosed clinical cases performing better on the knowledge test than those who did not diagnose cases. There was also an advantage for the students who diagnosed the cases through deliberate reflection, although the difference between deliberate reflection and differential diagnosis on test scores was marginally significant.

GENERAL DISCUSSION

The series of experiments that compose this thesis suggest that deliberate reflection upon clinical cases can be a relevant instructional approach to motivate medical students to subsequently engage in studying scientific texts that are relevant to those cases (Chapters 2 and 3). It also suggests that deliberate reflection, relative to differential diagnosis, can improve learning outcomes from the study of such texts (Chapters 3 and 4), and that cognitive processing of information is the mechanism that is mostly responsible for this positive effect of reflection (Chapter 4). Finally, the activation of and elaboration on prior knowledge was suggested to be one cognitive mechanism facilitated by reflection (Chapter 5). This section will discuss each research question in the light of the experiments' results.

Research question 1: Would deliberate reflection upon clinical cases foster medical students' motivation to engage in subsequent study of scientific texts relevant to these cases?

Reflection, defined by Dewey[3] as the conscious effort to systematically analyze a challenging problem to find a proper solution for it, is assumed to be a relevant instructional approach for learning. It is a reasonable assumption, considering the empirical evidence provided by reading research, in fields other than medical education, which has been shown that unexpectedness, novelty and challenges to be triggers of interest, particularly the short-living, context-related interest called situational interest.[6-10] Reflection is also expected to be triggered by unexpectedness, novelty and challenges.[3] In the medical practice, it would be the case when, for example, a student or a physician has a hard time to find out what is the most likely diagnosis for a patient and does not "jump" into premature conclusions. Reflecting upon clinical cases could, therefore, motivate students to engage in learning activities.

In Chapters 2 and 4, medical students' situational interest, measured through a self-reported questionnaire, was higher when they diagnosed clinical cases through deliberate reflection when compared to the more conventional approach of giving differential diagnosis, with a large effect size. This SI questionnaire measures both emotions, with questions like *I enjoyed working with this case,* and intention to learn, as measured by the questions like *I want to know more about this case,* and it has been shown to be a good predictor of students' engagement with learning activities.[6,7,12] Indeed, in Chapter 3, which measured students' interest through study-time, students who reflected upon the cases engaged longer with a text relevant to the cases than those who gave a differential diagnosis. These findings suggest that deliberate reflection upon clinical cases can trigger and sustain medical students' interest for scientific texts relevant to these cases. Challenging problems had already been shown to be triggers of interest among secondary and economy students.[12] The novelty in the experiments of this thesis is to expand this evidence to the medical clinical reasoning and reflection field.

In Chapter 4, within the groups who had no restriction of time to study the scientific text, reflection also fostered longer engagement with the text relative to differential diagnosis, but this difference was not significant. One possible explanation to this unexpected outcome is that volunteers of this experiment were more advanced in medical school (5^{th} *vs* 4^{th}-year) and showed higher initial diagnostic accuracy scores (suggesting higher knowledge about the cases) relative to the volunteers of Chapter 3. Our restriction of time, based on the results of Chapter 3, might have given the audience of Chapter 4 enough time to go through a material about a topic that probably presented information that was more familiar to them compared to their more novice peers. We cannot rule out, however, that the sample size of Chapter 4 was not large enough to allow significant differences regarding study-time to emerge.

It is noteworthy that Chapter 2 showed that this positive effect of reflection on students' motivation can be hindered if clinical cases are too difficult for students. Contrary to our expectations, students' SI was higher for easier cases than for difficult cases. Students perhaps felt that the knowledge needed to solve the difficult cases, for which their initial diagnostic accuracy was below 10%, was beyond their reach. Too much of a challenge could put a task beyond an optimal level of incongruity[9] and have a disruptive effect on students' interest.

In complementary analysis, Chapters 2 and 4 also explored the knowledge deprivation hypothesis[9] for SI. Alongside with measures of SI, measures of students' awareness of knowledge gaps were also obtained. In Chapter 2, students who diagnosed the cases through deliberate reflection showed higher scores on AKG. The difference between groups, however, was not significant, except for a *post hoc* analysis with two questions that seemed to measure awareness of gaps *per se*, in contrast with other questions that might have capture students' frustration about the clinical tasks they worked with. In Chapter 4, on the other hand, a significant effect of deliberate reflection emerged both for SI and AKG, this time considering all AKG questions. Chapters 2 and 4 used the same questionnaires translated to Portuguese from English (SI) and German (AKG). In chapter 4, however, slight adjustments were made in the adaptation of the AKG questionnaire to Portuguese, in an effort to get closer to the original one. For example, in Chapter 2 one question was *While working on this case, at certain moments I felt unsure about my diagnostic hypothesis* and in Chapter 4 it was changed for *While working on this task, at certain moments I felt unsure if my answers were correct*. These adjustments might have allowed Chapter 4 to better capture students' AKG regarding the tasks they were asked to do. Assuming that deliberate reflection on clinical cases is a more challenging task than giving differential diagnosis, these results are in line with the knowledge deprivation theory: more challenge would lead to higher levels of AKG that would, finally, lead to higher SI.[9,10,12] In Chapter 2, however, when the analysis also considered case difficulty, SI and AKG went in opposite directions. This unexpected result might have been a consequence of the aforementioned very high level of difficulty of the difficult cases used in Chapter 2.

In conclusion, deliberate reflection upon clinical cases can, indeed, foster medical students' motivation to engage in subsequent study of scientific texts relevant to these cases. The results of Chapter 4 support the assumption that awareness of knowledge gaps underly situational interest, as it has been previously observed with different audiences,[10,12] at least if clinical cases are within a medium level of difficulty.

Research question 2: Would deliberate reflection improve the learning outcomes of scientific text study?

Deliberate reflection has already been shown to improve medical students' diagnostic accuracy *per se*, without any additional information provided to students,[22,23] an effect probably derived from reorganization of students' current knowledge into better knowledge structures.[24] Including additional information to students while they reflect upon cases, such as guiding reflection towards plausible diagnostic hypothesis or offering students the chance to study the reflective analysis performed by experts, has been shown to improve outcomes on diagnostic accuracy as well.[25-27] However, if deliberate reflection would foster subsequent learning of scientific texts was yet to be explored.

In Chapters 3 and 4, students who diagnosed clinical cases through deliberate reflection and, subsequently, studied a text relevant to these cases, showed higher scores on the test about this text, relative to those who diagnosed the cases giving differential diagnosis. These results suggest that deliberate reflection upon clinical cases can improve medical students' learning from scientific texts. These results are similar to what have been observed with other learning strategies, such as self-explanation.[18,19,21] Although reflection and self-explanation are different strategies, both require that one generates, individually and using only one's prior knowledge, explanations for facts or problems.

Results from Chapters 3 and 4 suggest that reflection can indeed improve learning outcomes of scientific text study and expand empirical evidence on the positive effects of problem-solving on learning to clinical reasoning and reflection.

Research question 3: Which mechanisms would underly a positive effect of deliberate reflection on learning of scientific texts: motivation and/or cognition?

In Chapter 3 deliberate reflection upon clinical cases fostered students' longer engagement with a text relevant to these cases and increased students' scores on the test about this text, relative to differential diagnosis. Could this positive effect on students' learning be solely a consequence of students' longer engagement with the text, which suggests a motivational explanation for the effect? Could reflection, on the other hand, have affected students' cognitive processing? Problem-solving strategies, such as PBL and self-explanation, have been shown to foster learning through cognitive mechanisms.[16-18] Activation of prior knowledge and elaboration of knowledge through discussion of problems with peers, for example, are positive effects of PBL.[28] Effects at an individual level, however, have also been observed. In an experiment in which students watched a recorded group discussion about a problem on physics, those who provided their own explanations to the problems while watching the discussion performed better on a test about the problem than those who watched it passively.[16] In an experiment with first-year medical students, Larsen *et al*[29] observed that students who worked, individually, with self-explanation while reading texts about neurology scored higher on a test about the subject than those who did not, even six months after the intervention.[19] These studies suggest that the attempts to explain problems facilitates elaboration and retention of new knowledge relevant to the problems. This could also be true for deliberate reflection, which also asks for arguments that justify a diagnostic hypothesis.

Indeed, in Chapter 4, an effect that could not be attributed to study-time emerged on students' test scores. In fact, the fact that students who reflected upon the cases showed higher scores on the test regardless of their study-time condition, and the fact that no interaction between diagnostic strategy and study-time emerged suggests that, at least for this experiment, cognitive mechanisms were responsible for the positive effect of reflection on students' test scores, not motivation. This finding supports the assumption that reflection has cognitive effects that are similar to other problem-solving strategies. A motivation effect, however, cannot be excluded as results from Chapters 2 and 4 showed that reflection fostered students' motivation, measured through a self-reported questionnaire on SI, and Chapter 3 showed similar results, except that the measurement of motivation was behavioral (study-time).

Chapter 5 of this thesis explored the potential of deliberate reflection to facilitate students' activation of and elaboration on prior knowledge. Students who diagnosed clinical cases before the knowledge test, either through deliberate reflection or by providing differential diagnosis, cited more information about the "typical clinical pictures" of the diseases represented in the cases relative to students who did not diagnose any cases before the knowledge test. Participants who diagnosed cases received no information about their diagnostic accuracy or about the cases' analysis during the experiment. Therefore, their descriptions represent their prior knowledge about the diseases, alongside with possible elaborations on this knowledge that took place during the diagnostic task. The higher scores on the test observed on participants who diagnosed cases, relative to those who did not diagnose any cases, expands the current knowledge on the mechanisms through which problem-solving helps learning to the field of clinical reasoning. There was also an advantage to the group who diagnosed the cases through deliberate reflection when compared to those who provided differential diagnosis. This latter result supports the assumption that deliberate reflection can facilitate the activation of and elaboration on students' prior knowledge to a larger extent than the more conventional diagnostic procedure of providing differential diagnosis. The difference between these groups was, however, marginally significant.

In conclusion, the results of this thesis suggest that the main mechanism that underly the positive effect of reflection on learning of scientific text is cognitive, and that the activation of/ elaboration on prior knowledge can be one of such mechanisms. A motivation effect, however, seem to play an additional role.

RESEARCH IMPLICATIONS

This thesis provides empirical evidence that reflection can foster medical students' situational interest, their engagement with reading scientific texts that are relevant to these cases and their learning from the texts. It adds supports, therefore, to the assumption that reflection is an important instructional approach to foster learning.[1,2] Chapter 4 also suggest that the main mechanism through which deliberate reflection helped students to learn from the scientific texts is cognitive. Activation of/elaboration on prior knowledge are probably among such mechanisms, as the results of Chapter 5 suggest. Motivational mechanisms, however, could not be entirely excluded as mediators of students' learning, as previously discussed.

It is also relevant to discuss the contradictory findings of Chapter 4, in which higher levels of SI did not translate into subsequent longer engagement with the scientific text. In the experiments of this thesis, study-time was taken as the behavioral measure of interest. However, there is evidence, especially from reading research, that interest can be also translated into focus and concentration.[7,29] For example, in an experiment with $8^{th}/9^{th}$ grade students working with short texts about different topics, Ainley *et al* observed that SI influenced students' reading choices, with those who showed higher SI on X-rays than on body image topics choosing to read texts addressing the former topic before the latter.[7] A similar effect might explain the results of Chapter 4, in which participants in the restricted-time condition were aware of the limited time they had to study the text, and possibly focused their attention on information that was new to them. Moreover, new information is processed in working-memory, and, as working memory has a limited capacity, people facing a lot of data have to choose what to focus on and what to ignore.[15] This "selective-engagement" with the study material might have played a role to students in the restricted-time groups of this experiment. A student who realized s/he did not know the role of prothrombin activity in the diagnosis of jaundice, for example, knowing that study time would be restricted, might have focused his/her attention on it, possibly ignoring other information s/he already knew, thereby making study time more efficient. These assumptions, however, require further investigation.

The knowledge deprivation theory of SI[9,10,12] also requires further investigation regarding clinical reasoning. The results of Chapter 4 suggest that AKG can be, indeed, an underlying mechanism of SI triggered by clinical cases' analysis, but results of Chapter 2 are more difficult to interpret given the unexpected very high level of difficulty observed in some cases. The difficulty of the cases might have hindered students' SI despite the perception of knowledge gaps that emerged when students diagnosed these cases.

LIMITATIONS

This thesis did not investigate other cognitive mechanisms that could have facilitated students' learning from the scientific texts. For example, information processing about the clinical cases, i.e., the clinical data described in the cases that was actually moved from students' sensory to working memory, where it can be processed with new information from the texts.[15]

Volunteers of this series of experiments were on their 4th or 5th-year of medical training and jaundice was the main subject of clinical cases, texts and tests used in three of the four experiments. Although there is no reason to believe that the results would have been different if different topics were used, generalizability should be cautious, particularly to more novice or senior students, for which future research should explore the effects of deliberate reflection on motivation and learning within these specific audiences.

PRACTICAL IMPLICATIONS

Medical teachers can use reflection to foster their students' motivation to engage in studying scientific texts that are relevant for clinical cases, and also facilitate students' learning from these texts. The deliberate reflection procedure used in experiments is relatively simple and does not require much time to be carried out. It is feasible to be incorporated to the practice with outpatients, for example. The teacher can ask students to reflect upon the patients they have just assisted, or, at the end of the day, to select one or a few cases that seem more relevant for individual reflection. The procedure can also be used in simulated scenarios and even in large classes: a teacher can ask students to reflect upon a case before a seminar, for example, to trigger students' interest for the topic to be discussed and to make learning more efficient. Teachers should also have in mind that the level of difficulty of tasks involving clinical cases can influence students' motivation and should avoid tasks that are too difficult for the students.

CONCLUSION

The main purpose of this series of experiments was to study reflection as an instructional approach both to motivate medical students to engage in studying scientific texts relevant to clinical cases and to make learning from these texts more efficient. It also explored which mechanisms could underly such positive effect of reflection on students' learning. Results suggest that deliberate reflection can indeed trigger students' interest and facilitate their learning from scientific texts. They also suggest that cognitive processing is the main mechanism through which reflection can facilitate learning, and that the activation of/elaboration on prior knowledge are among such mechanisms. They add empirical support to reflection as an important instructional approach and open windows for future research that can better inform medical students and teachers on when and how reflection upon clinical cases fosters learning.

REFERENCES

1. Sandars J. The use of reflection in medical education: AMEE Guide No. 44. *Med Teach.* 2009;31(8):685-95.
2. Mann, K, Gordon, J, MacLeod, A. (2009). Reflection and reflective practice in health professions education: a systematic review. *Adv Health Sci Educ*, 14, 595–621
3. Dewey J. *How we Think*. Boston: D. C. Haeth and Co.; 1913.
4. Schön, D. A. Understanding The Need for Artistry in Professional Education *in* Schön, D. A. (1987). *Education The Reflective Practitioner: Toward a New Design for Teaching and Learning in Professions.* San Francisco Jossey-Bass Higher Education Series, pp 1-20.
5. Mamede S, Schmidt HG, Penaforte JC. Effects of reflective practice on the accuracy of medical diagnosis. *Med Educ*. 2008;42:468-75.
6. Hidi S. Interest and its contribution as a mental resource for learning. *Rev Educ Res*. 1990;60(4): 549-571.
7. Ainley M, Hidi S, Berndorff D. Interest, Learning and the Psychological Processes that Mediate their Relationship. *J Educ Psychol*. 2002;94(3):545-61.
8. Hidi, S, Renninger A. The Four-Phase Model of Interest Development. *Educ Psychol*. 2006;41(2), 111-127.
9. Loewenstein G. The psychology of curiosity: a review and reinterpretation. *Psychol Bull*. 1994;116(1):75-98.
10. Rotgans JI, Schmidt HG. Situational interest and learning: Thirst for knowledge. *Learn Instr*. 2014;32:37-50.
11. Glogger-Frey I, Fleischer C, Gruny L, Kappich J, Renkl A. Inventing a solution and studying a worked solution prepare differently for learning from direct instruction. *Learn Instr*. 2015;39:72-87.
12. Rotgans JI, Schmidt HG. Situational interest and academic achievement in the active-learning classroom. *Learn Instr*. 2011;21:58-67.
13. Ormrod JE. Introduction to memory and cognition. In: Ormrod JE, editor. *Human Learning*. 5th ed ed. Boston: Pearson Education; 2008. p. 150-92
14. Mayer RE. Information Processing. In: Harris KR, Graham S, Urdan T, eds. *APA Educational Psychology Handbook. Vol. 1. Theories, Constructs and Critical Issues,* 1st ed. Washington DC: American Psychology Association 2012; 85-99.
15. Ormrod JE. Long-term memory storage and retrieval processes. In: Ormrod JE, editor. *Human Learning.* 7th ed ed. Boston: Pearson Education; 2016. p. 206-48.
16. van Blankenstein FM, Dolmans DHJM, van der Vleuten APM, Schmidt HG. Which cognitive processes support learning during small-group discussion? The role of providing explanations and listening to others. *Instruc Sci*. 2011;39:189-204.
17. Schmidt HG, Rotgans J, Yew EH. The process of problem-based learning: what works and why. *Med Educ*. 2011;45(8):792-806.
18. Chi MTH, de Leeuw N, Chiu M-H, LaVancher C. Eliciting self-explanations improves understanding. *Cognitive Sci*. 1994;18:439-77.
19. Larsen DP BA, Roediger III HL. Comparative effects of test-enhanced learning and self-explanation on long term retention. *Med Educ*. 2013;47:674-82.
20. Schmidt HG, De Volder M, De Grave WS, Moust JHC, Patel VL. Explanatory model in the processing of science text: the role of prior knowledge activation through small-group discussion. *J Educ Psychol*. 1989;81(4):610-9.
21. VanLehn K CM. A model of the self-explanation effect. *J Learn Sci*. 1992;2(1):1-59.

22. Mamede S, van Gog T, Moura AS, de Faria RMD, Peixoto J, Rikers RMJP, *et al*. Reflection as a strategy to foster medical student's acquisition of diagnostic competence. *Med Educ*. 2012;46:464-72.
23. Mamede S, van Gog T, Moura AS, de Faria RMD, Peixoto JM, Schmidt HG. How can students' diagnostic competence benefit most from practice with clinical cases? The effects os structured reflection on future diagnosis of the same and novel diseases. *Acad Med*. 2014;89(1):1-7.
24. Schmidt HG, Norman GR, Boshuizen HP. A cognitive perspective on medical expertise: theory and implications. *Acad Med*. 1990;65(10):611-621.
25. Ibiapina C, Mamede S, Moura A, Eloi-Santos S, van Gog T. Effects of free, cued and modelled reflection on medical students' diagnostic competence. *Med Educ*. 2014;48(8):796-805.
26. Mamede S, Figueiredo-Soares T, Eloi-Santos S, Faria RMD, Schmidt HG, van Gog T. Fostering novice students' diagnos- tic ability: The value of guiding deliberate reflection. *Med Educ*. 2019;53(6):628-637.
27. Fernandes, RAF, Malloy-Diniz, LF, de Vasconcellos, MC, Camargos, PAM, Ibiapina, C. Adding guidance to deliberate reflection improves medical student's diagnostic accuracy. *Med Educ*. 2021; 00: 1– 11.
28. Dolmans DHJM, Schmidt, HG. What Do We Know About Cognitive and Motivational Effects of Small Group Tutorials in Problem-Based Learning? *Adv Health Sci Educ*. 2006; 11:321–336
29. Sandars J. The use of reflection in medical education: AMEE Guide No. 44. *Med Teach.* 2009;31(8):685-95.
30. Mann, K, Gordon, J, MacLeod, A. (2009). Reflection and reflective practice in health professions education: a systematic review. *Adv Health Sci Educ*, 14, 595–621
31. Schiefele U, Krapp A. Topic interest and free recall of expository text. *Learn Indiv Diff*. 1996;8(2):141-60.

Samenvatting en Algemene discussie

Reflectie wordt beschouwd als een belangrijke educatieve benadering in het medisch onderwijs. Medische scholen, docenten en besturen vertrouwen op reflectie om het leren door studenten van scenario's die voorkomen in de bachelor tot na het afstuderen te ondersteunen.[1,2] Hoewel een sterk theoretisch kader de veronderstelling ondersteunt dat reflectie het leren bevordert,[3,4] zijn zowel empirisch bewijs voor de resultaten van reflectie momenteel beperkt, evenals de mechanismen die ten grondslag kunnen liggen aan de potentiële positieve effecten van reflectie in de context van het medisch onderwijs.

In deze scriptie wordt door middel van een reeks experimenten onderzocht of opzettelijke reflectie op klinische casussen het latere leren van wetenschappelijke teksten die relevant zijn voor de casus voor geneeskundestudenten vergemakkelijkt, en welke mechanismen, motivatie en/of cognitie, ten grondslag kunnen liggen aan dit potentiële positieve effect van reflectie.

Dit laatste Hoofdstuk geeft een samenvatting van de vier experimenten uit dit proefschrift, analyseert de resultaten in zijn geheel en bespreekt hoe ze medische docenten kunnen ondersteunen in hun dagelijkse praktijk en toekomstig onderzoek kunnen informeren.

SAMENVATTING VAN DE BELANGRIJKSTE BEVINDINGEN

Hoofdstuk 2 presenteert een experiment dat een vergelijking maakt van zelfgerapporteerde situationele interesse van vierdejaars geneeskundestudenten die worden veroorzaakt door de analyse van klinische casussen via twee verschillende benaderingen: opzettelijke reflectie, wat vereist dat studenten vóór en tegen verschillende mogelijke diagnoses moeten pleiten voordat ze tot een conclusie komen, en differentiële diagnose, een meer conventionele benadering waarbij de meest waarschijnlijke en twee alternatieve diagnoses zonder verdere overwegingen worden genoemd. Situationele interesse (SI) is een kortdurende en contextgerelateerde interesse waarvan is aangetoond dat deze wordt veroorzaakt door uitdaging en onzekerheid.[6-8] Vierdejaars geneeskundestudenten hebben beperkte klinische ervaring en we verwachtten daarom dat het diagnosticeren van klinische casussen, ongeacht de gebruikte benadering, tot op zekere hoogte, uitdagend zou zijn voor hen. Echter, we geloofden dat opzettelijke reflectie een grotere uitdaging zou zijn dan differentiële diagnose, en daarom dat het de SI van studenten in grotere mate zou stimuleren. Resultaten kwamen overeen met onze hypothese dat opzettelijke reflectie onder de studenten hogere maten van SI zou cultiveren, in vergelijking met het geven van differentiële diagnose, een veelbelovend effect van reflectie, aangezien is aangetoond dat SI een goede voorspeller is van betrokkenheid bij leren en leerresultaten. [6,7]

Hoofdstuk 2 verkende ook de interessante theorie over "kennisonthouding",[9,10] d.w.z. de veronderstelling dat SI wordt veroorzaakt door het besef van hiaten in de kennis over een bepaald onderwerp in de context van klinisch redeneren. Het zelfgerapporteerde besef van hiaten in de kennis (awareness of knowledge gaps; hierna AKG) werd ook gemeten door middel van een vragenlijst met 9 items.[11] Studenten die nadachten over de casussen vertoonden hogere scores op AKG, maar het verschil tussen de groepen was niet significant, behalve een post-hoc analyse met twee geselecteerde vragen uit de vragenlijst die de perceptie van studenten van hiaten in de kennis leken te meten (bijvoorbeeld, *door aan deze casus te werken verduidelijkte dat ik bepaalde dingen nog niet weet*), in tegenstelling tot enkele vragen die mogelijk de frustratie van studenten vastleggen bij het analyseren van de casussen (zoals, *ik zat vast bij het diagnosticeren van deze casus*).

Naast het vergelijken van SI- en AKG-scores tussen groepen volgens de diagnostische benadering die werd gebruikt om de casussen te diagnosticeren, vergeleek Hoofdstuk 2 deze scores ook op basis van de moeilijkheidsgraad van de casussen. Drie van de zes klinische casussen die door studenten werden gediagnosticeerd werden als gemakkelijke casussen beschouwd, waarbij deelnemers een gemiddelde initiële diagnosescore van 0.78 lieten zien, en drie werden als moeilijke casussen beschouwd met een gemiddelde initiële diagnosescore van 0.07 (bereik 0-1). De meest onverwachte en misschien interessante bevinding van dit experiment was dat, ondanks dat AKG hoger was voor de moeilijke casussen, was SI dat niet: ongeacht de diagnostische benadering waaronder studenten presteerden, vertoonden studenten hogere niveaus van SI voor de gemakkelijkere casussen. Gezien de belangrijke theorie over kennisonthoudings[9] verwachtten we het tegenovergestelde resultaat. Deze "ontkoppeling" van situationele interesse en besef van hiaten in de kennis staat in contrast met de bevindingen van Rotgans en Schmidt in de context van PBL,[10,12] maar is vergelijkbaar met wat Glogger-Frey *et al.*[11] observeerden bij leerlingen uit leerjaar 8 in een experiment dat de uitkomsten van verschillende onderwijsbenaderingen, het bedenken van een oplossing en het bestuderen van een canonieke oplossing, vergeleek met een probleem. Zoals het experiment van Glogger-Frey *et al.* suggereert ligt het besef van hiaten in de kennis mogelijk niet ten grondslag aan situationele interesse voor sommige doelgroepen in sommige casussen. Geneeskundestudenten die klinische casussen diagnosticeren kunnen een van dergelijke condities zijn. We kunnen echter niet uitsluiten dat de onverwachte moeilijkheidsgraad van de moeilijke casussen in dit experiment, veel hoger dan we hadden geanticipeerd, de interesse van studenten voor deze casussen zou kunnen hebben belemmerd.

Hoofdstuk 3 beschrijft een experiment dat is ontworpen om te controleren of de positieve resultaten van opzettelijke reflectie op de situationele interesse van studenten, gemeten als een zelfgerapporteerde vragenlijst, zich reproduceren wanneer interesse werd gemeten aan de hand van waarneembaar gedrag: de tijd die studenten besteedden aan het bestuderen van een wetenschappelijke tekst die relevant is voor de casussen. De hypothese dat een grotere interesse zich zou vertalen in meer leren werd ook getest. In een elektronische omgeving diagnosticeerden studenten klinische casussen met geelzucht als de belangrijkste klinische bevinding, óf door opzettelijke reflectie, óf door differentiële diagnose. De casussen die zijn gekozen voor dit experiment werden geschat op een matige moeilijkheidsgraad voor de studenten, om het negatieve effect van hogere moeilijkheidsgraden op de interesse van leerlingen dat werd waargenomen in het vorige experiment te voorkomen. Vervolgens lazen alle studenten een tekst over de differentiële diagnose van geelzucht, gebaseerd op de casussen die ze zojuist hadden gediagnosticeerd, voor zo lang zij wilden. De betrokkenheid van studenten bij deze tekst, gemeten aan de hand van de tijd die ze besteedden aan het lezen, werd geregistreerd. Tenslotte deden de studenten een cued-recall-test over de informatie die aan hen in de tekst werd gepresenteerd. De scores van de studenten op deze test werden genomen als maatstaf voor het leren van studenten over de tekst. Zoals we verwachtten en in overeenstemming met de bevindingen van experimenten met lezen en problemen oplossen, waarvan is aangetoond dat uitdaging en onverwachtheid interessante triggers zijn, [6,7,12] waren studenten die bewust reflecteerden over de casussen langer bezig met de tekst en lieten zij hogere scores op de test zien in vergelijking met degenen die differentiële diagnose gaven. Deze bevindingen bieden empirische ondersteuning dat opzettelijke reflectie inderdaad het leren van wetenschappelijke teksten kan bevorderen.

Twee mechanismen zouden het positieve effect van opzettelijke reflectie op het leren van studenten dat in dit experiment werd waargenomen kunnen verklaren: motivatie en/of cognitie. Als het onderliggende mechanisme van het positieve effect van positieve reflectie op het leren van studenten puur motiverend was, zouden leerlingen die hebben gereflecteerd over casussen meer over de tekst hebben geleerd dan degenen die differentiële diagnose gaven, puur omdat ze meer tijd hebben besteed aan het lezen van de tekst. Anderzijds heeft bewuste reflectie de verwerking van de in de tekst gepresenteerde informatie door de studenten kunnen vergemakkelijken, door bijvoorbeeld hun voorkennis te activeren en daardoor de opname van nieuwe informatie door de studenten te vergemakkelijken.[13-15] Een derde hypothese is dat beide mechanismen, motivatie en cognitie, bijdroegen tot betere leerresultaten, gevoed door reflectie.

Hoofdstuk 4 onderzoekt welke mechanismen ten grondslag liggen aan het positieve effect van reflectie op het leren van studenten dat is waargenomen in Hoofdstuk 3: motivatie en/of cognitie. Daartoe werden geneeskundestudenten willekeurig ingedeeld in een van vier experimentele groepen. Groepen 1 en 2 diagnosticeerden klinische casussen met geelzucht als de belangrijkste klinische bevinding door middel van opzettelijke reflectie. Vervolgens las groep 1 zonder tijdsbeperkingen een tekst voor over de geelzucht diagnose. Groep 2 las dezelfde tekst voor, maar ze hadden daar een beperkte hoeveelheid tijd voor. Groepen 3 en 4 diagnosticeerden dezelfde klinische casussen door middel van de differentiële diagnose. Vervolgens last groep 3 een tekst over de diagnose geelzucht zonder tijdsbeperking, terwijl groep 4 maar een beperkte hoeveelheid tijd had om deze te lezen. We hebben dus twee variabelen gemanipuleerd: diagnostische benadering (opzettelijke reflectie versus differentiële diagnose) en studietijd (vrije tijd versus beperkte tijd). Ten slotte deden alle deelnemers een herinneringstest (recall-test) over de tekst die ze bestudeerden. De scores van studenten op deze test werden genomen als maatstaf voor het leren over de tekst. De resultaten lieten een hoofdeffect van de diagnostische benadering zien, geen effect van studietijd en geen interactie tussen diagnostische benadering en studietijd: studenten die reflecteerden over de casussen presteerden vergelijkbaar en beter dan degenen die differentiële diagnose gaven, ongeacht de studietijd omstandigheden waaronder ze presteerden. Deze resultaten suggereren dat het onderliggende mechanisme van het positieve effect van opzettelijke reflectie op de testscores van studenten cognitief was en niet motiverend. Het lijkt erop dat bewuste reflectie werkt via mechanismen zoals de verwerking van informatie over de klinische casussen, de activering van voorkennis over de ziekten en/of het proces van het integreren van nieuwe informatie die in de teksten wordt gepresenteerd in reeds bestaande kennisstructuren.

De rol van het motiverende mechanisme kon echter niet worden uitgesloten. Er werden ook metingen gedaan van de zelfgerapporteerde situationele interesse (SI) en het besef van hiaten in de kennis (AGK) van studenten. Zoals we verwachtten, en vergelijkbaar met wat werd waargenomen in Hoofdstuk 2, waren zowel de SI- als de AKG-scores hoger voor studenten die reflecteerden over de casussen, vergeleken met degenen die differentiële diagnose gaven, wat suggereert dat motivatie ook een rol zou kunnen hebben gespeeld bij de prestaties van studenten op de test. Deze keer was het verschil tussen de groepen significant voor beide metingen, inclusief de AKG-scores waarbij er rekening gehouden werd met alle vragen, zelfs de vragen die gevoelens van frustratie zouden kunnen hebben opgevangen in plaats van een meer emotioneel vrije perceptie van hiaten in de kennis.

Hoofdstuk 5 bouwt voort op de resultaten van Hoofdstuk 4, waarin werd gesuggereerd dat het belangrijkste mechanisme dat ten grondslag ligt aan het positieve effect van reflectie op het leren van studenten cognitief is. Om te testen of een van dergelijke mechanismen de activering en uitwerking van voorkennis zou kunnen zijn, vergeleek een experiment de beschrijving van studenten van het "typische klinische beeld" van ziekten, welke als maatstaf werd gebruikt voor hun kennis over de ziekten, na de diagnose van klinische casussen van deze ziekten door middel van verschillende diagnostische strategieën. Voordat deze kennistest gedaan werd, diagnosticeerde de ene groep studenten klinische casussen door middel van bewuste reflectie en de andere groep studenten door middel van differentiële diagnose. Een derde groep beschreef de typische ziektebeelden van de ziekten zonder vooraf een klinische casus te diagnosticeren. Aangezien eerder is aangetoond dat andere probleemoplossende strategieën voorkennis activeren,[13-17] verwachtten we dat studenten die die klinische casussen diagnosticeerden beter presteerden op de kennistest dan de groep die geen casussen diagnosticeerde. Als bewuste reflectie inderdaad de activering van de voorkennis van studenten beter faciliteert dan het stellen van een differentiaaldiagnose, dan is er ook een voordeel voor de eerste groep. De resultaten waren in lijn met onze verwachtingen, waarbij studenten die klinische casuïstiek diagnosticeerden beter presteerden op de kennistest dan studenten die geen casuïstiek diagnosticeerden. Studenten die bewust redeneren toepasten tijdens het diagnosticeren presteerden ook beter dan studenten die de instructie kregen om differentiaal diagnoses te stellen, hoewel dit verschil marginaal significant was.

ALGEMENE DISCUSSIE

De reeks experimenten waaruit dit proefschrift bestaat suggereert dat bewuste reflectie op klinische casussen een relevante educatieve benadering kan zijn om geneeskundestudenten te motiveren om vervolgens wetenschappelijke teksten te bestuderen die relevant zijn voor die casussen (Hoofdstukken 2 en 3). Het suggereert ook dat opzettelijke reflectie, vergeleken met differentiële diagnose, de leerresultaten van het bestuderen van dergelijke teksten kan verbeteren (Hoofdstukken 3 en 4), en dat cognitieve verwerking van informatie het mechanisme is dat grotendeels verantwoordelijk is voor dit positieve effect van reflectie (Hoofdstuk 4). Tot slot werd er gesuggereerd dat het activeren en uitwerken van voorkennis een cognitief mechanisme is dat wordt gefaciliteerd door reflectie (Hoofdstuk 5). In deze sectie wordt elke onderzoeksvraag besproken in het licht van de resultaten van de experimenten.

Onderzoeksvraag 1: Zou bewuste reflectie op klinische casussen de motivatie van geneeskundestudenten bevorderen om later wetenschappelijke teksten te bestuderen die relevant zijn voor deze casussen?

Reflectie, door Dewey[3] gedefinieerd als de bewuste poging om een uitdagend probleem systematisch de analyseren om er een goede oplossing voor te vinden, wordt beschouwd als een relevante educatieve benadering voor het leren. Het is een schappelijke aanname, gezien het empirische bewijs geleverd door leesonderzoek op gebieden anders dan medisch onderwijs, wat heeft laten zien dat onverwachtheid, nieuwigheid en uitdagingen interessante triggers zijn, met name de kortstondige, context gerelateerde interesse die situationele interesse wordt genoemd.[6-10] Ook is de verwachting dat reflectie wordt veroorzaakt door onverwachtheid, nieuwigheid en uitdagingen.[3] In de medische praktijk zou dit het geval zijn wanneer bijvoorbeeld een student of een arts moeite heeft om er achter te komen wat de meest waarschijnlijke diagnose voor een patiënt is en vervolgens niet in voorbarige conclusies "springt". Reflecteren op klinische casussen zou daarom studenten kunnen motiveren om deel te nemen aan leeractiviteiten.

In Hoofdstukken 2 en 4 was de situationele interesse van geneeskundestudenten, gemeten met een zelfgerapporteerde vragenlijst, groter wanneer ze klinische casussen diagnosticeerden door middel van opzettelijke reflectie, in vergelijking met de meer conventionele benadering van het geven van differentiële diagnose, met een grote effectgrootte. Deze SI-vragenlijst meet zowel emoties, met vragen zoals *ik vond het leuk om met deze casus te werken,* en de intentie om te leren, zoals gemeten door vragen als *ik wil meer te weten komen over deze casus*, en er is gebleken dat dat dit een goede voorspeller is voor de betrokkenheid van studenten bij leeractiviteiten.[6,7,12] Inderdaad, in Hoofdstuk 3, waarin de interesse van studenten werd gemeten via studietijd, waren studenten die reflecteerden over de casussen langer bezig met een relevante tekst voor de casussen dan degenen die een differentiële diagnose gaven. Deze bevindingen suggereren dat opzettelijke reflectie op klinische casussen de interesse van geneeskundestudenten voor wetenschappelijke teksten die relevant zijn voor deze casussen kan opwekken en behouden. Er was al aangetoond dat uitdagende problemen interesse wekken bij middelbare- en economiestudenten.[12] De noviteit in de experimenten van dit proefschrift is dat dit bewijs wordt uitgebreid tot het medisch klinische redeneer- en reflectieveld.

In Hoofdstuk 4 bevorderde reflectie binnen de groepen die geen tijdsbeperking hadden om de wetenschappelijke tekst te bestuderen ook een langere betrokkenheid bij de tekst vergeleken met differentiële diagnose, maar dit verschil was niet significant. Een mogelijke verklaring voor deze onverwachte uitkomst is dat vrijwilligers van dit experiment verder gevorderd waren in het medisch onderwijs (5[e] versus 4[e] jaar) en hogere initiële diagnostische nauwkeurigheidsscores vertoonden (wat suggereert dat er meer kennis over de casussen is) in vergelijking met de vrijwilligers van Hoofdstuk 3. Onze tijdsbeperking, gebaseerd op de resultaten van Hoofdstuk 3, zou de doelgroep van Hoofdstuk 4 genoeg tijd hebben gegeven om materiaal over een onderwerp door te nemen dat waarschijnlijk informatie biedt die hen meer vertrouwd was, vergeleken met hun minder gevorderde studiegenoten. We kunnen echter niet uitsluiten dat de steekproefomvang van Hoofdstuk 4 niet groot genoeg was om significante verschillen met betrekking tot studietijd naar voren te laten komen.

Het is noemenswaardig dat Hoofdstuk 2 liet zien dat dit positieve effect van reflectie op de motivatie van studenten kan worden belemmerd als klinische casussen te moeilijk zijn voor studenten. In tegenstelling tot onze verwachtingen was de SI van studenten hoger voor gemakkelijkere casussen dan voor moeilijkere casussen. Studenten waren wellicht van mening dat de kennis benodigd voor het oplossen van de moeilijke casussen, waarvoor hun initiële diagnostische nauwkeurigheid minder dan 10% was, buiten hun bereik lag. Een te grote uitdaging kan een taak voorbij een optimaal niveau van incongruentie[9] brengen en een verstoren effect hebben op de interesse van studenten.

In een complementaire analyse verkenden Hoofdstuk 2 en 4 ook de kennisonthoudingsmethode[9] voor SI. Naast metingen van SI werden ook metingen van het besef van studenten over de hiaten in de kennis verkregen. In Hoofdstuk 2 vertoonden studenten die casussen diagnosticeerden door middel van opzettelijke reflectie hogere scores op AKG. Het verschil tussen de groepen was echter niet significant, afgezien van een post-hoc analyse met twee vragen die het bewustzijn van hiaten in kennis op zich leken te meten, in tegenstelling tot andere vragen die mogelijk de frustratie van studenten over klinische taken waarmee ze werkten opvangen. Aan de andere kant kwam in Hoofdstuk 4 een significant effect van opzettelijke reflectie naar voren voor zowel SI als AKG, dit keer waarbij alle AKG-vragen werden behandeld. Hoofdstukken 2 en 4 gebruikten dezelfde vragenlijsten vertaald naar het Portugees vanuit het Engels (SI) en Duits (AKG). In Hoofdstuk 4 zijn echter kleine aanpassingen gemaakt in de vertaling van de AKG-vragenlijst naar Portugees, om dichter bij de oorspronkelijke te komen. In Hoofdstuk 2 was bijvoorbeeld een vraag *Terwijl ik aan deze casus werkte, voelde ik me op bepaalde momenten onzeker over mijn diagnostische hypothese*, en in Hoofdstuk 4 werd deze veranderd in *Tijdens het werken aan deze taak voelde ik me op bepaalde momenten onzeker of mijn antwoorden juist waren*. Deze aanpassingen zorgden er mogelijk voor dat Hoofdstuk 4 beter de AKG van studenten met betrekking tot de taak die ze moesten doen kon vastleggen. Ervan uitgaande dat opzettelijke reflectie op klinische casussen een grotere uitdaging is dan het geven van differentiële diagnose, zijn deze resultaten in overeenstemming met de kennisonthoudingstheorie: meer uitdaging zou leiden tot hogere niveaus van AKG, wat uiteindelijk zou leiden tot een hogere SI.[9,10,12] In Hoofdstuk 2 gingen SI en AKG echter in tegengestelde richting, toen de analyse ook rekening hield met de moeilijkheidsgraad van de casus. Dit onverwachte resultaat zou een gevolg kunnen zijn van de eerdergenoemde zeer hoge moeilijkheidsgraad van de moeilijke casussen die in Hoofdstuk 2 worden gebruikt.

Concluderend kan een bewuste reflectie op klinische casussen inderdaad de motivatie van geneeskundestudenten bevorderen om later wetenschappelijke teksten te bestuderen die relevant zijn voor deze casussen. De resultaten van Hoofdstuk 4 ondersteunen de veronderstelling dat het besef van hiaten in de kennis ten grondslag ligt aan de situationele interesse, zoals eerder is waargenomen bij verschillende doelgroepen,[10,12] tenminste als klinische casussen zich binnen een gemiddelde moeilijkheidsgraad bevinden.

Onderzoeksvraag 2: Zou bewuste reflectie de leerresultaten van wetenschappelijke tekststudie verbeteren?

Er is reeds aangetoond dat opzettelijke reflectie de diagnostische nauwkeurigheid van geneeskundestudenten op zich verbetert, zonder dat er aanvullende informatie aan studenten wordt verstrekt,[22,23] een effect dat waarschijnlijk wordt afgeleid van de reorganisatie van al vergaarde kennis naar betere kennisstructuren.[24] Van het opnemen van aanvullende informatie aan studenten terwijl ze reflecteren op casussen, zoals het begeleiden van reflectie naar plausibele diagnostische hypothesen of zoals de mogelijkheid aan studenten te geven om de reflectieve analyse uitgevoerd door experts te bestuderen, is ook aangetoond dat het de resultaten op diagnostische nauwkeurigheid verbetert.[25-27] Of opzettelijke reflectie het latere leren van wetenschappelijke teksten zou vergemakkelijken moest echter nog worden onderzocht.

In Hoofdstukken 3 en 4, lieten studenten die klinische casussen diagnosticeerden door middel van opzettelijke reflectie en vervolgens een tekst bestudeerden die relevant was voor deze casussen hogere scores op de test over deze tekst zien dan degenen die de casussen diagnosticeerden door differentiële diagnose te geven. Deze resultaten suggereren dat bewuste reflectie op klinische casussen het leren van wetenschappelijke teksten door geneeskundestudenten kan verbeteren. Deze resultaten zijn vergelijkbaar met wat is waargenomen met andere leerstrategieën, zoals zelfverklaring.[18,19,21] Hoewel reflectie en zelfverklaring verschillende strategieën zijn, vereisen beide dat men, individueel en met alleen de eigen voorkennis, verklaringen voor feiten en problemen genereert.

Resultaten uit Hoofdstukken 3 en 4 suggereren dat reflectie inderdaad de leerresultaten van wetenschappelijke tekststudie kan verbeteren en empirisch bewijs over de positieve effecten van probleemoplossing op leren kan uitbreiden naar klinisch redeneren en reflecteren.

Onderzoeksvraag 3: Welke mechanismen zouden ten grondslag liggen aan een positief effect van bewuste reflectie op het leren van wetenschappelijke teksten: motivatie en/of cognitie?

In hoofdstuk 3 stimuleerde de doelbewuste reflectie op klinische casussen studenten om langer betrokken te raken bij een tekst die relevant was voor deze casussen en verhoogde de score van leerlingen op de test over deze tekst, in vergelijking met differentiële diagnose. Zou dit positieve effect op het leren van studenten uitsluitend een gevolg kunnen zijn van een langere betrokkenheid van leerlingen bij de tekst, wat een motiverende verklaring voor het effect suggereert? Kan reflectie daarentegen de cognitieve verwerking van studenten hebben beïnvloed? Er is aangetoond dat probleemoplossende strategieën, zoals PBL en zelfverklaring, het leren bevorderen via cognitieve mechanismen.[16-18] Activering van voorkennis en uitwerking van kennis door bijvoorbeeld problemen te bespreken met leeftijdsgenoten zijn positieve effecten van PBL.[28] Er zijn echter ook effecten op individueel niveau waargenomen. In een experiment waarbij studenten naar een opgenomen groepsdiscussie over een natuurkundig probleem keken, presteerden degenen die tijdens het kijken naar de discussie hun eigen verklaring voor de problemen gaven, beter op een test over het probleem dan degenen die passief keken.[16] Larsen *et al.*[29] constateerden in een experiment met eerstejaars geneeskundestudenten dat studenten die individueel met zelfverklaring werkten tijdens het lezen van teksten over neurologie hoger scoorden op een toets over het onderwerp dan degenen die dat niet deden, zelfs zes maanden na de interventie.[19] Deze onderzoeken suggereren dat pogingen om problemen te verklaren de uitwerking en het behoud van nieuwe kennis die relevant is voor de problemen vergemakkelijken. Dit zou ook kunnen gelden voor opzettelijke reflectie, wat ook vraagt om argumenten die een diagnostische hypothese rechtvaardigen.

In Hoofdstuk 4 kwam inderdaad een effect naar voren dat niet kon worden toegeschreven aan studietijd op de testscores van studenten. Het feit dat studenten die reflecteerden over de casussen hogere scores lieten zien op de test, ongeacht hun studietijdconditie, en het feit dat er geen interactie tussen diagnostische strategie en studietijd naar voren kwam, suggereert dat, tenminste voor dit experiment, cognitieve mechanismen verantwoordelijk waren voor het positieve effect van reflectie op de testscores van studenten, niet motivatie. Deze bevinding ondersteunt de aanname dat reflectie cognitieve effecten heeft die vergelijkbaar zijn met andere probleemoplossende strategieën. Een motivatie-effect kan echter niet worden uitgesloten, aangezien de resultaten uit Hoofdstukken 2 en 4 lieten zien dat reflectie de motivatie van leerlingen bevorderde, gemeten met een zelfgerapporteerde vragenlijst over SI, en Hoofdstuk 3 liet vergelijkbare resultaten zien, behalve dat de meting van motivatie gedragsmatig was (leertijd).

Hoofdstuk 5 van dit proefschrift verkent het potentieel van bewust redeneren om de activering en verwerking van de voorkennis van studenten te bevorderen. Studenten die klinische casuïstiek diagnosticeerden, ongeacht de diagnostische methode die ze daarbij gebruikten, scoorden hoger op een test die hen vroeg om typische klinische beelden te beschrijven van ziektes die ze zojuist gediagnosticeerd hadden, in vergelijking met studenten die voorafgaand aan de test niet waren blootgesteld aan klinische casuïstiek. Dit verschil duidde op een groot effect. Verder presteerden studenten die bewust redeneren toepasten tijdens het diagnosticeren beter dan studenten die differentiaal diagnoses opstelden, hoewel dit verschil marginaal significant was. Dit ondersteunt onze hypothese dat één van de cognitieve mechanismen waardoor bewust redeneren op klinische casuïstiek het leren van relevante wetenschappelijke teksten faciliteert, de activering en verwerking van de voorkennis van studenten is.

Concluderend suggereren de resultaten van dit proefschrift dat het belangrijkste mechanisme dat ten grondslag ligt aan het positieve effect van reflectie op het leren van wetenschappelijke tekst cognitief is, en dat het activeren/uitwerken op voorkennis een van dergelijke mechanismen kan zijn. Een motivatie-effect lijkt echter een extra rol te spelen.

IMPLICATIES VOOR ONDERZOEK

Dit proefschrift levert empirisch bewijs dat reflectie de situationele interesse van geneeskundestudenten, hun betrokkenheid bij het lezen van wetenschappelijke teksten die relevant zijn voor deze casussen en hun leren van de teksten kan bevorderen. Het voegt daarom ondersteuning toe aan de aanname dat reflectie een belangrijke educatieve benadering is om het leren te bevorderen.[1,2] Hoofdstuk 4 suggereert ook dat het belangrijkste mechanisme waardoor opzettelijke reflectie studenten hielp om van de wetenschappelijke teksten te leren, cognitief is. Activering/uitwerking van voorkennis is waarschijnlijk een van die mechanismen, zoals de tussentijdse resultaten van Hoofdstuk 5 suggereren. Zoals eerder besproken konden motivatiemechanismen echter niet volledig worden uitgesloten als bemiddelaars van het leren van studenten.

Daarnaast is het relevant om de tegenstrijdige bevindingen van Hoofdstuk 4 te bespreken, waarin hogere niveaus van SI zich niet vertaalden in een daaropvolgende langere betrokkenheid bij de wetenschappelijke tekst. In de experimenten van dit proefschrift werd de studietijd als gedragsmaatstaf van belang genomen. Er zijn echter aanwijzingen, vooral uit leesonderzoek, dat interesse ook kan worden vertaald in focus en concentratie.[7,29] In een experiment met leerlingen uit leerjaar 8 en 9 die werkten met korte teksten over verschillende onderwerpen, merkten Ainley *et al.* op dat SI de leeskeuzes van studenten beïnvloedde, waarbij degenen die een hogere SI lieten zien op röntgenfoto's dan op onderwerpen over beelden van het lichaam eerder kozen om teksten te lezen over het eerste onderwerp dan over het laatste onderwerp.[7] Een soortgelijk effect zou de resultaten van Hoofdstuk 4 kunnen verklaren, waarin deelnemers in de beperkte tijd conditie zich bewust waren van de tijdslimiet die ze hadden om de tekst te bestuderen en mogelijk hun aandacht richtten op informatie die nieuw voor hen was. Bovendien wordt nieuwe informatie in het werkgeheugen verwerkt en omdat het werkgeheugen een beperkte capaciteit heeft moeten mensen die met veel informatie worden geconfronteerd kiezen waar ze zich op willen concentreren en wat ze negeren.[15] Deze "selectieve betrokkenheid" bij het studiemateriaal zou een rol kunnen hebben gespeeld voor studenten in de beperkte tijdgroepen van dit experiment. Een leerling die zich realiseerde dat hij/zij de rol van protrombineactiviteit bij de diagnose van geelzucht niet kende, bijvoorbeeld wetende dat de studietijd beperkt zou zijn, zou zijn/haar aandacht erop kunnen hebben gericht, waarbij hij mogelijk andere informatie negeerde die hij/zij al wist, waardoor de studietijd efficiënter werd. Deze aannames vergen echter nader onderzoek.

De kennisonthoudingstheorie van SI[9,10,12] vereist ook nader onderzoek naar klinisch redeneren. De resultaten van Hoofdstuk 4 suggereren dat AKG inderdaad een onderliggend mechanisme van SI kan zijn dat wordt opgewekt door analyse van klinische casussen, maar de resultaten van Hoofdstuk 2 zijn moeilijker te interpreteren gezien de onverwachte, zeer hoge moeilijkheidsgraad die bij sommige casussen werd waargenomen. De moeilijkheidsgraad van de casussen kan de SI van de studenten hebben belemmerd, ondanks de perceptie van hiaten in de kennis die naar voren kwamen wanneer studenten deze casussen diagnosticeerden.

BEPERKINGEN

Dit proefschrift heeft geen andere cognitieve mechanismen die het leren van studenten wetenschappelijker teksten hadden kunnen vergemakkelijken onderzocht. Bijvoorbeeld, informatieverwerking over de klinische casussen, d.w.z. klinische informatie die beschreven werd in de casussen die daadwerkelijk van het sensorische naar het werkgeheugen is verplaatst, waar het kon worden verwerkt met nieuwe informatie uit de teksten.[15]

Vrijwilligers van deze serie experimenten zaten in hun 4^{e} of 5^{e} jaar van de medische opleiding en geelzucht was het belangrijkste onderwerp van de klinische casussen, teksten en testen die in drie van de vier experimenten werden gebruikt. Hoewel er geen reden is om aan te nemen dat de resultaten anders zouden zijn geweest als er verschillende onderwerpen zouden zijn gebruikt, moet men voorzichtig zijn met generaliseerbaarheid, met name voor meer beginnende of oudere studenten waarvoor toekomstig onderzoek de effecten van opzettelijke reflectie op motivatie en leren binnen deze specifieke doelgroepen zou moeten onderzoeken. Tot slot is het experiment beschreven in Hoofdstuk 5 nog lopende en tussentijdse gegevens moeten zorgvuldig geïnterpreteerd worden.

PRAKTISCHE IMPLICATIES

Docenten geneeskunde kunnen reflectie gebruiken om de motivatie van hun studenten om wetenschappelijke teksten te bestuderen die relevant zijn voor klinische casussen te stimuleren, en ook om studenten te helpen bij het leren van deze teksten. De opzettelijke reflectieprocedure die in deze experimenten wordt gebruikt is relatief eenvoudig en kost niet veel tijd om uit te voeren. Het is bijvoorbeeld mogelijk om deze op te nemen met poliklinische patiënten. De docent kan studenten vragen om te reflecteren over de patiënten die ze zojuist hebben geholpen, of aan het eind van de dag een of enkele casussen te selecteren die relevanter lijken voor individuele reflectie. De procedure kan ook worden gebruikt in gesimuleerde scenario's en zelfs in grote klassen: een docent kan bijvoorbeeld studenten vragen om vóór een seminar over een casus te reflecteren, om de interesse van studenten voor het te bespreken onderwerp op te wekken en om het leren efficiënter te maken. Docenten moeten er ook rekening mee houden dat de moeilijkheidsgraad van taken waar klinische casussen bij betrokken zijn de motivatie van studenten kan beïnvloeden en taken die te moeilijk zijn voor studenten vermijden.

CONCLUSIE

Het belangrijkste doel van deze reeks experimenten was om reflectie te bestuderen als een educatieve benadering, zowel om geneeskundestudenten te motiveren om wetenschappelijke teksten die relevant zijn voor klinische casussen te bestuderen, en om het leren van deze teksten efficiënter te maken. Het onderzocht ook welke mechanismen ten grondslag kunnen liggen aan een dergelijk positief effect van reflectie op het leren van studenten. De resultaten suggereren dat opzettelijke reflectie inderdaad de interesse van studenten kan opwekken en het leren van wetenschappelijke teksten kan vergemakkelijken. Ze suggereren ook dat cognitieve verwerking het belangrijkste mechanisme is waardoor reflectie het leren kan vergemakkelijken, en dat de activering/uitwerking van voorkennis een van die mechanismen is. Ze voegen empirische ondersteuning toe aan reflectie als een belangrijke educatieve benadering en openen vensters voor toekomstig onderzoek dat geneeskundestudenten en docenten beter kan informeren over wanneer en hoe reflectie op klinische casussen het leren bevordert.

Curriculum Vitae and Publications

CURRICULUM VITAE

Ligia Ribeiro was born in Belo Horizonte, Brazil, in 1977. She graduated from Medical School in 2001 and completed her training in Internal Medicine in 2004. Just after completing this training program, she began working as a medical doctor at a local, small and recently inaugurated public facility, Contagem City Hospital (HMC). At HMC she joined the team that implemented the Hospitals' first medical post-graduation training programs, including Internal Medicine, for which she was responsible until 2015. In 2007, she started working as a medical teacher at UNIFENAS Medical School, where she currently teaches undergraduate and master students and collaborates with curriculum and faculty development. In 2013 she got a Masters' degree in Health Professions Education (MHPE) from Maastricht University with a dissertation on students' reflective portfolios. In 2015 she became a PhD candidate at Erasmus MC, at the Institute of Medical Education Research Rotterdam (iMERR), in clinical reasoning research. In this same year she discontinued her work with patients to be full time dedicated to teaching and researching. During her PhD, she visited Rotterdam four times to plan her research projects, analyze its results, draft manuscripts and enjoy the scientific activities of the iMERR. From 2016-2019, she also presented her PhD results in four European conferences. Unfortunately, Covid-19 prevented her to visit Rotterdam more times, but on the other hand allowed her to join the iMERR scientific meetings, now online, more often. In 2020 she joined the faculty of UNIFENAS' MHPE and currently supervises 4 master students. She recently collaborated with medical education journals as a referee.

PUBLICATIONS

1. **Ribeiro, L M C**; Mamede, S; de Brito, E. M.; Sampaio, A. M.; Faria, R. M. D.; Schmidt, H G. Exploring mechanisms underlying learning from deliberate reflection: An experimental study. *Medic Educ*, 2021, v.55, 404-412
2. Moreira, L. M.; Sampaio, A. M.; de Brito, E. M.; Junqueira, F. F. P.; **Ribeiro, L. M. C.** Impact of a Faculty Development Strategy Aiming at Mitigating Erosions in a PBL-Based Medical School *J Problem Based Learn in Higher Educ*, 2021, v.9, 69-96
3. Mamede, S; de Carvalho-Filho, M A; de Faria, R M D; Franci, D; Nunes, M P T; **Ribeiro, L M C**; Biegelmeyer, J; Zwaan, L; Schmidt, H G. 'Immunising' physicians against availability bias in diagnostic reasoning: a randomised controlled experiment. *BMJ Quality & Safety*, 2020, v.29, 550-559.
4. **Ribeiro, L M C**; Moura, A S. 'Hands-on' ideas to provide student-targeted clinical reasoning educational interventions. *Medic Educ*, 2020 v.54, 680-682
5. **Ribeiro, L M C**; Mamede, S; de Brito, E. M.; Sampaio, A. M.; de Faria, R. M. D.; Schmidt, H G. Effects of deliberate reflection on students' engagement in learning and learning outcomes. *Medic Educ*, 2019, v.53, 390-397
6. **Ribeiro, L M C**; Mamede, S; Sampaio, A. M.; de Brito, E. M.; Faria, R. M. D.; Schmidt, H G. Effect of reflection on medical students' situational interest: an experimental study. *Medic Educ*, 2018, v.52, 488-496

PRESENTATIONS

1. Ribeiro, L. M. C.; Mamede, S.; Sampaio, A. M.; de Brito, E. M.; Faria, R. M. D.; Schmidt, H. G. The effects of Reflection on clinical problems on medical students' awareness of knowledge gaps and situational interest. Poster presentation. AMME conference: Barcelona: Spain (2016 August 27-29).
2. Ribeiro, L. M. C.; Mamede, S; Sampaio, A. M.; de Brito, E. M.; Faria, R. M. D.; Schmidt, H. G. The effects of Reflection on clinical problems on medical students' awareness of knowledge gaps and situational interest. Rogano Meeting: Barcelona: Spain (2016, September 1-2).
3. Ribeiro, L. M. C.; Mamede, S; Sampaio, A. M.; de Brito, E. M.; Faria, R. M. D.; Schmidt, H. G. The effects of reflection on clinical problems on medical students' engagement in a learning activity and learning outcomes. Short communication. AMEE conference: Basel: Switzerland (2018, August 25-29).
4. Ribeiro, L. M. C; Mamede, S; Sampaio, A. M.; de Brito, E. M.; Faria, R. M. D.; Schmidt, H. G. Is it about getting interested? Exploring how deliberate reflection might help medical students learn from clinical cases. Short communication. AMEE conference: Vienna: Austria (2019, August 24-28)

5. Ribeiro, L M C; Mamede, S; Sampaio, A. M.; de Brito, E. M.; Faria, R. M. D.; Murta, C. B.; Schmidt, H. G. Can deliberate reflection on clinical cases help medical students to activate and elaborate on their prior knowledge? Short communication. AMEE conference: Lyon, France (2022, August 27-31)

Acknowledgments

This thesis is the result of a collaborative work of many people. Here are my special acknowledgments to some of them.

Fellow teachers and dear friends Alexandre Moura, Eliza Brito and Rosa Malena contributed a lot to each chapter, helping me to plan the experiments, recruit participants and analyze data. Denise Lemos, Ruth Dias and Maria Turci gave great contributions motivating students to volunteer to the experiments. As scientific initiation students, Uslaine Lima and Camila Murta supported some of the experiments' planning and execution. Fellow PhD students at iMERR (Institute of Medical Education Research Rotterdam) helped me find my ways at Erasmus MC and the city of Rotterdam. My special thanks to Josepha Kuhn, who shared her process of transforming texts into idea units (which I did a lot!) and Justine Staal, who patiently and kindly helped me with many different things, from Qualtrics' software support to Dutch translations. I am also grateful to each and every medical student who dedicated their time and effort to participate in my experiments.

The thesis would not be possible without the support of UNIFENAS-BH Medical School. At UNIFENAS-BH I learned that being a medical teacher goes way beyond being a medical doctor. Teaching at UNIFENAS inspired me to pursue an academic career in the medical education field, culminating in this PhD journey. The school has always supported my research projects and provided the necessary adjustments in my routine as a teacher that allowed me to visit iMERR yearly (until Covid-19 emerged).

At iMERR I always felt welcome. The exchanges with iMEER's faculty, secretaries and PhD students made my academic PhD experience richer, as well as my social experience. My special thanks to Professor Walter van den Broek, for his kindness and institutional support, and Rita de Kimpe, who helped me navigate documents and bureaucracy in my first days at iMERR.

My very special thanks to my promotors. Silvia guided me closely throughout the whole process. Always scientifically accurate, she was never personally rigid and was sensitive to my professional and personal troubles and needs. It was also comforting to rely on Henk's guidance along the way, especially in the most challenging steps of the research. I felt reassured, and quite privileged, to count on his experience and kindness whenever a difficult question or decision emerged.

Last, but not least, my gratitude to my parents, friends, colleagues and research partners for all the happy moments we shared in the past years. These were times of joy!

Appendices

APPENDIX 1: DELIBERATE REFLECTION'S PROCEDURE AND INSTRUCTIONS.

Please read carefully the instructions below.

You will be presented to (x) clinical cases on the following pages. We would like you to work on these cases, one by one, through a series of steps, following this sequence:

1. First, read the case and write down the most likely diagnostic hypothesis to it. We are interested here in your first diagnostic impression. The first diagnosis that comes to a doctor's mind is often the correct one. Please try to read the case and write down the diagnosis as fast as you can but without making mistakes. You have 2 minutes for this first diagnosis. I will tell you when you have to start and when you should finish and move to the next page.
2. On the next page, the same case will be presented again. Now we would like you to reflect on the case by following a structured procedure. By strictly following this procedure, you can make an accurate differential diagnosis, which will provide you with the basis for your diagnostic decision. After the case, there will be a blank table. Fill it with the information requested on it:

A) In the first row of the column "Diagnostic hypotheses" write the diagnosis that you gave for the case on the previous page.

B) Write down the findings in the case that corroborate your hypothesis and those that refute it in the respective columns. List also the findings that you would expect to be present if your hypothesis were true, but are absent in the case.

C) Now, suppose your first hypothesis proved to be incorrect. What other hypotheses would you consider? Write these alternative hypotheses down, one in each line of the table. Now proceed in the same way for analyzing each hypothesis; Indicate the findings that reaffirm and those that refute the hypothesis, as well as the findings that you expected to encounter if the hypothesis were true but are not present in the case.

D) Based on this analysis, now indicate in the first column the order of likelihood of the diagnostic hypotheses you considered. Write 1 for the most likely hypothesis, 2 for the second one, and so on.

The progress time will be controlled, so please follow the instructions on it.

Next, you will find an example of a case with the responses provided by a student in another study. Read it carefully so that you can understand what you are requested to do.

Case-example

Read the case below and write down your first diagnosis hypothesis.

Try to read it quickly and, as soon as you do it, write your hypothesis.

A 48 years old man goes to an emergency room complaining about dyspnea and chest discomfort in the last 24 hours. He reports two similar episodes last month: one of them when he was gardening and the other, painting. In both occasions, the symptoms disappeared spontaneously in a few hours, with no interventions. He denies orthopnea. In between the episodes, he felt fine, with no restrictions to exercise. He smoked for 20 years, one pack a day, but quit five years ago. He is not on medications and has no record of allergies. He reports "bronchitis" in his childhood. His father died with a heart attack at the age of 61.

Physical examination:

PA: 120/80mmHg; Pulse: 90 bpm; Temp:36,50°C; RP: 20bpm

Patient with a healthy look, in respiratory discomfort. Flat jugular veins. Normal heart sounds. Lungs: sibilant both sides.

Laboratory tests results	*References*	*Laboratory tests results*	*References*
Hb: 13,5g/dL	13,5 a 17,5g/dL	Mon 5%	3-10%
Leukocytes: 13.500/µL	4.000-11.000/ µL	Eosinophils 2%	1-5%
Segmented 82%	45-75%	Throponin	negative
Lymphocits 11%	22-40%		

Eletrolytes, urea, creatinin, glucose: normal.

Blood gases (ambient air) - pO2: 52 mm Hg, pCO2: 35 mm Hg, pH: 7.44.

ECG: left ventricular hypertrophy; unspecific repolarization changes.

Blood gases after recovery (ambient air) - pO2: 104 mm Hg, pCO2: 33 mm Hg, pH: 7.41.

Chest X-Ray: cardiac area slightly enlarged; normal lungs.

What's your first diagnostic hypothesis for this case?

Asthma

The previous case will be presented to you again, as follows.

A 48 years old man goes to an emergency room complaining about dyspnea and chest discomfort in the last 24 hours. He reports two similar episodes last month: one of them when he was gardening and the other, painting. In both occasions, the symptoms disappeared spontaneously in a few hours, with no interventions. He denies orthopnea. In between the episodes, he felt fine, with no restrictions to exercise. He smoked for 20 years, one pack a day, but quit five years ago. He is not on medications and has no record of allergies. He reports "bronchitis" in his childhood. His father died with a heart attack at the age of 61.

Physical examination:

PA: 120/80mmHg; Pulse: 90 bpm; Temp:36,50°C; RP: 20bpm

Patient with a healthy look, in respiratory discomfort. Flat jugular veins. Normal heart sounds. Lungs: sibilant both sides.

Laboratory tests results	*References*	*Laboratory tests results*	*References*
Hb: 13,5g/dL	13,5 a 17,5g/dL	Mon 5%	3-10%
Leukocytes: 13.500/µL	4.000-11.000/ µL	Eosinophils 2%	1-5%
Segmented 82%	45-75%	Throponin	negative
Lymphocits 11%	22-40%		

Eletrolytes, urea, creatinin, glucose: normal.

Blood gases (ambient air) - pO2: 52 mm Hg, pCO2: 35 mm Hg, pH: 7.44.

ECG: left ventricular hypertrophy; unspecific repolarization changes.

Blood gases after recovery (ambient air) - pO2: 104 mm Hg, pCO2: 33 mm Hg, pH: 7.41.

Chest X-Ray: cardiac area slightly enlarged; normal lungs.

Now, you will be asked to work fill out a table following this procedure:

A) Step 1: In the first row of the column "diagnostic hypothesis", write the diagnosis that you gave for the case on the previous page.

B) Step 2: In the respective columns, write the findings in the case that corroborates your hypothesis, those that refute it and those you expected to be present if your hypothesis were true, but are absent in the case.

C) Step 3: suppose your first hypothesis proved to be incorrect. Think of other hypothesis you should consider, write them on the next rows and repeat step 2 for each one of them.

D) Step 4: based on this analysis, indicate in the first column the order of the likelihood of the diagnostic hypothesis you considered. Write 1 for the most likely hypothesis, 2 for the second one, and so on.

Diagnostic hypoyhesis	**Findings that corroborate this hypothesis**	**Findings that refute this hypothesis**	**Findings that you expected in this diagnosis, but are absent in the case.**	**Likelihood of diagnosis**
Asthma	*Chest discomfort, dyspnea, sibling, history of bronchitis.*	*Age at symptoms, no history of allergy.*	*Good response to short action bronchodilator.*	1
Angina pectoris	*Chest discomfort, dyspnea, family history of heart attack, smoking.*	*Non typical chest pain, long term pain, sibling.*	*Ischemic findings on ECG; positive throponin.*	3
Congestive heart failure	*Dyspnea, enlarged heart on X-ray, left ventricular hypertrophy on ECG.*	*No orthopnea.*	*Congested jugular veins; B3, pulmonary crepitating.*	2

APPENDIX 2: SITUATIONAL INTEREST AND AWARENESS OF KNOWLEDGE GAPS QUESTIONNAIRES AS PRESENTED TO PARTICIPANTS

Instructions to participants

Reflection groups: You just performed a task that consisted of giving an initial diagnosis to a clinical case of a patient with jaundice, reflecting in a structured way on the plausible alternatives for the differential diagnosis and choosing the most likely one. Please analyze you experience in solving this task. Be careful and consider not only your final diagnosis but understand "this task" as the process of reflection on diagnostic alternatives that underlie your decision. Read each statement below carefully and indicate, on a scale from 1 (*not true at all for me*) to 5 (*very true for me*), how true each statement is for you **right now**.

Differential diagnosis groups: You just performed a task that consisted of giving an initial diagnosis to a clinical case of a patient with jaundice, thinking about alternative diagnoses for the case and choosing the most likely one. Please analyze you experience in solving this task considering not only your final diagnosis but understand "this task" as the process of analyzing the diagnostic alternatives that underlie your decision. Read each statement below carefully and indicate, on a scale from 1 (*not true at all for me*) to 5 (*very true for me*), how true each statement is for you **right now**.

Questions

Questions on Situational Interest (SI) and Awareness of Knowledge Gaps (AKG) were merged in a single questionnaire. Questions 1, 4, 7, 10, 14 and 15 measured SI. Questions 2, 3, 5, 6, 8, 9, 11, 12 and 13 measured AKG.

1.	*I enjoyed working on this task.*	1 *Not true at all*	2 *Not true for me*	3 *Neutral*	4 *True for me*	5 *Very true for me*
2.	Working on this task revealed I don't know certain things yet.	1 *Not true at all*	2 *Not true for me*	3 *Neutral*	4 *True for me*	5 *Very true for me*
3.	My knowledge was insufficient to complete this task.	1 *Not true at all*	2 *Not true for me*	3 *Neutral*	4 *True for me*	5 *Very true for me*
4.	*I think this task was interesting.*	1 *Not true at all*	2 *Not true for me*	3 *Neutral*	4 *True for me*	5 *Very true for me*

5.	This task was too hard to be solved completely.	1 *Not true at all*	2 *Not true for me*	3 *Neutral*	4 *True for me*	5 *Very true for me*
6.	It seems to me that my solution for this task is incomplete.	1 *Not true at all*	2 *Not true for me*	3 *Neutral*	4 *True for me*	5 *Very true for me*
7.	*I was totally focused while working on this task; I was not distracted by other things.*	1 *Not true at all*	2 *Not true for me*	3 *Neutral*	4 *True for me*	5 *Very true for me*
8.	Solving this task was easy for me.	1 *Not true at all*	2 *Not true for me*	3 *Neutral*	4 *True for me*	5 *Very true for me*
9.	While solving this task, I sometimes briefly felt that I encountered difficulties	1 *Not true at all*	2 *Not true for me*	3 *Neutral*	4 *True for me*	5 *Very true for me*
10.	*Presently, I feel bored.*	1 *Not true at all*	2 *Not true for me*	3 *Neutral*	4 *True for me*	5 *Very true for me*
11.	At certain times, I was stuck with solving this task.	1 *Not true at all*	2 *Not true for me*	3 *Neutral*	4 *True for me*	5 *Very true for me*
12.	I am not sure if I have found the right solution for this task.	1 *Not true at all*	2 *Not true for me*	3 *Neutral*	4 *True for me*	5 *Very true for me*
13.	While working on this task, at certain moments I felt unsure if my answers were correct.	1 *Not true at all*	2 *Not true for me*	3 *Neutral*	4 *True for me*	5 *Very true for me*
14.	*I want to know more about the differential diagnosis of patients with jaundice.*	1 *Not true at all*	2 *Not true for me*	3 *Neutral*	4 *True for me*	5 *Very true for me*
15.	*I want to master the differential diagnosis of patients with jaundice well.*	1 *Not true at all*	2 *Not true for me*	3 *Neutral*	4 *True for me*	5 *Very true for me*

APPENDIX 3: CUED RECALL-TEST QUESTIONS.

List all the relevant clinical history information to the evaluation of patients with jaundice and explain how they help on the differential diagnosis.

1. Explain how the characteristics of urine and feces can help differentiating the causes of jaundice.
2. List all the relevant physical examination findings to the evaluation of patients with jaundice and explain how they help on the differential diagnosis.
3. Explain how direct and indirect bilirubin levels help differentiating the causes of jaundice.
4. Explain how the levels of enzymes AST and ALT help differentiating the causes of jaundice.
5. Explain how the levels of alkaline phosphatase and gamma-glutamyl transferase help differentiating the causes of jaundice.
6. Explain how prothrombin activity helps differentiating the causes of jaundice.
7. List all the causes of conjugated hyperbilirubinemia that you remember.

PPENDIX 4: EXAMPLE OF PARTICIPANT'S ANSWER SCORING TO THE CUED RECALL TEST.

Q1 ANSWER	Q1 SCORING	Q2 ANSWER	Q2 SCORING	Q3 ANSWER	Q3 SCORING	Q4 ANSWER	Q4 SCORING
Adequate assessment of symptoms (for example, a detailed description of pain with regard to its onset, character, relief and worsening factors, etc.) may favor the narrowing of the diagnosis. In addition, data from previous history, such as previous STI* may indicate the etiology of a possible hepatitis. The evolution of the jaundice can also favor the accurate diagnosis (The appearance of jaundice about 12 hours after pain in the epigastrium that radiates to the right hypochondrium may indicate choledocholithiasis for example). *sexually transmited infection	1-pain; 2-onset of pain; 3-previous history; 3-previous STI; 4- hepatitis; 5- STI indicates hepatitis; 6- choledocholithiasis; 7- jaundice that appears 12 hours after pain suggests choledocholithiasis; 8-epigastric pain; 9-pain in right hypochondrium; 10-pain radiating from the epigastrium to the right hypochondrium suggests choledocholithiasis	urobilinogen is responsible for causing choluria in times when there is hyperbilirubinemia and this leads us to investigate the causes of this accumulation of bilirubin in the blood (excess of red blood cell degradation? hepatic malabsorption?). Acholic stool happens by reducing the amount of stercobilin in the digestive tract and this occurs due to the absence of bilirubin in the intestine.	1-color of urine; 2-choluria; 3-color of the stool; 4- Acholic stool; 5- Acholic stool is due to the reduction of stercobilin in the digestive tract; 6- Acholic stool is due to the absence of bilirubin in the intestine.	Murphy's sign may indicate cholecystitis; Examination of the liver may facilitate the diagnosis with regard to size, consistency, edges and the presence of pain (which may help us to differentiate between acute and chronic liver disease, for example). jaundice intensity is important to be assessed.	1-Murphy's sign, 2-Musphy's sign may indicate cholecystitis, 3- Examination of the liver, 4- liver size, 5-liver consistency, 6-liver edges, 7-liver pain, 8-liver pain may help to differentiate between acute and chronic liver disease.	High levels of unconjugated bilirubin may indicate pre-hepatic causes of jaundice. High levels of conjugated bilirubin may indicate obstructive causes. High levels of both may indicate hepatocyte damage.	1-High levels of unconjugated bilirubin; 2- pre-hepatic causes of jaundice. 3-high levels of unconjugated bilirubin may indicate pre-hepatic causes of jaundice. 4-High levels of conjugated bilirubin, 5- obstructive causes, 6-high levels of conjugated bilirubin may indicate obstructive causes., 7- hepatocyte damage, 8-High levels of both conjugated and unconjugated bilirubin may indicate hepatocyte damage.

Q5 ANSWER	Q5 SCORING	Q6 ANSWER	Q6 SCORING	Q7 ANSWER	Q7 SCORING	Q8 ANSWER	Q8 SCORING
alkaline phosphatase (AP) is excreted with bile and is therefore increased in cases of biliary stasis (however, it may be increased in osteopathy or in pregnancy as well) .Gamma-glutamyl transferase (GGT) is also produced in the bile ducts and it is an indicator of its injury (it also serves as a marker of alcohol consumption).	1-alkaline phosphatase (AP) is produced in bile ducts, 2-AP is excreted with bile, 3-elevated levels of AP 4- AP is increased in cases of biliary stasis, 5-GGT is produced in the bile ducts, 6-GGT is an indicator of bile ducts injury, 7-AP may be increased in osteopathy, 8-AP may be increased in pregnancy.	Prothrombin is a protein synthesized by the liver, therefore, reduced prothrombin activity may indicate liver damage, as this substance is produced in the liver.	1-Prothrombin is a protein, 2-Prothrombin is synthesized by the liver, 3-reduced prothrombin activity, 4- reduced prothrombin activity may indicate liver damage	ALT is an enzyme present in the hepatocyte cytoplasm and may indicate liver damage when it is above its reference values. AST is present in both the cytoplasm and hepatocyte mitochondria (but is less specific because it is also present in skeletal muscles and myocardium). Elevation of these enzymes indicates liver damage (sudden and significant increase may indicate viral hepatitis)	1-ALT is an enzyme 2- ALT is an enzyme present in the hepatocyte cytoplasm, 3-ALT may indicate liver damage, 4- ALT may indicate liver damage when it is above its reference values, 5- AST is present in the cytoplasm, 6-AST is present in hepatocyte mitochondria, 7-AST is present in skeletal muscles, 8- AST is present in the myocardium, 9-AST is less specific than ALT, 10- AST is less specific than ALT because it is also present in skeletal muscles and myocardium, 11- increased levels of AST, 12- AST increased levels indicate liver damage, 13- ALT sudden and significant increase may indicate viral hepatitis, 14- AST sudden and significant increase may indicate viral hepatitis.	viral hepatitis, pancreatic head tumor, Vater's papillary tumor	1-Hepatites, 2-viral hepatitis, 3-pancreatic tumor, 4-pancreatic head tumor, 5-Vater's papillary tumor

APPENDIX 5: THE ORIGINAL QUESTIONNAIRE ON AWARENESS OF KNOWLEDGE GAPS, IN GERMAN.

Bitte lies jede Aussage sorgfältig durch und entscheide, in welchem Ausmaß die Aussage auf Dich zutrifft oder nicht zutrifft. Es gibt sechs verschiedene Antwortmöglichkeiten, von „trifft überhaupt nicht zu" (1) bis „trifft völlig zu" (6).

Bitte gib an, inwieweit die folgenden Aussagen auf Dich zutreffen…	
1. Mir ist durch die Aufgabe klar geworden, dass ich bestimmte Dinge noch nicht weiß.	1 2 3 4 5 6
2. Um diese Aufgabe zu lösen, fehlt mir notwendiges Wissen.	1 2 3 4 5 6
3. Diese Aufgabe war zu schwer, um sie vollständig lösen zu können.	1 2 3 4 5 6
4. Meine Lösung der Aufgabe kommt mir unvollständig vor.	1 2 3 4 5 6
5. Mir fiel es leicht, die Aufgabe zu lösen.	1 2 3 4 5 6
6. Ich hatte beim Lösen der Aufgabe manchmal kurz das Gefühl, auf Schwierigkeiten zu stoßen.	1 2 3 4 5 6

7. Ich bin beim Lösen der Aufgabe mal nicht weitergekommen.	1 2 3 4 5 6	
8. Ich bin mir nicht sicher, ob ich eine richtige Lösung für die Aufgabe gefunden habe.	1 2 3 4 5 6	
9. Beim Bearbeiten der Aufgabe war ich mir manchmal unsicher, ob meine Antworten richtig sind.	1 2 3 4 5 6	