Dissertation zum Erwerb des Doctor of Philosophy (Ph.D.)
am Munich Center of the Learning Sciences

München, November 9, 2016

Watching people fail

Fostering diagnostic competences with peer feedback on erroneous cognitive modeling examples

Christian Günther Strobel

First Supervisor: Professor Doctor Martin Fischer

Second Supervisor: Professor Doctor Jan-Willem Strijbos

Submitted by: Christian Günther Strobel

Date of defense: 02.02.2017

Experience is simply the name we give our mistakes.

Oscar Wilde

Acknowledgements and Dedication

I would like to begin with expressing my deepest gratitude to my supervisors Prof. Martin Fischer and Prof. Jan-Willem Strijbos. They have been loyal, patient and supportive academic mentors throughout my work, making this dissertation possible. I would also like to emphasize my thankfulness for helping me with amazing support and input to Prof. Ingo Kollar and Dr. Nicole Heitzmann, as well as to Prof. Kevin Eva as my peer reviewer and to Prof. Birgit Neuhaus as my forth examination committee member. Thanks to the Elite-network of Bavaria for generously funding my position, conferences and stays abroad [Project Nr: K-GS-2012-209].

Further, I would like to thank Prof. Frank Fischer who believed in me as researcher and whose support enabled me to pursue my academic career until this PhD. Thanks to all the professors and colleagues from Lehrstuhl für Empirische Pädagogik und Pädagogische Psychologie and the Institut für Didaktik und Ausbildungsforschung in der Medizin at the LMU. Particular regards to Anja Härtl, Thomas Brendel, Diana Ouellette, Mia Wermelt and Benedikt Lenzer. I am grateful for my colleagues from the REASON doctoral program for being such a formidable, supportive and joyful company.

Without the support of my family throughout my career, I would not have been able to pursue this endeavor. Therefore I would like to thank my granny Josefine for believing in me, my parents Klaus-Günther and Claudia for their proud words, my sister Iris for her emergency hotline as well as my aunt Birgit for her open ear and shoulder to lean on.

Many thanks for their contributions to my student helpers Max Fink, Rebecca Thaller, Isabell Wenning, Ceyda Göcmen, Ina Bechter and Veronika Maricic.

Lastly, thanks to my great friends Selcan Özbay, Christina Färber, Klaus Schmidt, Alejandro Mankowski, Leo Horsch, Dominik Hofmeister, Sarah Severin, Kati and Nils Ledig, Vanessa Middendorf, Malte and Felix Lachmann and especially Pezhman Nasirifard for the amazing encouragement through all phases of my PhD. You rock.

Abstract

Utilizing effective instructional means to teach medical students how to diagnose accurately and successfully is a major goal of medical education. The two empirical experimental studies of this thesis examine how learning with cognitive modeling examples might foster diagnostic competences consisting of diagnostic knowledge types and meta-cognitive error detection skills. The first study examined if providing peer feedback and erroneous or correct cognitive modeling examples are effective for learning diagnostic competences. Results showed that errors in examples positively influenced error detection skills whereas the task to provide peer feedback on the examples was detrimental for both diagnostic knowledge and error detection skills. Based on these results, the second study examined if a variation of communication medium (spoken versus written) or additional incorporated expert feedback on the erroneous cognitive modeling examples positively influences degree of elaboration and learning. While expert feedback showed no effect on acquiring diagnostic competences, providing spoken feedback positively influenced the degree of elaboration, which positively influenced the acquisition of diagnostic competences. The results of both studies suggest that erroneous cognitive modeling examples offer advantages for learning diagnostic knowledge and error-related skills for instructional design in medical education. The integration of the task to provide peer feedback on observed diagnoses in teaching should be handled with care as feedback medium and degree of elaboration are influential for learning outcome.

Contents

1 Introduction

1.1 Aim of this thesis

Clinical reasoning as part of the every-day job of physicians remains one of the most complex cognitive tasks (G. J. Kuhn, 2002). Due to its complexity, it is prone to errors. Flawed clinical reasoning comes with costly consequences, since it results in flawed diagnoses: 10-15% of erroneous medical diagnoses are a result of incorrect or incomplete reasoning (Elstein, 2009). This is why error prevention in clinical decision settings on the ground of scientific reasoning has gained significant attention in recent years. Researchers and educators have also called for interventions to decrease the number of errors that result from flawed reasoning processes (Brennan, 2000).

However, designing such interventions is not an easy task, as intervention studies aimed at fostering diagnostic or meta-cognitive skills show inconclusive results. For instance, research on training of concrete diagnostic skills like overall diagnostic performance or knowledge, (Eva, Link, Lutfey, & McKinlay, 2010; Mamede et al., 2012; Ziv, Ben-David, & Ziv, 2005) and meta-cognitive skills such as cognitive forcing strategies and reflection (Sherbino, Kulasegaram, Howey, & Norman, 2014) yield inconclusive empirical evidence. These studies reflect the difficulty of developing effective trainings to teach clinical reasoning and to reduce errors occurring within the reasoning process. Sherbino et al. (2014) for example conducted a large-scale 4-weeks training with 145 medical students to teach them cognitive forcing strategies. Cognitive forcing strategies are meta-cognitive self-monitoring strategies that can be applied during diagnostic reasoning and are hypothesized to counter errors in diagnoses. Results, however, showed no significant impact of the training.

Apart from the empirical research, there are suggestions based on theoretical work about what kinds of interventions might work (Croskerry, Singhal, & Mamede, 2013a, 2013b;

Sherbino, Dore, Siu, & Norman, 2011). These suggestions are yet to be empirically tested. One of those promising approaches is observational learning with cognitive modeling examples (Van Gog, 2015). The observed models, in this case physicians, can either display a correct performance or one that commits one or more errors. In theory, watching models whose performance includes errors might teach meta-cognitive *negative knowledge* about how and where errors can happen (Oser & Spychiger, 2005) – referred to as *error detection skills* (Nyssen & Blavier, 2006) – and prove effective to avoid errors in one's own clinical reasoning. Hence, observing other models' reasoning mistakes that lead to false diagnoses should result in a better understanding of reasoning processes and the errors that can occur along the way. While erroneous cognitive modeling examples have been shown effective in other settings (Van Gog, 2015) and seem to be a promising approach to enhance clinical decisions, there is a research gap with respect to testing the approach's viability in medical settings. How to actively engage students into elaborating on the observed cognitive modeling example is open for discussion (Renkl & Atkinson, 2003). One approach might be providing peer feedback on the observed performance (Li, Liu, & Steckelberg, 2010).

To summarize, this thesis investigates if observing erroneous cognitive modeling examples is a valid approach to teach diagnostic competences. It is further explored if and how the task to provide peer feedback on these examples influences the learning gain of own diagnostic competences. How this task should be designed in order to be effective will also be examined. The two main research questions of this thesis are:

(1) Can medical students acquire diagnostic competences from providing peer feedback on erroneous or correct cognitive modeling examples of clinical reasoning?

(2) Does expert feedback or the medium by which peer feedback is provided on erroneous cognitive modeling examples influence the degree of elaboration and the acquisition of diagnostic competences?

To give a brief overview, the following section provides an outline of the chapters of this thesis.

1.2 Outline of the thesis

The first chapters of this dissertation, Chapters 2 to 4, are dedicated to clarify and describe the theoretical and empirical background of important concepts that are the basis of the empirical studies. An overview of these studies is presented in Chapter 5 and in detail in chapter 6 and 7. Chapter 8 summarizes and compares both empirical studies, with a general discussion provided in Chapter 9.

Chapter 2 focuses on scientific reasoning and highlights arguments why it is essential to emphasize the relevance of research about this skill for domains such as medicine and why a clinical reasoning model shows great similarity to a cross-domain model of scientific reasoning. For clinical reasoning, early research highlighted cross-domain scientific reasoning skills underlying successful diagnosis (Patel, Arocha, & Zhang, 2005). This view however, was challenged by multiple theories, scholars and studies in recent years (Ilgen et al., 2012). First, similarities of a general model of scientific reasoning and models of clinical reasoning are outlined. A recent cross-domain model of scientific reasoning by F. Fischer et al. (2014) is depicted to explain basic steps in the process of conscious scientific reasoning. Differences in the proposed modes of operation of reasoning by Stokes (1997) and studies emphasizing the importance of domain specific factors, knowledge and skills highlight the differences between clinical and scientific reasoning and challenge a view of domain-generality of reasoning (Krasne & Stevens, 2010; Kruglanski & Gigerenzer, 2011).

Chapter 3 describes the concept of diagnostic competences as necessary factors that constitute successful clinical reasoning as it relies heavily on domain specific content knowledge (Schmidt, Norman, & Boshuizen, 1990). As medical expertise grows, this knowledge changes its form of representation from basic semantic structures over more elaborated knowledge networks to encapsulated illness scripts that are based on a large experience base (Boshuizen & Schmidt, 2008). This knowledge is divided into three main parts: conceptual knowledge about what to do, strategic knowledge about how to do it and conditional knowledge about why to do it (Van Gog, Paas, & Van Merriënboer, 2004). For diagnostic accuracy however, domain specific content knowledge seem to not be the only influential factor (Clarke et al., 2000; Graber, 2009). Most diagnostic errors are due to flaws in the clinical reasoning (Schiff et al., 2009). These flaws might be explained by the increased automatization of clinical reasoning with more expertise: experts use more automated pattern recognition and only use conscious clinical reasoning when the automated reasoning is likely to fail (Hammond, 1990; Kahneman, 2011; Moulton, Regehr, Mylopoulos, & MacRae, 2007). Errors do not happen less in expert clinical reasoning, but experts outperform novices in error detection and management (Wilkinson, Cauble, & Patel, 2011). Therefore, I propose to add a fourth dimension of error detection skills to a holistic framework of diagnostic competences. How to teach these competences however, is still debated.

Chapter 4 focuses on how to foster diagnostic competences. Due to medical expertise (and illness scripts) and usable knowledge being based on experience with patient cases, case-based learning methods are advocated to be fruitful for learning diagnostic competences (Schmidt & Rikers, 2007; Stark, Kopp, & Fischer, 2011). These cases can be presented as written worked examples (Heitzmann, Fischer, Kühne-Eversmann, & Fischer, 2015) or video-based as cognitive modeling examples for more authenticity (Renkl, 2014; Van Gog & Rummel, 2010). Correct examples are good for beginners, but for intermediate and more advanced learners incorporating errors in the observed performance has been shown to increase motivation and attention and ultimately learning gain (Große & Renkl, 2007; Renkl & Atkinson, 2003). These errors in cognitive modeling examples are hypothesized to be beneficial for learning error detection skills as they teach the student what possible error pitfalls might be (Blandin & Proteau, 2000; Gartmeier, Bauer, Gruber, & Heid, 2008).

However, these erroneous examples though need to be accompanied by instruction on how to reflect and elaborate on the content. Self-reflection prompts have been found to be beneficial to actively engage the student in the elaboration (Atkinson, Renkl, & Merrill, 2003; Craig, Chi, & VanLehn, 2009). For a more constructive or even interactive approach, the task to provide peer feedback on the observed might lead to better learning (Chi, 2009; Y. H. Cho & Cho, 2011; Nicol, Thomson, & Breslin, 2014). Therefore, two studies were conducted which are summarized in a brief overview in Chapter 5.

The first empirical study in Chapter 6 examined the effect of erroneous (versus correct) examples on the acquisition of diagnostic competences and error detection skills and if providing peer feedback on these examples as an interactive instructional method is beneficial for learning diagnostic competences. This was examined with a 2x2 quasi-experimental study with a control group with 121 medical students in the seventh semester. The students either provided peer feedback on or just observed three erroneous or correct cognitive modeling examples of peer students' differential diagnoses. The results indicated that students learning with erroneous cognitive modeling examples acquire higher error detection skills than students learning with correct ones. No difference was found for knowledge parts of diagnostic competences. Against the hypotheses, providing peer feedback was detrimental for the acquisition of both knowledge and meta-cognitive error detection parts of diagnostic competences. This was explained by the unfamiliarity of the task to provide peer feedback and therefore too high mental effort was caused.

Based on the results from Study 1, Study 2 in Chapter 7 examined how to alter the intervention to make providing peer feedback beneficial for learning as described in the literature. For that, the effects of different communication media through which learners provide spoken or written peer feedback and receiving expert feedback (or not) on the observed performance on degree of elaboration and acquisition of diagnostic competences were investigated. The design was a 2x2 quasi-experimental study with a control group as well. Participants (N = 134) either provided spoken or written peer feedback on three erroneous cognitive modeling examples and received expert feedback or did not. Results showed that receiving expert feedback did not make a difference in terms of degree of elaboration of provided peer feedback or acquisition of diagnostic competences. Medium of communication did influence learning indirectly: Students providing spoken peer feedback provided more elaborated peer feedback, which led to a better acquisition of diagnostic competences.

In Chapter 8, both studies are shortly summarized (sections 8.1 through 8.2) and then compared (section 8.3). Results of this comparison indicate that samples were similar in terms of demographics, but significantly different in terms of prior knowledge. However, this difference was not responsible for differences in degree of elaboration of provided peer feedback. An empirical comparison showed that students in Study 2 provided more elaborated peer feedback, which might be explained by the short peer feedback training incorporated in Study 2. However, the samples of Study 1 and Study 2 did not differ in terms of acquisition of diagnostic competences overall.

Chapter 9 offers a general discussion with an interpretation, discussion and implications of the findings in section 9.1, conclusions in section 9.2, which offer answers to the two main research questions of this thesis, and limitations and an outlook on possible future research in section 9.3 in 9.4. The closing thoughts are outlined. In section 9.1 depicted results are shortly interpreted and discussed. First, the effects of cognitive modeling examples on the acquisition of diagnostic competences are discussed. However, it is questioned if the benefits of using cognitive modeling examples in medical education outweigh the costs. Using OSCE settings to create cognitive modeling examples for learning could counter high costs. Erroneous cognitive modeling examples were found to be beneficial for acquiring conceptual parts of diagnostic competences when compared to textbook learning. The impact of the findings of positive effects of erroneous cognitive modeling examples on meta-cognitive error detection skills are integrated into the existing literature. Practical implications for medical education are drawn: Learning with errors should be integrated into teaching methods to teach students error-related skills. The effects of providing elaborated peer feedback on (erroneous) cognitive modeling examples are then discussed: While providing corrective peer feedback was found to be detrimental for acquiring diagnostic competences in Study 1, Study 2 showed that the degree of elaboration of provided peer feedback positively influences learning. However, providing spoken peer feedback led to a higher degree of elaboration, which in turn led to more acquisition of diagnostic competences. Receiving expert feedback did neither influence elaboration of provided peer feedback nor acquisition of diagnostic competences. For educators in practice, the results imply that using peer feedback as intervention for learning should be done with care, as training and medium play an important role for the degree of elaboration, which is essential for learning gain.

In section 9.2 answers to the two main research questions of this thesis are concluded. Erroneous cognitive modeling examples can be used in medical education to learn distinct parts of diagnostic competences. Providing peer feedback was found to be detrimental, but it makes a difference if it is elaborated or corrective. Research question two about the influence of expert feedback and medium of peer feedback provision could be answered with yes and no: Receiving expert feedback while providing peer feedback does not increase degree of elaboration or learning gain. In contrast, the medium of communication of provided peer feedback did have an influence. Providing spoken compared to written peer feedback increases degree of elaboration of peer feedback and therefore learning gain.

The limitations presented in section 9.3 firstly address the instrument limitations: The pen-and-paper assessment of diagnostic competences, suggesting a multi-trait multi-method assessment. Performance-based assessment like OSCE or Script-concordance testing could be included in testing for triangulation to counter critique concerning validity. Further, error detection skills need to be validated with locating it empirically (e.g. by factor or function analysis) in a framework of diagnostic competences and should be related to accuracy and performance in the clinical setting. Furthermore, the covariate of mental effort must be measured differently to distinguish different types of cognitive load that might be responsible for parts of the negative findings of peer feedback provision as discussed earlier. A limitation concerning learning material stems from the video-based nature of the intervention. The results can only be carefully transferred into real-life settings.

Overall, future studies might explore the relationship of error detection skills and other knowledge types or competences in medicine, but also across domains. Further research is needed to entangle if and how a combination of correct erroneous cognitive modeling examples is better for teaching diagnostic competences than erroneous examples alone. Further, the effects of incorporating practice phases to rehearse the observed performances might be beneficial. To unveil what happens in cognition of students during providing peer feedback, retrospective analysis methods could be applied. The results concerning feedback leave the question what characteristics of providing peer feedback influence learning positively or negatively. Future studies might also focus on the effects of peer feedback trainings on elaboration to find out what influences the degree of elaboration and therefore learning gain. In closing, the suggestion is provided that peer feedback provision for learning must be adequately designed to influence learning positively. Errors in medical education should gain more attention as this thesis highlights possible advantages for diagnostic accuracy.

2 The Relevance of Scientific Reasoning for Medical Practitioners

The following three sections are dedicated to explaining the importance of scientific reasoning in clinical contexts and showing the close link between scientific and clinical reasoning. While clinical reasoning shares distinct aspects with scientific reasoning, scholars highlight the domain-specific aspects of clinical reasoning. Hence, the coming sections not only draw attention to the links between these concepts, but also point out to the differences and the domain specific aspects of clinical reasoning. It begins with why it is important to study scientific reasoning for society and medicine in particular (section 2.1). A recent across-domain model of scientific reasoning is presented (section 2.2) and domain specific aspects are highlighted (section 2.3) that impede the complete adaptation of a cross-domain scientific reasoning model for clinical reasoning.

2.1 Importance of Scientific Reasoning in Society and Medicine

In 1979, Daniel Kahnemann and Amos Tversky published an article in Econometrica about their *Prospect Theory* (Kahneman, 2011; Kahneman & Tversky, 1979) investigating the cognitive processes underlying decision making processes in economics. With their Nobel-prize winning extension of the existing idea of people as irrational decision makers with cognitive biases, they undertook a first step in experimental psychology to see decision making as a multifactorial, complex endeavor, sensitive to manifold factors. A clear-cut, easy model of how reasoning ‘in general’ works was neglected. Decision-making and reasoning since then has been target of a vast amount of theoretical essays, philosophical discourses and empirical studies, shedding light on the processes involved. Especially scientific reasoning as the basis for the logic in science has been emphasized as important research topic (Popper, 1959).

The value of scientific reasoning as a skill derives from its imminent implications for individuals and society. For individuals, it serves as an important 21st century skill to cope with the modern world as an information society and navigate in it fluently (F. Fischer et al., 2014; Trilling & Fadel, 2009). Since scientific reasoning is a precondition to conduct research and to create new knowledge, it is crucial for the development of our society (Morris, Masnick, Zimmerman, & Croker, 2012). Therefore, competences connected to scientific reasoning and argumentation, as for instance evaluating evidence or setting and testing hypotheses, have been the target of different educational institutions, which have been calling for an explicit facilitation of reasoning in schools and higher education in different domains such as mathematics, engineering and general science education (Hersh, 2005; KMK, 2004; Quinn, Schweingruber, & Keller, 2012).

In the domain of medicine, scientific reasoning is also of major importance: During diagnosing, the physician engages in several scientific reasoning tasks such as identifying a problem, generating possible explanations and testing them (Patel et al., 2005). However, it is an open empirical question whether clinical reasoning is congruent to scientific reasoning in the field of medicine and whether rules and frameworks can be applied by with adapting a scientific reasoning model for this context. To examine if a domain-general scientific reasoning model could be suitable to explain the reasoning during the diagnostic process, a detailed recent domain-general scientific reasoning model by F. Fischer et al. (2014) is explained in the following.

2.2 A Domain-General Model of Scientific Reasoning

Due to the significance of understanding scientific reasoning in different domains such as medicine, scholars thrive for unraveling its complexity. Scientific reasoning and its underlying cognitive processes have been examined with multiple approaches by many scholars over the last decades (F. Fischer et al., 2014). The complexity of scientific reasoning makes it hard to grasp as it constitutes of multiple components such as problem identification, hypothesis generation and testing (Schunn & Anderson, 1999) in multiple domains such as medicine (Patel et al., 2005), biology (Lawson et al., 2000) or physics (Bao et al., 2009). Therefore, many models of a general scientific reasoning process exist describing the process of reasoning in varying complexity and domain-specificity.

These conceptual frameworks aim to explain what makes a scientist successful when identifying and explaining a scientific phenomenon (Klahr & Dunbar, 1988). Scientific reasoning can be defined as reasoning that establishes hypothesis based on “intentional knowledge seeking to test theories and hypotheses and to evaluate evidence with respect to a hypothesis or theory” (F. Fischer et al., 2014, p. 30). D. Kuhn (2002) describes in a basal model the basic steps of scientific reasoning across domains with identification and investigation of the problem, access and analysis of data, inference and argument deduced from the observed. F. Fischer et al. (2014) reformulated and extended these steps by eight epistemic activities in a non-linear model of scientific reasoning. The model is one of the most holistic and recent ones, as it depicts the scientific reasoning process in great detail: It starts with (1) problem identification: Identifying a practical or scientific problem that needs advancement in given theories and methods to solve or explain it. After problem identification, (2) questions are formulated that lead to the (3) generation of hypotheses that already suspect a possible answer to the question based on existing knowledge, methods and explanations. (4) Artefacts are constructed and redesigned which are methodological tools to examine the generated hypotheses scientifically. Through that, (5) evidence is generated by (deductive) scientific inquiry which is then (6) evaluated for quality and validity with existing scientific standards of a field. By weighting and analyzing the significance of the evidence, (7) conclusions are drawn under consideration of previously established hypotheses. In a last step, the results are (8) communicated and scrutinized. Throughout these reasoning steps, domain specific knowledge is needed to perform each step successfully (Schauble, 1996), such as diagnostic knowledge in medicine (Norman & Eva, 2010).

The framework of epistemic steps by F. Fischer et al. (2014) depicts scientific reasoning as a deliberate and conscious process, based on explicit knowledge. However, other scholars also advocate intuitive parts of decision-making within the scientific reasoning process: Kruglanski and Gigerenzer (2011) see scientific reasoning as relying on a two-process model of decision-making based on intuitive and deliberate judgements. A person uses both heuristics for intuitive and domain-specific rules and knowledge for deliberate judgements. According to the person’s processing capacities, the rules and knowledge are matched to the existing problem. Decisions are made based on a fit or no-fit of evidence and knowledge with problems to advance further in the scientific reasoning process. These decisions during epistemic steps however are based on domain-specific knowledge to reach the aim of the reasoning process (Kruglanski & Gigerenzer, 2011; Schauble, 1996). Both knowledge and aim can vary across domains: In medical education research, adopting a general scientific reasoning model for clinical reasoning has been largely challenged due to the importance of domain-specific factors (Ilgen et al., 2012; Patel et al., 2005). In the following section, this debate will be explained, drawing on two major differences between clinical (which is reasoning about issues connected to patient care) and a scientific reasoning models.

2.3 Domain-specific aspects of scientific reasoning in medicine

Whereas F. Fischer et al. (2014) claim that the eight epistemic steps of their scientific reasoning model can be found across different domains and reasoning tasks, aims of the scientific reasoning process can differ. This means the epistemic steps of scientific reasoning can be executed in different scientific modes (depending on the aim) as proposed by Stokes (1997), either more aiming at theoretical or practical implications. Three modes of scientific reasoning are described to classify scientific reasoning according to its relevance for theory advancement (generalized knowledge) or practical immediate use (F. Fischer et al., 2014; Stokes, 1997):

(a) Basic research with the goal to advance existing theory (generalized knowledge) about phenomena (Bohr’s Quadrant),

(b) applied research with the goal to advance existing practice (Edison’s Quadrant) and

(c) a combination of both as use-inspired basic research (Pasteur’s Quadrant).

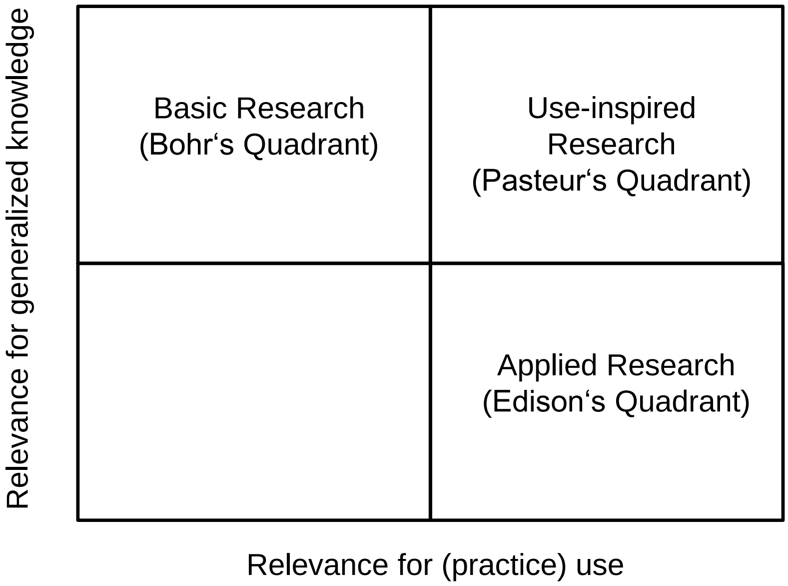

Figure 1. Epistemic modes of scientific reasoning after Stokes (1997)

To exemplify Bohr’s quadrant, basic research in how the digestive system of a specific breed of micromammals during ice age was functioning does not immediately impact humankind or a general practice but broadens the knowledge about this time. Contrary, research in Edison’s quadrant for immediate use might be found in engineering: A study about two types of engine gasket to see which one erodes slower might have an impact on car drivers but does not advance generalized knowledge in that area. Pasteur’s Quadrant combines both approaches: For example, research about how science lessons in Germany could be designed by examining the cognitive processes involved in science learning can be located in Pasteur’s quadrant as it has both relevance for generalized knowledge (theories about cognition and learning) and practice use (better science classes). Davis (1997) locates clinical reasoning in a special paradigm of practical reasoning as it draws on basic and applied research but is used for the goal to heal the patient. The clinician acts as a scientist trying to solve a problem, but the ‘sample’ is the patient. Compared to a general concept of scientific reasoning, clinical reasoning of patient cases differs in terms of its aim but, due to its goal to heal the patient, cannot be clearly located in the proposed paradigms depicted in Figure 1.

Apart from the difference in modus operandi, studies examining scientific reasoning and its underlying processes often seem to neglect domain-inherent factors such as knowledge, culture, situations or problems of different domains (F. Fischer et al., 2014; Morris et al., 2012; Schunn & Anderson, 1999):

> “Although much of the validity in studying scientific reasoning processes lies in the real-world nature of the task, it is an interesting fact that the great majority of psychological research on scientific reasoning has studied neither actual scientists nor used actual scientific tasks. Instead much of the research has used undergraduates working on artificial problems” (Schunn & Anderson, 1999, p. 388).

Due to the lack of generalizability across domains, Ilgen et al. (2012) highlight the importance of a domain-specific clinical reasoning model for medicine that adopts parts of a general scientific reasoning model but also incorporates domain-specific aspects of the clinical reasoning task. Elstein, Shulman, and Sprafka (1978) for example found that expert clinicians were more efficient and accurate due to better domain knowledge, not due to better general problem solving abilities. An example: It is important that a physician knows that s/he has to base her/his hypotheses on reliable evidence and test these hypothesis based on ruling out alternatives, but s/he needs to be able to draw valid conclusions based on test results by knowing what a certain test result implies for further reasoning.

To summarize, it is of major importance to keep in mind the domain-specific parts of scientific reasoning when studying it in specific contexts that are heavily based on specific content knowledge. Especially since the emergence of evidence-based medicine as “the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients“ (Sackett, 1997, p. 3), psychologists and educators have become interested in the domain-specific processes underlying clinical reasoning and how to foster them. A general model of scientific reasoning seems insufficient to represent clinical reasoning as there is important domain-specific knowledge that is essential to the reasoning process and that needs to be incorporated in order to understand and to enhance reasoning processes in clinical decision contexts. In the following chapter, clinical reasoning is first described from a process perspective, proposing an integrated model of diagnosing in the medical context, based on domain-specific factors. Then, the domain-specific competencies needed for successful clinical reasoning are explained.

3 Clinical Reasoning as a Requirement and Challenge for Physicians

The following chapter offers a definition of clinical reasoning in section 3.1 that incorporates cognitive and meta-cognitive processes which are based on knowledge. The cognitive processes are described with three different models of scientific and clinical reasoning which are then compared and analogies and differences outlined in section 3.2. Due to the emphasis on domain-specific knowledge in both clinical reasoning models, the nature of diagnostic knowledge is explained in section 3.3. A three-component framework of distinct parts of diagnostic competences is outlined in section 3.4. This concept is expanded with meta-cognitive error detection skills in section 3.5 as forth component.

3.1 Definition of Clinical Reasoning

In the field of medical education, various names have been used to label and subsume this critical process involved in evidence-based medical decision-making in diagnosis with the most prominent being: Clinical problem solving (Elstein & Schwarz, 2002), clinical decision making (Gill, Miller, Boucher, & Strauss, 1986), diagnostic reasoning (Croskerry, 2009), and clinical reasoning (Higgs, Jones, Loftus, & Christiensen, 2008). In this work, the term *clinical reasoning* is used to describe the process behind the diagnosis made by a clinician, to highlight its practical implications for clinical work and differentiate ‘clinical reasoning’ from the term ‘diagnostic reasoning’ used in domains like teacher education (Hesse & Latzko, 2009).

For advancements in teaching in medical education, scholars have been trying to understand the nature of the clinical reasoning process underlying successful medical diagnosing for several decades now (Norman, 2005). The findings that expert clinicians differed from novice physicians in their reasoning in the late 1970’s sparked the idea to examine the process of clinical reasoning what they labeled as medical problem-solving more closely by observational studies (Elstein et al., 1978; Neufeld, Norman, Feightner, & Barrows, 1981).

Most of the theory that serves as a basis for clinical reasoning models evolved from such observational studies as evidence suggested that interviews with physicians about their process of clinical reasoning seemed unreliable (Nisbett & Wilson, 1977). Methods to unravel the processes of clinical reasoning have become more and more refined, from using methods like think-aloud protocols or observing simulation-based performance (Elstein et al., 1978; Kassirer & Gorry, 1978) to participatory action research with computational mapping and modeling (Charlin et al., 2012). Therefore, models used to describe the clinical reasoning process vary greatly in complexity and detail, depending on the method of assessment. A comprehensive definition that describes what factors constitute clinical reasoning as a multi-level construct by Higgs et al. (2008) is:

> "Clinical reasoning (or practice decision making) is a context-dependent way of thinking and decision making in professional practice to guide practice actions. It involves the construction of narratives to make sense of the multiple factors and interests pertaining to the current reasoning task. It occurs within a set of problem places informed by the practitioner's unique frames of reference, workplace context and practice models, as well as by the patient's or client's contexts. *It utilizes core dimensions of practice knowledge, reasoning and meta-cognition* and draws on these capacities in others. Decision making within clinical reasoning occurs at micro, macro and meta levels and may be individually or collaboratively conducted. It involves metaskills of critical conversations, knowledge generation, practice model authenticity and reflexivity" (Higgs et al., 2008, p .4, after Higgs, 2006). [Emphasis added]

Thus, a holistic construct of clinical reasoning does not only describe the cognitive processes that happen when a physician is confronted with a complex patient case, but also considers other factors that influence the reasoning process. As Higgs et al. (2008) state, three different core dimensions are involved in successful clinical reasoning: The cognitive process of reasoning, practice knowledge needed for successful reasoning and meta-cognition to reflect and monitor. This may serve as a basic framework of what constitutes clinical reasoning: It is a cognitive process, based on practical knowledge and skills which is monitored by metacognitive control. It is important to take into consideration the multi-dimensional nature of clinical reasoning so that educators design effective learning methods for students to be good on all levels of successful clinical reasoning. To outline the basic construct of clinical reasoning and its components used in this thesis, the following sub-chapter describes models of clinical reasoning from a cognitive process perspective, followed by a (practice) knowledge perspective of diagnostic competences needed for successful clinical reasoning and error-related skills of meta-cognition to monitor the diagnostic reasoning process.

3.2 Clinical Reasoning as Cognitive Process: Models in Comparison

The cognitive view on clinical reasoning regards reasoning as a precursor for a diagnostic decision (Simmons, 2010). It evolves around solving a problem-solving task (Gilhooly, 1990) and follows a hypothetico-deductive structure (Charlin, Tardif, & Boshuizen, 2000). This means that the physician weights different hypotheses until one can be selected. This is done by searching, choosing and validating a hypothesis by matching it with knowledge representations of illnesses. The physician then decides for the most likely disease explaining the symptoms and history of the patient. This is followed by further steps of diagnostic validation or therapy (Rajkomar & Dhaliwal, 2011; Schmidt et al., 1990). In a simple basal sequential model, conscious clinical reasoning of a physician can be depicted as consisting of three consecutive cognitive steps. First, in the step of *data collection*, the physician gathers information to perceive and identify key-features of the disease with the help of guiding symptoms. During the second step of *clinical reasoning processing*, possible hypotheses of diseases are generated that might explain the observed symptoms and data from diagnostics tests (e.g., blood work) is interpreted. In the final step referred to as *validation*, the physician either neglects all hypotheses and begins with data collection again or accepts the most likely one and processes further (Graber, Gordon, & Franklin, 2002; Groves, O'rourke, & Alexander, 2003).

Based on these basic steps, Bowen (2006) proposes a more explicit clinical reasoning model that constitutes of steps analogous to the scientific reasoning steps as proposed by F. Fischer et al. (2014). It highlights the dynamic hypothetico-deductive nature of the clinical reasoning process that scholars in medical education advocate for (Coderre, Mandin, Harasym, & Fick, 2003; Eva, 2005): First, the physician is (1) confronted with the patient's history, which leads to immediate initial simple representations of what might be the case for the symptoms. Next, relatively unstructured (2) data-acquisition begins with for example physical examination, testing imaging and further questions to rule out some of the hypotheses (Rajkomar & Dhaliwal, 2011). Subsequently a more (3) accurate problem representation evolves and leads to the (4) generation of one or two main hypotheses (working diagnoses). Next, for these hypotheses, (5) matching knowledge representations in form of an illness script (see chapter 3.3) are selected. If this step fails and no matching illness script is found for the disease and the temporary working diagnoses are invalid and must all be neglected, the physician returns to (2) data acquisition. The emphasis of this iterative constant 'jumping' from (5) selecting a matching hypothesis to (2) multiple times during the process of diagnosing highlights the hypothetico-deductive nature of clinical reasoning. Finally, after finding a matching illness script, the validated (6) diagnosis is selected and communicated to colleagues for further treatment and possibly the patient. (Bowen, 2006)

Compared to F. Fischer et al. (2014) model of scientific reasoning, the emphasis lies on the dynamic reshaping of the hypothesis by reoccurring of data acquisition and refinement or negligence of the initial hypotheses as a constant chain until a final decision is reached (see figure 2 for comparison). The recurring chain of reasoning steps and the meta-cognitive processes involved in monitoring these steps are complex and make attempts to depict the full clinical reasoning process difficult. Clinical reasoning models therefore need to describe manifold contextual, cognitive and meta-cognitive variables involved clinical reasoning in a decision-dense, complex environment (Ilgen et al., 2012).

A recent attempt to depict the full variety of processes during clinical reasoning is the 'MOT Model of Clinical Reasoning*'* by Charlin et al. (2012) who argue in line with Nendaz and Bordage (2002) that the dynamic nature of the problem representation is the core of the diagnostic process. The model therefore gives a multifaceted picture of clinical reasoning across medicine. It is analogous to the proposed model by Bowen (2006), following a circular reasoning process but offering a holistic, detailed structure along the shaping of a dynamic representation during the diagnostic process. It emphasizes the multidimensional non-linearity of clinical reasoning by incorporating and highlighting context variables, cognition and *meta-cognitive monitoring* along the diagnosis to decide which step should follow the last one. The meta-cognitive processes monitoring the dynamic diagnostic process is highlighted in both models of Bowen (2006) and Charlin et al. (2012): The physician 'jumps' back and forth between the reasoning steps until the initial representation of the problem becomes more refined and further actions can be taken. This change in representation of the problem happens by monitoring the searching and selecting of matching illness scripts to explain the patient's problems. Clinical reasoning therefore is inseparable from meta-cognition and knowledge (Schmidt & Boshuizen, 1992). This dynamic change in representation relies on the mobilization and organization of specific knowledge types (Charlin et al., 2012). Therefore, medical educators advocate assessing specific, action-related knowledge types to determine a physician's ability to be accurate and successful in clinical reasoning (Ilgen et al., 2012; Page, Bordage, & Allen, 1995).

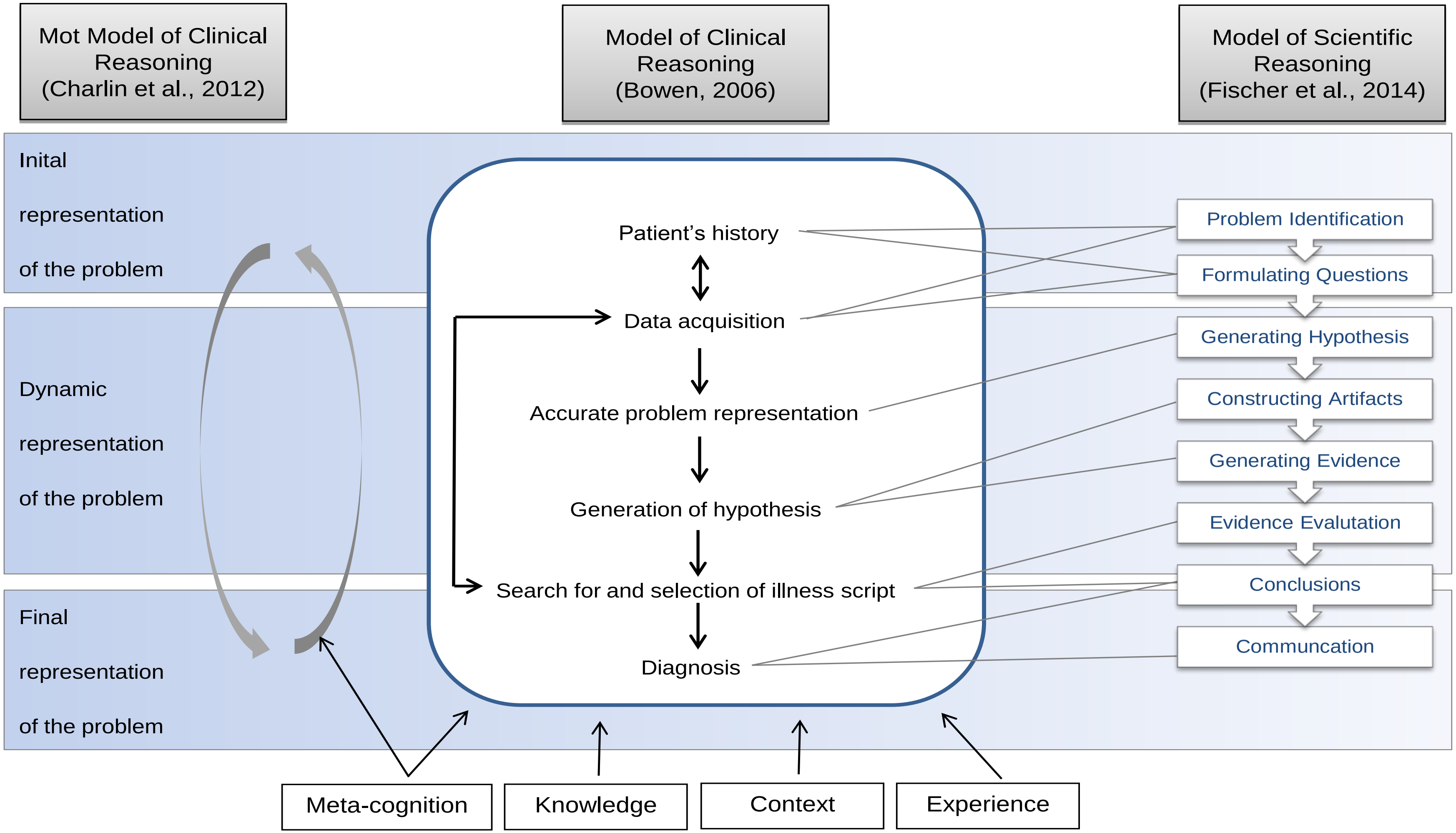

Figure 2. Integrative model of clinical reasoning. Comparison and integration of a proposed clinical reasoning model of ahypothetico-deductive reasoning model by Bowen (2006), the dynamic multi-level MOT model of clinical reasoning by Charlin et al. (2012) and the scientific reasoning model of epistemic steps by F. Fischer et al. (2014)

In summary, clinical reasoning has been a target of medical education research for decades now, leading to manifold synonymously used labels and different conceptualizations of the cognitive process. In this thesis, clinical reasoning is used to make a clear distinction towards reasoning in other domains such as teaching. It is the precursor to a final decision of a diagnosis and consists of three dimensions of cognition, (practical) knowledge and metacognition. The cognitive parts are often described as a process-model: it is a hierarchical and complex, non-linear chain of processes that involve the general steps of data collection, reasoning and validation. These steps have multiple sub-steps that can be depicted in varying detail (Bowen, 2006; Charlin et al., 2012). A domain-general model of reasoning like the one of F. Fischer et al. (2014) is, has similarities to a general process model of clinical reasoning steps. However, the two described models of clinical reasoning (Bowen, 2006; Charlin et al., 2012) emphasize the reliance of the reasoning process on domain-specific knowledge, meta-cognition, experience and the situation. All but the latter can be targeted by instructional interventions. Studies examining a domain-general reasoning model seem to try to generalize a process that is indisputably relying on domain and *content specific characteristics* (Elstein et al., 1978; Schmidt et al., 1990). In medicine, the strong reliance on biomedical content knowledge is essential for successful clinical reasoning (Ilgen et al., 2012). Therefore, the domain-specific knowledge and skills that enable successful clinical reasoning need to be addressed further (Higgs et al., 2008). The following paragraph describes a framework to classify the knowledge involved in diagnostic competences as prerequisite for successful clinical reasoning which is important to consider when aiming at interventions to foster clinical reasoning.

3.3 Diagnostic Competences as Knowledge Base for Clinical Reasoning

Early research on clinical problem solving in medical education aimed for generalizing the idea of a prototypical model of clinical reasoning (Elstein et al., 1978). In their study, Elstein et al. (1978) compared high-performing clinicians and low-performing clinicians on how their clinical reasoning is shaped in terms of hypothesis generation and evidence generation and evaluation. They found no difference between both groups in a general reasoning skill but in their knowledge base across patient cases. Their results gave a clear answer to neglecting the hypothesis that successful clinical reasoning is relying on domain-general reasoning skills. Clinical reasoning was found to have high *content specificity* as this reasoning process is highly dependent on biomedical knowledge. To understand how to support and enhance clinical reasoning, one has to understand the nature of clinical knowledge as prerequisite for successful, accurate clinical reasoning:

In recent years, there have been several scholars examining the question of what constitutes successful clinical reasoning: To a certain extent, clinical reasoning is based on the general ability to reason scientifically (Krasne & Stevens, 2010). This is due to its hypothetico-deductive nature (Charlin et al., 2000). However, medical expertise and reasoning success is heavily dependent on content knowledge (Schmidt et al., 1990): Expert clinicians and novices do not differ much in terms of general reasoning ability but rather in terms of clinical knowledge and knowledge representation (Feltovich, Coulson, Spiro, & Dawson-Saunders, 1992; Schmidt & Boshuizen, 1992).

This difference is rooted in the nature of how physicians shape the initial problem representation during the above-mentioned dynamic reasoning: The clinician constantly compares his/her mental representation of the problem with his/her internal representation of diseases (Bowen, 2006), also called *illness scripts* (Schmidt & Rikers, 2007). The structural theory of *scripts* as mental representations of ‘event knowledge’ (Nelson & Gruendel, 1986) comes from the results of an explorative study of Schank and Abelson (1977). In their conceptual framework, scripts are needed to successfully handle any situation the person is exposed to. These scripts consist of procedural knowledge about action and behavior (Schank & Abelson, 1977). Scripts as representations of procedural knowledge can be general or more specific towards a situation and vary in their flexibility to be reused or adapted to a similar (transfer) task (Schank, 1999). The more a person is exposed to a specific situation, the more his or her knowledge about what to do in that situation is refined and automated.

In the field of medicine, the automatization and refinement (*knowledge encapsulation)* happens when the physician is more and more exposed to a specific disease. As mentioned before, illness scripts are mental representations of diseases, constituted by clinical knowledge and constructed from past encounters with patients (Schmidt et al., 1990). They differ from prototypical representations by being enriched with extensive biomedical procedural and declarative factual knowledge behind that representation (Schmidt & Rikers, 2007; Woods, 2007). This elaborated knowledge about illnesses includes enabling factors (e.g., unprotected sexual contact for HIV), the pathophysiological processes of the illness (T-Helper cells are attacked by the virus, amount significantly drops, viral load increases) and the symptoms of the disease (fatigue, flu-like symptoms, swollen lymph nodes) (Boshuizen & Schmidt, 2008). The structure and complexity of illness scripts can be classified in three expertise stages:

Novice stage: The medical student learns biomedical conceptual knowledge about diseases. This knowledge is stored in a semantic network of representations. These networks are continuously validated and enriched. Reasoning in this stage is the inference between these concepts, so mostly linear and simple reasoning can successfully be accomplished. At the end of this stage though, linear reasoning chains become faster, concepts become more closely connected and indirect connections become direct connections (Boshuizen & Schmidt, 2008; Schmidt & Rikers, 2007).

Intermediate stage: After building up an extensive network of biomedical knowledge in the novice stage of learning, this knowledge base is now more and more connected to clinical knowledge of features of diseases experienced by patient contact. In this stage, medical students during clinical reasoning still consider and are able to name elaborated pathophysiological knowledge behind their diagnostic decisions in contrast to novices or experts ((Rikers, Schmidt, & Boshuizen, 2000; Schmidt & Boshuizen, 1992, 1993; VanLehn, 1996).

Expert stage: Biomedical knowledge and clinical knowledge are now being fully encapsulated and connected to experience with patients which constitute fully developed illness scripts. When an expert now is confronted with a patient case, one or several illness scripts are activated. In contrast to novices and intermediate learners, the illness script as a whole is activated and contains all knowledge connected to a specific disease. For finding the matching hypothesis, the encapsulated knowledge in the illness script is compared to the knowledge about the representation of the patient's problem. Non-matching illness scripts are deactivated again (Boshuizen & Schmidt, 2008).

In summary, diagnostic competences that are a prerequisite for successful clinical reasoning are based on domain-specific knowledge encapsulated in illness scripts. Illness scripts are built and extended throughout the medical expertise stages by being confronted with cases. Illness scripts are based on different types of knowledge. The following section describes the theoretical model of diagnostic competences based on three distinct types of knowledge.

3.4 Knowledge Components of Diagnostic Competences

As mentioned before, illness scripts incorporate knowledge about biomedical facts as well as knowledge about procedures. This is in line with a dualistic model of memory where knowledge is stored as different knowledge types, which are declarative and procedural knowledge. Declarative knowledge means knowledge about facts and is stored in a semantic network of propositions. Procedural knowledge means knowledge about situations and actions and is stored in episodic memory (Anderson, 1990). It is triggered as soon as a person enters a specific or a similar situation.

Paris, Lipson, and Wixson (1983) describe procedural knowledge as being composed of strategic and conditional knowledge. They found that experts differ from novices in terms of the use of strategies. Experts are more advanced and efficient with the use of problem-solving strategies than novices. This advanced use of strategies is explained by three factors. First, experts have a larger *conceptual* (or declarative) *knowledge* base about *what or that*. Second, they seem to be more advanced in procedural knowledge about *how* to do a task. This is referred to as *strategic knowledge.* However, merely the knowledge about what to do and how to do that is not sufficient for strategically performing a cognitive task for which *conditional knowledge*, the knowledge about the when and why an action might be successful, is needed. This distinction has been used to accurately describe and assess the diagnostic knowledge needed for diagnostic reasoning (Stark et al., 2011; Van Gog et al., 2004) and can be applied to any given diagnostic situation as exemplified by the following:

A patient comes to the emergency room with heavy breathing (dyspnea) lasting for several days already. The doctor assesses vital parameters like breathing frequency, d-dimers, O^2 concentration and heart rate and continues with a physical examination. He/She also asks several questions as part of the holistic data gathering, for example if the dyspnea occurs only under strain or if the patient had a long flight. He/She determines due to high d-dimers and a long flight in connection with a slightly swollen leg pulmonary embolism as most prevalent diagnosis. He/She knows that a high concentration of d-dimers in a blood sample is a sign for a clot that is dissolved (conceptual knowledge). The swollen leg is a sign for a deep vein thrombosis in this leg. These two symptoms lead to the working hypothesis of pulmonary embolism. To validate or rule out this hypothesis, further strategic actions need to be taken: he orders a computer tomography (validation on the basis of strategic knowledge) to see if there are clots in the lung that cause the heavy breathing (conditional knowledge).

To summarize, in the domain of medicine, domain-specific knowledge is represented in illness scripts. They evolve during learning from linear basic networks of biomedical facts that allow causal chains over a combination of clinical-situational and biomedical knowledge in the intermediate stage of the learner to a rich encapsulated illness script about a disease.

When confronted with a clinical situation, the physician activates one or more of these illness scripts as a whole and matches it with the characteristics of the patient's problem. This knowledge can be classified as conceptual knowledge about biomedical facts, strategic knowledge about medical tests and treatments and conditional knowledge about why these medical tests and treatments need to be done. As expertise grows, illness scripts are fully developed and the physician is able to diagnose complex cases. However, it has been shown that expert diagnostic performance still is subject to errors (Clarke et al., 2000; Graber, Franklin, & Gordon, 2005). To understand how diagnostic errors happen, it is important to differentiate types of errors and what cognitive processes are responsible for erroneous diagnoses. The following paragraph first defines diagnostic errors, depicts a taxonomy of kinds of errors and outlines the cognitive processes involved in erroneous diagnosing.

3.5 Meta-Cognition: Errors and Error Detection Skills

Experts are believed to be faster and solve problems better in their domain, with fewer errors involved, when compared to novices or intermediate learners; assuming they are more experienced and have more action routine (Chi, Glaser, & Farr, 2014). However, diagnostic errors happen also in an advanced stage of expertise, but experts seem to manage errors better than novices (Wilkinson et al., 2011). Due to the occurrence of diagnostic errors even in expert diagnostic reasoning, it is essential to identify what diagnostic errors are and what factors contribute to these errors.

Diagnostic errors are a subset of medical errors. Medical errors can be defined as "the failure of a planned action to be completed as intended or the use of a wrong plan to achieve an aim" (Kohn, Corrigan, & Donaldson, 2000, p. 4). The definitions of diagnostic errors specify that a diagnostic error is any mistake that happens during diagnosing leading to delay in the diagnostic process, missing a diagnose or misdiagnosing (Graber, 2005; Schiff et al., 2009; Zwaan & Singh, 2015).

Diagnostic errors can be classified in three major categories:

Non-fault errors happen if the disease is silent or masked, meaning it is hardly possible to diagnose it with the given data. *System errors* are rooted in weak spots in the health care system like malfunctioning communication between doctors and nurses or training and supervision. Mostly context accounts for this type of error. *Cognitive errors* are bound to lack of knowledge or flaws in the reasoning process within its three main steps: Data collection, clinical reasoning and validation. System 1 and 2 reasoning flaws account for this type of error mostly. In a systematic literature review about diagnostic errors, Schiff et al. (2009) identified *system errors* and *cognitive errors* as the most prevalent errors. *Non-fault errors* occurred the least. This is consistent with the study from Graber et al. (2005) reviewing erroneous patient cases to depict the following proportional share of error categories: They found that non-fault errors account for 7% of errors, errors that were system-related for 19%, errors that are based only on cognitive flaws account for 28% and both system-related and cognitive errors accounting for 46% of all diagnostic errors. Concluded, cognitive errors (alone and in combination) may be evidenced as the source in the majority of all diagnostic errors.

Therefore, cognitive diagnostic errors have been target of educational strategies for prevention for several years now (Croskerry et al., 2013a, 2013b). However, it is important to examine the underlying processes of diagnostic errors and account for the sources of these errors. A random diagnostic error can hardly be countered. However, systematically occurring errors can be prevented. Systematically occurring errors are called biases and can be differentiated into affective and cognitive biases (Croskerry et al., 2013a). Affective biases refer to the influence of emotions on the decision-making process, whereas cognitive biases are due to cognitive flaws acquired by experience. Examples of the most common cognitive biases influencing decision-making in the diagnostic reasoning process according to Norman and Eva (2010) are

- *availability bias*: Easily recalled diagnose is preferred;
- *premature closure*: Closure of the diagnose process without further tests etc.;
- *confirmation bias*: Data collection to confirm an initial hypothesis;
- *base rate neglect*: Ignoring the chances of different diseases and favor more unlikely ones; and
- *representativeness*: Recognizing prototypical features of diseases more than atypical ones.

The most common of these biases is premature closure (Graber et al., 2005). It is “the failure to continue considering reasonable alternatives after an initial diagnosis was reached” (Graber et al., 2005, p. 1493). This means that premature closure happens when the physician does not think about his/her reasoning process but reasons in an automated manner.

The models shown in figure 2 depict the process of clinical reasoning as conscious hypothetico-deductive reasoning. This conscious reasoning consists of forward and backward reasoning of the clinician until the problem is solved and the diagnosis can be validated (Groen & Patel, 1988; Patel & Groen, 1991). However, as expertise increases, the structure of illness scripts changes to increase speed and efficiency (Boshuizen & Schmidt, 2008). For greater efficiency in diagnosing, experts seem to rely on conscious clinical reasoning but also on a more intuitive (reasoning) approach of the clinician enabling shortcuts:

“The *intuitive approach* leans heavily on the experience of the decision maker and, therefore, uses reasoning that depends on inductive logic. Experienced decision makers recognize overall patterns (gestalt effects) in the information presented and act accordingly—action is recognition primed” (Croskerry, 2009, p. 1022). In the frameworks of decision making by Kahneman and Tversky (1979) and Kruglanski and Gigerenzer (2011) explained in Chapter 2.1, they subsumed that people use a dual-system processing when deciding during reasoning. The conscious part of decision-making is connected to reasoning and rationale, interpreting values, variables and symbols, making ‘sense’ of a patient’s history and symptoms: Analytic *system 2* mode of reasoning following the hypothetico-deductive reasoning structure described in figure 2. As medical expertise grows and structure of illness scripts change, the conscious reasoning is complemented more and more by context-based, fast but non-analytic *system 1* reasoning (Kahneman, 2011). Experts in medical decision making and clinical reasoning are making more use of system 1 reasoning system in later diagnosing, using more shortcuts in reasoning (Hammond, 1990; Wiswell, Tsao, Bellolio, Hess, & Cabrera, 2013). This rather unconscious mode of reasoning follows pattern recognition and is connected to associative memory (Lieberman, Gaunt, Gilbert, & Trope, 2002). Therefore, the clinician may skip steps of reasoning described in the hypothetico-deductive reasoning model summarized in figure 2. Both systems can operate rather independently and can also have different results of reasoning. When System 1 reasoning as the predominant reasoning strategy fails or is likely to produce erroneous results, System 2 reasoning becomes predominant to prevent errors (Kahneman, 2011; Kahneman & Frederick, 2002; Sloman, 1996).

When describing the modus operandi of expert reasoning, Ericsson (2004) states that experts use both system 1 and 2 processes during a task of problem-solving. Clinical reasoning becomes more efficient by becoming less reflective and more automated and intuitive. Expert reasoning is based on pattern recognition, but at the same time experts develop a ‘sensitivity’ when analytic reasoning and time to reconsider is needed in the diagnostic process when an error is likely to happen (Moulton et al., 2007; Rajkomar & Dhaliwal, 2011).

Even though both system 1 and system 2 reasoning are based on knowledge in an either implicit or explicit form, diagnostic errors cannot be countered by a large knowledge base alone as faulty or insufficient knowledge does merely account for diagnostic errors in an intermediate or expert stage (Croskerry et al., 2013a; Dror, 2011; Graber, 2009; Graber et al., 2005). Scholars advocate that the cognitive and meta-cognitive skills of experts are responsible for low diagnostic errors (Croskerry et al., 2013a, 2013b; Graber, 2009; Trowbridge, 2008). This might be due to the fact that the dynamic diagnostic process depicted in the MOT Model of diagnostic reasoning is monitored by meta-cognitive skills that monitor if the working diagnosis fits or needs to be neglected, an error has occurred or more data should be gathered (Charlin et al., 2012). This error management is connected to experts' diagnostic reasoning: Experts might not be more accurate in diagnostic reasoning but better in error management when an error in diagnosing has happened or is about to happen (Patel & Cohen, 2008; Patel et al., 2011): "Experts make errors, but surpass nonexperts theoretically with superior error management strategies that help them detect and recover from them more effectively" (Wilkinson et al., 2011, p. 42). Especially error detection is hypothesized to be a key skill that leads to experts making fewer diagnostic errors than novices. Even though the outcomes of experts' final diagnoses include fewer mistakes, they can make mistakes during the process of reasoning, but use (meta-cognitive) error management skills to avoid an actual mistake (Nyssen & Blavier, 2006; Wilkinson et al., 2011).

Based on the idea that accurate diagnostic reasoning requires meta-cognitive error detection skills, I propose a framework of diagnostic competences that not only incorporates three different types of diagnostic knowledge (Schmidmaier et al., 2013), but also meta-cognitive error detection skills in addition.

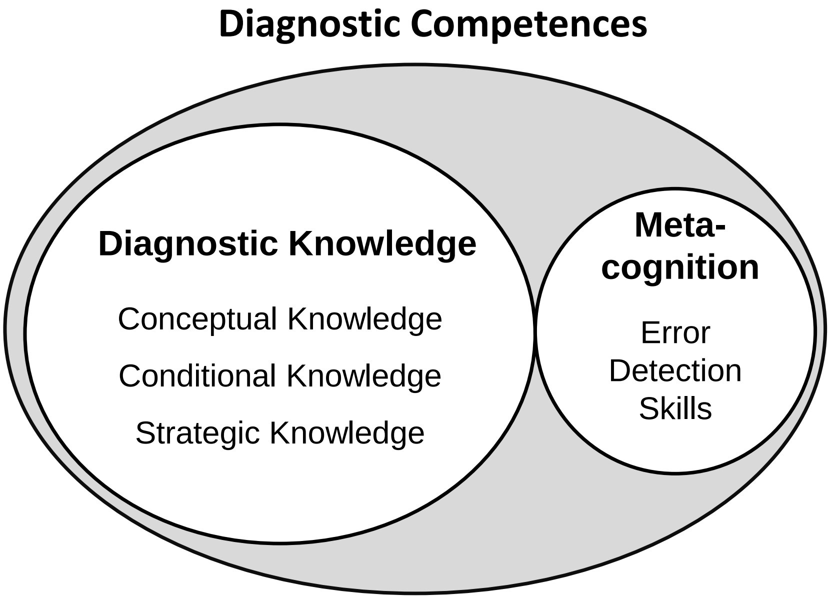

Figure 3. Holistic model of diagnostic competences. Model of diagnostic competences as proposed by Stark et al. (2011) with addition of meta-cognitive error detection skills for accurate diagnostic reasoning

How to teach error detection (and correction) is still an open question: Interventions to increase meta-cognition and cognition such as teaching the use of checklists or cognitive-forcing strategies that tried to teach students to stop and reconsider the diagnosis have been scarce but ineffective (Ely & Graber, 2015; Sherbino et al., 2014). Although the prevention of errors is of special interest in medical education, only assumptions and theories about possible underlying mechanisms have been discussed. Reason and Zapf (1994) argue for focusing on error management as a distinct key ability of experts to avoid and deal with errors in decision making. Error management consists of error detection and error recovery (Patel & Cohen, 2008; Reason & Zapf, 1994; Zapf & Reason, 1994). To increase diagnostic success, physicians must be knowledgeable of where to switch from non-analytic system 1 reasoning to analytic system 2 reasoning. This switch happens when a pitfall and possible error is detected (Dror, 2011). This skill might be taught by erroneous examples of what *not* to do, rather than by correct examples: Gartmeier, Bauer, Gruber, and Heid (2010) argue that experts excel in accuracy based on their experience with errors that enable them to perceive what might endanger a correct performance.

To summarize, clinical reasoning relies on two processes: Intuitive non-analytic system 1 and conscious analytic system 2 reasoning. Flaws in both modes lead to cognitive errors. These errors are the most prevalent causes (alone and in combination with system errors) of misdiagnosing and are caused by cognitive biases of physicians. Premature closure is the most common bias and leads to finalizing the reasoning process too soon and accepting a false hypothesis. There have been explanations of meta-cognitive and cognitive mechanisms that lead to fewer mistakes such as cognitive forcing strategies. So far, interventions either focused on fostering diagnostic knowledge or to reduce diagnostic errors which has been ineffective (Sherbino et al., 2014) or hypothesized to be effective (Croskerry et al., 2013a, 2013b). Therefore, scholars advocate error management skills consisting of error detection and error recovery to be the key to reducing diagnostic errors (Dror, 2011).

The previously discussed models of clinical reasoning do not incorporate error management, neither when adopting a process perspective (see Bowen, 2006) nor knowledge perspective on clinical reasoning (Stark et al., 2011). For a holistic assessment of important competences for accurate diagnostic reasoning, error detection skills need to be considered – especially when evaluating for effectiveness of an intervention in medical education. Learning through erroneous examples from the past might be key element of error detection in diagnosing. This dimension has so far been neglected. The following paragraph describes research on interventions to increase diagnostic competences and error detection skills of young physicians with interactive, erroneous example-based methods.

4 Fostering Diagnostic Competences and Error Detection Skills

The following chapter describes methods to teach diagnostic competences in medical education. First, section 4.1 provides existing literature on using erroneous cognitive modeling examples for teaching diagnostic competences. Therefore, the underlying assumptions of case-based learning methods are described and why it is beneficial for acquiring applicable, transferrable knowledge in section 4.1.1. Findings concerning incorporating written worked examples are outlined in 4.1.2. The advantages of case-based learning on the basis of cognitive load theory are explained in section 4.1.3. In section 4.1.4, video-based cognitive modeling examples, their use in education and influential characteristics are provided. These cognitive modeling examples have been assumed to be effective for learning correct and erroneous, which is summarized in section 4.1.5. Section 4.2 addresses why the task for students to provide peer feedback can be good for learning. First, feedback and peer feedback are defined in section 4.2.1. Then, studies on the effects of feedback on learning are highlighted in section 4.2.2. In section 4.2.3, the assumed positive effects of providing peer feedback on peer cognitive modeling examples on learning diagnostic competences are explained with its interactive nature and summarized.

4.1 Erroneous Cognitive Modeling Examples for Teaching

4.1.1 Teaching complex cognitive skills with case-based methods

The ‘structural knowledge perspective’ on clinical reasoning based on diagnostic competences as described before in chapter 2.2 assumes that successful clinical reasoning is based on clinical knowledge that is encapsulated into illness scripts (Schmidt & Rikers, 2007). This means an enrichment of declarative knowledge with clinical contexts of experience (Boshuizen & Schmidt, 2008).

Successful reasoning of experts relies on illness-scripts based on cases from past experience and is described as *case-based reasoning* (Aamodt & Plaza, 1994). Case-based reasoning means clinical reasoning that uses general clinical knowledge but incorporates case-specific knowledge from past experienced diagnosed patient cases stored in memory. When a physician is confronted with a new patient case, s/he uses this context specific knowledge and applies it to the new context (Kolodner, 2014; Leake, 1996).

To help students with the transition from factual declarative to practical clinical knowledge, educational methods that incorporate patient cases embedded in a clinical context should be used for teaching (Schmidt & Rikers, 2007). As expert reasoning is based on experienced cases in the past, it is important to use authentic patient cases in medical training. To ensure patient safety and avoid fatal outcomes of practicing as a physician, it is possible to use substitutes to real-life scenarios for learners. Early and constant exposure to patient cases in medical education is important for context-related reasoning abilities of medical students (Charlin et al., 2000). Case-based learning (CBL) proves an effective possibility to teach context-embedded knowledge for successful case-based reasoning (e.g. Stark et al., 2011).

Cases are not only vignettes or anecdotes of an incident that happened in practice, but also follow a pedagogical taxonomy or theory – the underlying story is planned and designed to match a typical yet important scene of practice while important parts for learning are highlighted (Shulman, 1992).

Case-based learning is a form of inquiry or problem-based learning similar in its constructivist nature of goals: Finding a solution to a problem. CBL is a more structured form of learning, guiding the learner through the problem-oriented process (Thistlethwaite et al., 2012). In a CBL learning environment, “a case, problem, or inquiry is used to stimulate and underpin the acquisition of knowledge, skills, and attitudes” (Williams, 2005, p. 577). These case-based teaching methods are designed to lead to applicable, active knowledge. This means knowledge that can be used in a real-life setting. Educational methods should be designed to foster active and avoid inert, ‘passive’ knowledge that is not applicable in real-life situations (Gruber, Mandl, & Renkl, 2000).

CBL environments can vary in the degree of explanation given to the student. They can be highly structured, guiding students through the case step-by-step (Kirschner, Sweller, & Clark, 2006) or less structured with an inquiry or problem-based approach to learning (Hmelo-Silver, Duncan, & Chinn, 2007; Schmidt, Loyens, Van Gog, & Paas, 2007). An advantage of the highly structured approach is that learners can focus on the learning content and do not need to worry about structure as it is a given (see cognitive load theory, 4.1.3) (Kirschner et al., 2006; Mayer, 2002). A less structured approach of CBL highlights the importance of self-directed, constructivist knowledge building (meaning the learner constructs the knowledge by him/herself) during learning, leading to complex, well-connected contextual knowledge and other skills learnt like self-directing. Both forms need guidance and instruction to be effective (Hmelo-Silver et al., 2007). Depending on the prior knowledge of the learner, the degree of structure should be varied from high degrees of provided instruction for beginners to less instruction for intermediates.

4.1.2 Incorporating worked examples in case-based learning

A written, text-based form of case-based learning that has been shown to be effective for complex skills (Atkinson, Derry, Renkl, & Wortham, 2000; Sweller & Cooper, 1985) in domains such as mathematics, physics and programming and also in medicine for the acquisition of diagnostic competences (Heitzmann et al., 2015; Kopp, Stark, & Fischer, 2008; Kopp, Stark, Kühne-Eversmann, & Fischer, 2009; Stark et al., 2011) is *worked examples* (or worked-out examples). They are often referred to as an “ultimate form of direct instruction” (Sweller, 2003, p. 250).

Worked examples are text-based examples that begin with a problem statement and depict an expert model’s procedure how to solve that problem that the learner can adapt. They should be designed to help learners solve similar transfer problems (Atkinson et al., 2000).

In an unstructured problem-oriented learning environment, students might focus on superficial features of the problem which takes up mental effort. Worked examples are advantageous over problem-based learning methods because the student can focus on content and schema acquisition rather than superficial features or structure. The learner can focus on cognitive skill acquisition and is guided through that process. This frees up mental capacity by reducing learning-detrimental parts cognitive load for learning relevant schema acquisition (Atkinson et al., 2000; Paas & Van Merriënboer, 1994; Renkl, 2014).

4.1.3 Explaining benefits of CBL with cognitive load theory

Cognitive load means the mental effort that is caused by different characteristics and the content of a learning environment. According to cognitive load theory, a learner has a certain amount of mental capacity of working memory when learning (Sweller & Cooper, 1985; Sweller, Van Merrienboer, & Paas, 1998). The learner is confronted with three types of cognitive load during learning that are limited by mental capacity:

Intrinsic cognitive load is caused by the difficulty and complexity of the learning task itself. It varies depending on the task and element interactivity of it (Brünken, Moreno, & Plass, 2010; Sweller, 2003). Intrinsic load needs to be managed to gain an optimal learning environment.

Extraneous cognitive load is caused by the design of the learning environment. Irrelevant or dissonant information of the presentation presented in the learning material or instructions that focus learners’ attention might be an example for this. This should be reduced to a minimum to leave mental capacity for learning-relevant processes of germane load (Brünken et al., 2010; Sweller & Cooper, 1985).

Germane cognitive load refers to the cognitive load that is caused by schema acquisition during learning. It is the learning-relevant type of load that should be focused on and maximized (Brünken et al., 2010; Paas & Sweller, 2014).

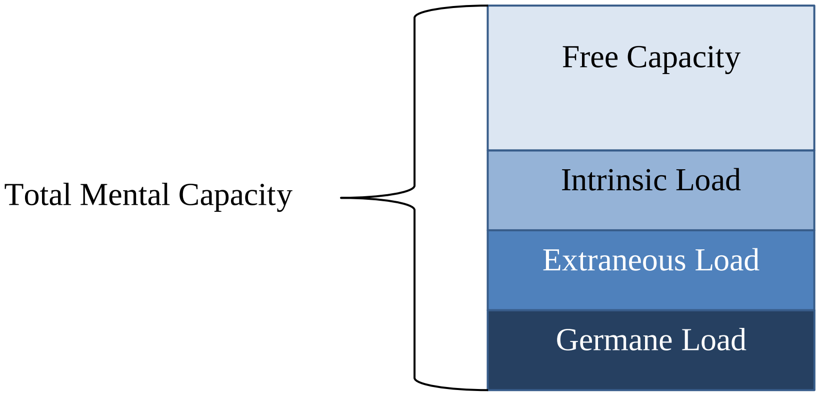

Figure 4. Schema of three types of load and total mental capacity

Cognitive load theory also received criticism as it seems to be impossible to falsify due to post-hoc explanation: The concept of a limited mental capacity is not challenged, but a clear distinction of different loads and how cognitive load must be reliably measured is criticized by scholars like De Jong (2010): Different cognitive load types are either measured via post-hoc explanation based on post-test performance or self-report. The bias of post-test performance measurement relates to an unfalsifiable explanation of performance with cognitive load types: If a student is well-performing, the increased load is germane whereas when s/he is underperforming it is extraneous load. Criticism of a self-report measurement of cognitive load highlights the inconsistent across studies (De Jong, 2010; Paas, Tuovinen, Tabbers, & Van Gerven, 2003). The debate how to measure different types of cognitive load is still open (Schnotz & Kürschner, 2007; Sweller, 2010), as this introspection method seems to be adding to the problems of subjectivity of perceived cognitive load by the students and differentiation in three distinct types of load: While some studies suggest a distinct three-factor structure of cognitive load in intrinsic, extraneous and germane load (Hadie & Yusoff, 2016; Leppink, Paas, Van der Vleuten, Van Gog, & Van Merriënboer, 2013), others argue for a reconceptualization of the construct in a two-factor construct with intrinsic and extraneous load (Leppink, Paas, Van Gog, van Der Vleuten, & Van Merrienboer, 2014; Sweller, 2010). Others argue that cognitive load measured with subjective introspection that cannot be differentiated should be labeled mental effort which is the perceived engagement in a task determined by self-efficacy, demand characteristics of the task and depth of processing (Kirschner & Kirschner, 2012).

4.1.4 Cognitive modeling examples

Based on cognitive load theory, educators should design the example-based learning environment so that working memory capacity is used optimally to increase learning gain. An important factor for this can be using a medium that is most suitable for the domain and learning task (Chandler & Sweller, 1991). Renkl (2014) distinguishes written worked examples, video-based modeling examples and analogic reasoning (i.e., comparing cases). While written worked examples are effective for learning in early stages, they lack authenticity when depicting patient cases and advanced students might lose motivation and attention (Van Gog & Rummel, 2010). Video based *modeling examples* are similar to written worked examples but differ in terms of format. They are videos of an agent performing a cognitive task and explaining the rationale behind the task (Hoogerheide, Loyens, & Van Gog, 2014). Video-based modeling examples can be used to increase the attention of both novice and non-novice learners and can offer high realism and authenticity of presented cases which is beneficial for learning (Große & Renkl, 2007; Renkl, 2014; Scheiter, Gerjets, Huk, Imhof, & Kammerer, 2009; Valmont, 1995; Van Gog & Rummel, 2010). Multi-modality of splitting up information in spoken text and video might enhance this effect: Due to auditory and visual working memory being autonomous, presenting information on both processing channels might reduce cognitive load that might overload the student when it is presented in just one medium (Leahy & Sweller, 2011). Lastly, video-based modeling examples show another person performing the task which might be used to compare one’s own performance with the performance of another person similar to oneself, leading to higher self-efficacy and an expectation to be able to perform the task as well (Bandura, 1981; Hoogerheide et al., 2014).

There are at least six important design options that need to be taken into account when creating modeling examples for learning, as each of them can differentially affect learning: (a) Form of presentation, (b) visibility of the agent, (c) hierarchical position of the model, (d) gender of the model and (e) model-observer-similarity, and (f) perceived expertise. The modeled task can be (a) live or on video, for increased accessibility and productivity (Blomberg, Renkl, Sherin, Borko, & Seidel, 2013). (b) The modeling agent performing the task can be visible or not visible. Van Gog, Verveer, and Verveer (2014) found that the visibility of the performing model can be important as it can serve as a cue for important features such as gestures. (c) The model depicted can be either a peer or an expert. While there is evidence that expert models should be used to teach skills that do not differ much in terms of automatization, abstraction and vocabulary (Boekhout, Gog, Wiel, Gerards-Last, & Geraets, 2010), Van Gog and Rummel (2010) argue that peer students should be used as a model to teach skills where experts and students differ in terms of automatization, vocabulary and knowledge structure – such as diagnostic competences. However, the task needs to match the age of the depicted model, leading to a better learning outcome: For example a young student performing a complex heart surgery process might not match (Hoogerheide, van Wermeskerken, Loyens, & van Gog, 2016). (d) The gender of the model on the other hand was not connected to identification or learning outcome (Hoogerheide, Loyens, & Van Gog, 2016). (e) Contrary to the assumption by various scholars that *model-observer-similarity* positively influences learning (Bandura, 1986; Schunk, 1987; Schunk, Hanson, & Cox, 1987), Hoogerheide, Loyens, Jadi, Vrins, and van Gog (2015) found no effect for a similarity of observer and model for modeling examples compared to similarity of an author of a written worked example.

Content-wise, Braaksma, Rijlaarsdam, and Van den Bergh (2002) found that (f) perceived competence of the model is important for a student to identify with the model. In their study, high-performing students learnt more from a high-performance model whereas weaker students learnt more from a weak model. This seems to be true for students who did not acquire high levels of prior knowledge before observing the performance. However, advanced prior knowledge conflicts with correct examples: When students already have acquired cognitive schemata on how to solve a problem, but the presented step-by-step solution of the example conflicts with their own knowledge, learning gain decreases (Kalyuga, 2007; Kalyuga, Ayres, Chandler, & Sweller, 2003; Renkl & Atkinson, 2003). Therefore, errors can be integrated into the examples to increase motivation and gain attention from advanced students to decrease conflict and redundancy of the observed performance and own prior knowledge (Große & Renkl, 2007).

4.1.5 Errors in cognitive modeling examples

The positive effects of incorporating errors in worked examples have been exemplified in several empirical studies by for example Große and Renkl (2007), Siegler (2002) or Stark et al. (2011). Erroneous worked examples are engaging for intermediate learners when the *expertise reversal effect* would decrease learning gain from worked examples (Kalyuga et al., 2003): The expertise reversal effect means that with increased expertise, high-structured learning environments lose effectiveness: In a novice stage of learning, problem-solving activities impose high amounts of cognitive load to the learner and step-by-step worked-out solutions are highly useful to guide through the process. In an intermediate stage of knowledge, a presented step-by-step solution with correct worked examples is not as beneficial as it is for beginners. Intermediate learners were found to profit more from problem-oriented learning tasks the higher their prior knowledge is (Große & Renkl, 2007;

Kalyuga et al., 2003; Renkl & Atkinson, 2003). Therefore, problems in form of for example errors integrated into worked examples offer problem-solving learning benefits.

Große and Renkl (2007) point out that learning from erroneous examples offers three advantages. First, being exposed to an error can engage the learner in reflection of what s/he has observed. Second, presence of errors can lead to cognitive conflicts of existing knowledge and what has been observed, enriching or even correcting existing knowledge (Kapur, 2008). Third, errors can decrease the probability of making this error yourself (Gartmeier et al., 2008; Siegler, 2002). This effect can be explained as follows: Being exposed to errors during learning leads to the acquisition of *negative knowledge* (knowledge needed for error detection skills described in chapter 3.5), which is knowledge about how to detect and avoid errors (Gartmeier et al., 2008). Observing erroneous motor skill execution has been shown to lead to a decrease of errors in own motor abilities in a post-test transfer task (Blandin & Proteau, 2000). Observing an erroneous performance of another person was found to be as beneficial as making the error oneself for decreasing the likelihood of these errors in one’s own later performances (Badets, Blandin, Wright, & Shea, 2006).

These findings support the hypothesis of the crucial beneficial effects of learning by observing erroneous modeling examples to increase future correct own performance and decrease future incorrect own performance via error detection skills. This, however, is only true if the error is found and triggers elaboration to be correcting the error (Renkl & Atkinson, 2003). There are guidelines how to guide students through observing erroneous performance in order to be effective for learning: The first step of finding the error can be supported by providing the learner with correct knowledge of results (*KR – knowledge of results*) when being exposed to erroneous examples (Badets & Blandin, 2004, 2005).

When the error is found, the second step of learning-relevant elaboration of the error is also not intuitively done by the learner: VanLehn (1999) highlights the importance of instruction to elaborate on erroneous learning material as otherwise learners follow shallow processing strategies. Elaboration of learning content does not happen automatically when the learner is confronted with the learning material. In order to construct sound knowledge from the erroneous examples, guiding learners to self-explain on the rationale behind the error was found to be beneficial for learning. Self-explanation can be achieved either via training or prompting students to do so, ultimately leading to deeper elaboration and therefore higher learning gain (Atkinson et al., 2003; Chi & Wylie, 2014; Renkl, 2002, 2014; Renkl & Atkinson, 2003; Renkl, Stark, Gruber, & Mandl, 1998; Stark et al., 2011).

On the basis of the results of the effectiveness of self-explanation on learning, there might be similar or higher effects of explaining the observed to others, based on a taxonomy by Chi (2009) of activity and interactivity. She points out in her *ICAP-Framework* that the engagement and learning outcome varies according to certain activity modes triggered by a learning activity that are linked to different cognitive processes: These modes range from the least effective *passive* over *active* and *constructive* to most effective *interactive* (Chi, 2009; Chi & Wylie, 2014). Passive learning offers the least outcome and means that the learner just passively is presented with learning material. Active learning means active engagement of the learner and promotes learning better than passiveness as it incorporates learning new knowledge and activating own prior knowledge. Constructive learning is more effective as it means higher-order cognitive processes such as newly producing something own which involves integrating, re-organizing or repairing own knowledge. Interactive learning activities are proposed to be the most effective as they are similar in nature as constructive ones but incorporate the partner’s contribution as well (Chi, 2009). This beneficial effect is based on the idea of Vygotsky (1987) that a learner in a social situation learns more than learning alone by being provided by another person with a ‘zone of proximal development’. By having differences in prior knowledge and approach to learning, building on each other’s’ ideas and elaborating on the topic, students during interaction learn not only from the content but from each other (Weinberger, Ertl, Fischer, & Mandl, 2005). As described earlier, just exposing students to this zone of proximal development is not enough for learning as the learner might still remain in a passive mode. Instructing the learner to engage in a constructive or interactive task might lead to higher-order cognitive processes. One effective approach to engage students in interactive learning activities and therefore better elaboration of the observed example is the task to *provide peer feedback* on the observed student's performance (Narciss, 2008; Nicol et al., 2014). This might trigger reflection and deeper elaboration which then might lead to more learning gain (K. Cho & MacArthur, 2011). Therefore, the following paragraph examines the beneficial effect of feedback on the provider.

4.2 Learning by Providing Feedback on Peer Modeling Examples

4.2.1 Definition of peer feedback

Feedback on students' performances in (for example simulation-based) medical education methods is an essential practice to foster learning (McGaghie, Issenberg, Petrusa, & Scalese, 2010). It was even referred to as the "heart of medical education" (Branch & Paranjape, 2002, p. 1185). Feedback means "information provided by an agent (e.g., teacher, peer, book, parent, self, experience) regarding aspects of one's performance or understanding" (Hattie & Timperley, 2007, p. 81). *Peer feedback* in that respect means that a student judges another student's performance. This can happen either as score (summative) or qualitatively (formative) with oral or written feedback (Strijbos & Sluijsmans, 2010; Topping, 1998). Peer feedback tasks interactively engage students to reflect and collaborate (Strijbos & Sluijsmans, 2010). Peer feedback and peer assessment are sometimes differentiated. Liu and Carless see peer feedback as the process of giving detailed comments without grades while peer assessment involves grading (Falchikov, 2001; Liu & Carless, 2006). Topping (1998, 2009) uses the term peer assessment to describe the task, while peer feedback is the outcome of assessment task. A distinction of peer assessment and feedback varies across studies and is not consistently done (Liu & Carless, 2006), leading to a synonymous use in this thesis.

4.2.2 Effects of (peer) feedback on learning

Hattie and Timperley (2007) suggest that feedback should aim towards positive and negative aspects of performance and ideally offer suggestions for improvement. A review by Van Zundert, Sluijsmans, and Van Merriënboer (2010) about the beneficial conditions and the effect of peer feedback shows that most studies focus on fostering skills related to peer feedback and very few on the effect of peer feedback on the acquisition of domain-specific skills. The five studies that did examine the effect on domain-specific skills demonstrate the potential of peer feedback for learning. They conclude that there is a clear lack of studies examining the added benefit of peer feedback as interactive learning opportunity for domain-specific skills. Out of the studies they reviewed, only one examined the effect of learners' thinking style on the process of feedback reception during peer feedback rounds (Lin, Liu, & Yuan, 2001). K. Cho and MacArthur (2011) emphasize the beneficial learning effect of receiving peer feedback over expert feedback if quality criteria are met. Gielen, Peeters, Dochy, Onghena, and Struyven (2010) found that feedback with justification (elaboration) enhanced learning compared to feedback without justification – with accuracy of the peer feedback playing a minor role. This is contrary to the findings by Strijbos, Narciss, and Dünnebier (2010) who found beneficial effects of receiving general over-elaborated feedback, possibly due to perceived competence level of the provider. The results of quality criteria for receiving peer feedback are mixed (Kollar & Fischer, 2010; Topping, 2010). The studies depicted mostly focused on peer feedback reception. However, few studies examine the isolated effect of *providing* peer feedback as interactive task for learning described in the following.

4.2.3 Effects of providing peer feedback on learning

Concerning the effects of peer feedback on the provider, studies that only investigated the effects of feedback provision isolated from reception are scarce, even though there is evidence that the provision of feedback as an interactive learning activity is claimed to be fruitful for own learning gain (Nicol et al., 2014). Providing peer feedback can provoke higher level thinking processes such as diagnosing and correcting errors in assessed performance (Patchan & Schunn, 2015) with deeper elaboration and judgement of the assessed performance (Nicol et al., 2014) by comparing the peer's performance with internal quality criteria and own performance (Topping, 1998).

Studies that investigate the isolated effect of providing peer feedback found that it positively influences learning under certain circumstances. For example, Althauser and Darnall (2001) found that the quality of provided peer feedback on the essay of a peer has a positive effect on the quality of the student's own work. Davies (2000) found similar results on subjective gain of own domain-specific skills through the process of peer feedback provision. Li et al. (2010) and Y. H. Cho and Cho (2011) both examined the effect of providing peer feedback on the development of the providing students' own abilities to design technical research drafts. Both studies found a beneficial effect of providing feedback to peers' research drafts on the providing students' performance if the student was engaged and provided high quality (elaborated) feedback. High quality of peer feedback in both studies meant degree of elaboration, incorporating both positive and negative aspects while giving constructive feedback about what to improve. Low quality meant superficial comments with no explanation. Providing superficial peer feedback was not found to be fruitful for learning due to the assumed lack of engagement and elaboration by the provider. They concluded that providing feedback on a weak peer performance plays a major role in reviewing one's own weaknesses and to be(come) able to avoid own errors (Y. H. Cho & Cho, 2011) – which is in line with the assumption that weak (peer) models lead to the acquisition of negative knowledge to avoid errors (see chapter 4.1.2) if students really engage in the feedback task and elaborate on strengths and weaknesses.

The results by Li and colleagues (2010) and Y. H. Cho and Cho (2011) are in line with the ICAP Framework about interactive learning activities in the sense that the added value of interactive learning environments depends on students' elaboration and engagement (Chi & Wylie, 2014). Students lack the ability to engage in essential collaboration activities (such as constructive elaboration) in interactive learning environments and need external guidance how to do so (Kollar, Fischer, & Slotta, 2007; Rummel & Spada, 2005; Weinberger, Stegmann, Fischer, & Mandl, 2007). Prins, Sluijsmans, and Kirschner (2006) found that medical students lack the ability to give elaborated, high-quality feedback, lacking structure, reflection and suggestions for improvement. They advise educators to train providing peer feedback beforehand with clear quality criteria or prompt reflective feedback activities to increase the quality of provided peer feedback during the provision task.

In summary, case-based learning activities in medical education are necessary and have been shown to be effective to teach diagnostic competences for successful clinical reasoning. Worked examples that are integrated in case-based learning offer an effective way to teach these skills. Authentic, video-based cognitive modeling examples could be a tool to teach diagnostic competences in a live-like setting as they have been used successfully in other domains already. Erroneous cognitive modeling examples might prove to be advantageous over correct cognitive modeling examples as advanced students benefit more from learning with them. Additionally, error detection skills could be fostered by being exposed to negative knowledge about diagnostic errors in erroneous examples. A downside of example-based learning that was highlighted is that students tend to be passive and do not actively elaborate on the examples without (inter-)active tasks given to them. This elaboration though is necessary in order to learn effectively. Students' elaboration can be triggered by self-explanation prompts. However, the interactive task to providing peer feedback on the observed model peer could yield higher learning gains due to an assumed advantage of interactive over active learning. The effect of providing peer feedback on cognitive modeling examples for learning complex cognitive skills in a domain such as medicine is yet to be tested. If and how an instructional intervention based on these assumptions can be used, is target of two empirical studies conducted. The following paragraph gives an overview of both studies.

5 Overview of the empirical studies of this thesis

Based on the theoretical assumptions of the beneficial effects of erroneous cognitive modeling examples and providing peer feedback on learning gain, two empirical studies were conducted which examined how providing peer feedback on observed erroneous cognitive modeling examples of complex differential diagnoses influences the providers' acquisition of diagnostic competences.

Study 1 examined if and how cognitive modeling examples could be used for acquiring diagnostic competences in medical education. Therefore, it was investigated if (a) erroneous or correct cognitive modeling examples are better for learning diagnostic competences and if (b) providing peer feedback on the examples adds to this assumed beneficial effect. Diagnostic competences were operationalized as diagnostic conceptual, strategic and conditional knowledge and error detection skills.

The hypotheses were created based on the theoretical foundation in chapter 4.1 that erroneous examples are better than correct examples for acquiring diagnostic competences. Further, it was assumed that students providing peer feedback to the peer students in the cognitive modeling examples show higher gains of diagnostic competences. To examine the relation to mental effort and degree of elaboration, post-hoc analysis of the provided peer feedback was related to acquiring diagnostic competences. The sample consisted of 121 medical students in their seventh semester randomly assigned to five groups in a 2x2 plus control group design. The groups either only watched erroneous (group 1) or correct examples (group 2), or provided peer feedback on erroneous (group 3) or correct examples (group 4). A control group learning with a textbook (group 5) did not watch any of the cognitive modeling examples.

The results of Study 1 indicated that cognitive modeling examples were better for acquiring conceptual knowledge parts of diagnostic competences compared to the control group, with no overall increase in diagnostic competences. Erroneous cognitive modeling examples were better than correct ones for acquiring error detection skills. Providing peer feedback on cognitive modeling examples led to a detrimental effect on the acquisition of overall diagnostic competences. Students providing peer feedback reported more mental effort. The degree of elaboration of peer feedback was only related to the acquisition of diagnostic competences when students learned with erroneous cognitive modeling examples. However, this was not the case for correct examples. It was hypothesized that the task to provide peer feedback led to higher mental effort which was detrimental for learning.

Due to the surprising negative influence of providing peer feedback on the acquisition of diagnostic competences and the positive relationship between the degree of elaboration of the provided peer feedback on erroneous cognitive modeling examples and the acquisition of diagnostic competences in Study 1, a follow-up study was conducted. It was examined how the task to provide peer feedback on erroneous cognitive modeling examples can be designed to be beneficial for acquiring diagnostic competences based on theoretical assumptions of possible influential factors for peer feedback provision.

In Study 2, a short training how to effectively provide peer feedback was integrated at the start of the intervention that was assumed to increase degree of elaboration of peer feedback. Further, the effects of written versus spoken medium of expert feedback and receiving expert feedback or not were examined. We contrasted the task to provide (a) written or spoken peer feedback on erroneous cognitive modeling examples and (b) students received – or did not – a written expert feedback on the observed diagnosis after each peer feedback provision task. The sample in Study 2 consisted of 134 students from seventh to eleventh semester. They were randomly assigned to four experimental and a control group. All four groups watched erroneous cognitive modeling examples and either provided spoken peer feedback (group 1), written peer feedback (group 2), or provided spoken peer feedback and received expert feedback (group 3) or provided written peer feedback and received expert feedback (group 4). The control group in Study 2 watched the erroneous cognitive modeling examples and did not provide peer feedback. Results revealed that receiving expert feedback had no effect on degree of elaboration or the acquisition of diagnostic competences, which might have been due to redundancy or inadequate timing of expert feedback. A significant main effect of communication medium was found for spoken peer feedback provision on the acquisition of diagnostic competences. This effect was explained by the degree of elaboration: students providing spoken peer feedback provided more elaborated feedback, which influences the acquisition of diagnostic competences.

In the following Chapter 6, the first empirical study is described in detail. In Chapter 7, the second empirical study is explained, followed by a comparison of both studies in Chapter 8.

6 Study 1: Watching People Fail

Improving Diagnostic Competences by Peer Feedback on Erroneous Modeling Examples

Context: Fostering diagnostic competences is a major concern in medical education. Effective interventions to foster diagnostic competences to decrease errors in diagnosing need to be examined. Interactive learning with video examples is hypothesized to be effective for learning these competences.

Objectives: This study investigated whether students learn more from providing feedback (vs. just observing) on erroneous or correct video examples of clinical reasoning.

Methods and Sample: The sample consisted of advanced medical students ($N = 121$) randomly assigned to the four conditions of a 2x2 factorial design and a control condition. It was investigated if (a) erroneous or correct video examples of diagnosing a patient with acute dyspnea and (b) the students providing peer feedback on the observed reasoning process in the examples increased their diagnostic competences which consist of conceptual, strategic and conditional diagnostic knowledge as well as error detection skills.

Results: Video-based cognitive modeling examples across experimental groups were superior for learning relative to a control group that learned by textbook reading, $t(40) = 2.651$ $p = .011$, $d = .84$. Observing erroneous examples was more effective for the acquisition of error detection skills than observing correct examples. $F(1,95) = 4.290$, $p = .041$, $\eta_p^2 = .046$. Surprisingly, students who provided peer feedback acquired less overall diagnostic competences than students who just observed the model, $F(1,95) = 5.066$, $p = .027$, $\eta_p^2 = .014$. A negative main effect of the provision of peer feedback on modeling examples (compared to observing the modeling examples) was found for conceptual, $F(1,95) = 6.546$, $p = .012$, $\eta_p^2 = .068$, and strategic parts of diagnostic knowledge, $F(1,95) = 8.827$, $p = .004$, $\eta_p^2 = .089$. There was no difference for conditional knowledge and no interaction effect between the factors for any outcome. Unlike for error detection skills, it made no difference for the three diagnostic knowledge components whether examples were erroneous or correct.

Conclusions: Erroneous examples of diagnoses deserve a standing in medical education practice and should be incorporated into medical education because they help students acquire error detection skills. Providing peer feedback as a learning intervention should be used carefully as the results of this study show a detrimental effect for the provider on error detection skills, conceptual knowledge and strategic knowledge – which is possibly due to cognitive overload. Training in feedback provision might decrease this effect.

6.1 Theoretical Framework

6.1.1 Diagnostic competences: Diagnostic knowledge in medical education

A recent study about the estimates of annually serious medical diagnostic errors reports that 5% – 12 million US adults – experience misdiagnosing in outpatient care (Singh, Meyer, & Thomas, 2014). Diagnosing is a difficult, complex and error-prone task for physicians in their everyday work (G. J. Kuhn, 2002). Reducing errors with educational means is emphasized as the key (Norman & Eva, 2010). How to effectively teach against errors is still an open question: Successful, accurate diagnostic reasoning relies on diagnostic competences: Diagnostic competences consist of diagnostic knowledge and meta-cognitive skill (Higgs et al., 2008). Gaining diagnostic knowledge means the acquisition of *advanced illness scripts*, i.e. representations of knowledge about facts and procedures connected to patient cases and situations (Mamede et al., 2012). Advanced illness scripts consist of conceptual knowledge and procedural (strategic and conditional) knowledge on how to apply the known concepts (Schmidmaier et al., 2013). Conceptual knowledge refers to knowledge about biomedical facts, strategic knowledge to knowledge about procedures to validate or reject possible hypotheses and conditional knowledge to knowledge about the reasons behind these procedures (Kopp et al., 2009; Stark et al., 2011). However, increased accuracy of expert physicians might not be just a function of superior knowledge, but also be related to advanced error-related meta-cognitive skills that make experts calibrate their error-prone decision making patterns (Graber, 2005).

6.1.2 Error detection skill as part of diagnostic competences

In addition to the diagnostic knowledge types proposed by Stark et al (2011), diagnostic competences involve meta-cognitive skills (Higgs et al., 2008). I propose a holistic diagnostic competences framework with meta-cognitive '*error detection skills'* as depicted in Figure 3 based on knowledge about possible pitfalls that may lead to errors during diagnosing (Dror, 2011). Scholars like Wilkinson et al. (2011) and Nyssen and Blavier (2006) claim that error detection skills are a distinct set of subskills of medical expertise, leading to less errors and more accurate and successful diagnostic reasoning.

The assumption that there is more involved in accurate diagnosing than diagnostic knowledge can be based on the process idea of diagnostic reasoning: Conscious diagnostic reasoning follows typical hypothetico-deductive reasoning steps: (1) (data) collection of the most important features of the patient case, (2) reasoning and creating hypotheses about underlying diseases on the basis of medical knowledge, and (3) validating these hypotheses by testing them (Lawson & Daniel, 2011).

As expertise increases, diagnostic reasoning process becomes more and more automated by pattern recognition for more efficiency (Anderson, 1987; VanLehn, 1996; Wiswell et al., 2013). Nonanalytic, (automated) system 1 processing is the predominant reasoning strategy. Physicians use a more analytic reasoning mode (system 2) when nonanalytic processing as predominant strategy is insufficient and leads to incorrect results. To switch from a rather nonanalytic to analytic reasoning mode, a physician must become aware that an error is likely to happen and automated processing might fail (Kahneman, 2011). Meta-cognitive *error detection skills* are required to 'assist' nonanalytic reasoning (system 1) with conscious hypothetico-deductive analytic reasoning (system 2) at the right time when an error might happen or has happened (Dror, 2011; Mamede et al., 2010). Studies examining medical expertise have shown advanced error detection skills in expert clinicians compared to novices (Gartmeier et al., 2010; Wilkinson et al., 2011). To my knowledge, medical educational studies have so far not focused on fostering error detection skills with interventions.

In summary, it is important to teach diagnostic competences. To reduce diagnostic errors, both knowledge and meta-cognition must be taught. Meta-cognitive error detection skills are needed to know when to use analytic reasoning based on negative knowledge during automated diagnostic reasoning. The meta-cognitive part has mostly been neglected in frameworks depicting diagnostic competences as a three component knowledge construct (Stark et al., 2011). Therefore, educational methods that focus on this part of diagnostic competences need to be examined. Especially interactive learning interventions (Chi & Wylie, 2014) with feedback rounds (Hattie & Timperley, 2007; Vollmeyer & Rheinberg, 2005) and model learning (Hoogerheide et al., 2014) were proven to be strong educational methods in other domains. Situations such as OSCE assessments, simulation training or medical clerkships, which are often used in medical training, might be video recorded and utilized as cognitive modeling examples for learning.

6.1.3 Teaching diagnostic competences with cognitive modeling examples

Experts use case-based reasoning as a predominant strategy to solve problems effectively by using past experience to solve newly presented problem cases (Aamodt & Plaza, 1994; Riesbeck & Schank, 2013). To develop essential medical reasoning skills, students should be exposed to authentic case-based learning scenarios so that they acquire knowledge in a clinical setting where they later have to apply and transfer it (Schmidt & Rikers, 2007). Learning scenarios which are based on real-life cases help students acquire knowledge needed for solving similar transfer cases and situations (Gruber et al., 2000; Stark, Mandl, Gruber, & Renkl, 1999). However, current medical education has been shown to expose students far too little to clinical cases (Schmidt & Mamede, 2015; Wimmers, Schmidt, & Splinter, 2006). Such lack to exposure leads to inert (passive) knowledge, which is knowledge that is remembered but cannot be applied in a practical situation (Gruber et al., 2000). Example-based instructional methods use real patient cases to teach students active knowledge that is applicable in practical situations without imposing too much cognitive load on the students (Cooper & Sweller, 1987; Stark et al., 2011). This means that students have more cognitive resources available for relevant processes during learning. Examples can be text-based written worked examples. For more authenticity, fidelity and increased focus of attention, video-based *cognitive modeling examples* can be utilized (Renkl, 2014). Cognitive modeling examples show a real or fictional agent performing a cognitive task while explaining the rationale behind it (Hoogerheide et al., 2014). Such videos have been used successfully to foster various behavioral and meta-cognitive skills such as self-regulation, self-assessment and task-selection skills (Braaksma et al., 2002; Kostons, Van Gog, & Paas, 2012; Van Gog et al., 2014; Zimmerman & Kitsantas, 2002). The educational use of cognitive modeling examples has not yet been examined for complex cognitive skills such as diagnostic competences necessary for accurately diagnosing complex medical problems.

6.1.4 Learning from errors in cognitive modeling examples

A cognitive modeling example shows an agent performing a cognitive task while explaining the rationale behind it. This real or fictional agent can be an expert (Boekhout et al., 2010) or a peer (Van Gog & Rummel, 2010). Cognitive modeling examples can be erroneous or correct when utilized for learning. A correct example (i.e., a well-performing model) should be used when teaching novice students, because correct examples do not confuse the student with conflicting information and leave cognitive resources to be used for learning. An erroneous example (i.e., a weak-performing model) should be used for students who are on an advanced level of knowledge to increase their engagement, motivation and learning gain (Van Gog & Rummel, 2010). Stark and colleagues (Stark et al., 2011) report beneficial effects of learning from erroneous examples for both conceptual and procedural diagnostic knowledge, whereas effects on reduced errors in own performance have been hypothesized but not yet studied. Furthermore, Gartmeier et al. (2010) claim that being exposed to erroneous cases might create *negative knowledge* about how not to diagnose and detect and avoid these errors in the future. Especially differential diagnosis of pulmonary embolism and similar respiratory diseases are error-prone and difficult to diagnose and therefore need educational attention (Schiff et al., 2009).

Erroneous modeling examples are beneficial for knowledge acquisition when the learner recognizes the error and elaborates on them (Renkl & Atkinson, 2003). Such elaboration leads to a contrasting of the observed erroneous example with internal standards of a correct procedure for the observed performance or external standards provided by a teacher or expert (Van Gog, 2015). Elaboration can be achieved via active self-explanation (Renkl, 2002) or explaining it to others. Explaining it to others might be a beneficial approach to engage students in interactive-constructive elaboration which leads to better learning (Chi & Wylie, 2014).

6.1.5 Learning through providing peer feedback on modeling examples

Explaining important aspects of an observed performance to someone else leads to engagement and elaboration (Topping, 1998). Providing peer feedback is challenging as medical students often face problems to give useful, elaborated feedback (Prins et al., 2006). Studies examining the beneficial effects of feedback rounds on learning found that feedback-provision can enhance one’s own learning gain (Y. H. Cho & Cho, 2011; Li et al., 2010; Rouhi & Azizian, 2013). Students providing peer feedback elaborate more on the content than when receiving it, because they have to diagnose weaknesses in the performance (Lundstrom & Baker, 2009; Patchan & Schunn, 2015): They compare the peer's performance with the criteria for successful performance and then compose and deliver a synthesis to the peer receiving the feedback. This active comparison should lead to an in-depth elaboration of the observed behavior which should interact with the beneficial aspects of elaborating on erroneous cognitive modeling examples. For example, when asked to provide feedback to video-based cognitive modeling examples, learners with sufficient prior knowledge are more likely to recognize errors in the examples and elaborate more deeply which leads to better learning (Renkl & Atkinson, 2003). However, the quality of the provided peer feedback in terms of elaboration is a key characteristic of peer feedback as a learning-enhancing activity (Y. H. Cho & Cho, 2011). Such elaboration can be that the peer-feedback provider offers reasons and further explanations for remarks, critique or suggestions (Strijbos et al., 2010). The degree of elaboration is important because providing elaborated peer feedback was found to be learning-enhancing, whereas providing shallow peer feedback has been shown to have a detrimental effect on the providers' learning (Y. H. Cho & Cho, 2011; Gielen et al., 2010).

6.2 Research Questions and Hypotheses

This study investigates the effects of video-based erroneous and correct cognitive modeling examples and provision of peer feedback on the development of diagnostic competences consisting of conceptual, strategic and conditional knowledge and error detection skills.

To examine whether learning with cognitive modeling examples is superior to learning without cognitive modeling examples *(RQ1)*, all groups learning with cognitive modeling examples were compared to a group that learned by reading a textbook chapter about dyspnea. The hypothesis is that video-based cognitive modeling examples are superior to learning without such cognitive modeling examples for the acquisition of diagnostic competences *(hypothesis 1).*

To examine the beneficial characteristics of cognitive modeling examples and the provision of peer feedback, research question two focuses on what the differences in effects are of watching correct versus erroneous examples, providing feedback to the video example (or not) and the possible interaction between these variations on the acquisition of diagnostic competences *(RQ2).*

It was hypothesized that erroneous cognitive modeling examples will lead to a higher learning outcome than correct cognitive modeling examples when acquiring diagnostic competences *(hypothesis 2a).* When students provide feedback on the performances by the peers in the examples, they should learn more from it than by just observing them *(hypothesis 2b).* Furthermore, I hypothesized an interaction effect: Provision of feedback might further enhance the effectiveness of erroneous cognitive modeling examples for the acquisition of diagnostic competences *(hypothesis 2c).*

To explore possible explanations for the effects of the two factors on the acquisition of diagnostic competences, two potential moderating processes and their relation to learning outcomes in two exploratory research questions were analyzed: If mental effort is related to providing peer feedback, type of cognitive modeling examples, and the acquisition of diagnostic competences (RQ3) and if content and degree of elaboration of provided peer feedback statements influence the acquisition of diagnostic competences (only for the conditions in which students were asked to provide feedback)

6.3 Method

6.3.1 Participants and design

One hundred twenty five advanced medical students in the seventh semester (70 females) participated in the study. Participation was voluntary, with informed consent and approved by the ethics committee of the medical faculty. Four students were removed due to language problems and incomplete data, leaving 121 valid cases. The mean age of the remaining 121 participants was 24.38 (*SD* = 2.80). They were randomly assigned to four conditions of a 2x2 factorial design (Table 1) and an additional control condition of 24 students.

Table 1

Design of the study

		Providing Peer-Feedback	
		yes	no
Type of modeling example	erroneous	23	25
	correct	24	25

The first factor refers to the comparison between erroneous and correct modeling examples. The second factor refers to whether the students only observed the videos or whether they were asked to provide peer feedback to the student they observed after watching each video. Therefore, five groups are compared: providing peer feedback on correct examples (CMF+), observing erroneous modeling examples without providing peer feedback (CMF-), providing peer feedback on erroneous examples (EMF+), observing erroneous examples without providing peer feedback (EMF-), and a control group (CG).

6.3.2 Learning environment

The participants watched three videos of students diagnosing dyspnea on the computer. The three (erroneous or correct) cognitive modeling examples were embedded into the CASUS online learning environment, which has been effectively used for case-based learning and video-based examples (M. Fischer, 2000). CASUS includes the possibility to administer surveys and assignments such as peer feedback provision and this feature was used in the present study. CASUS also allows for experimenter control over the pacing of the materials and when participants can proceed to the next case.

6.3.3 Materials

The three cognitive modeling examples showed a fictional peer medical student performing differential diagnoses, one senior physician introducing the patient with symptoms of dyspnea, and a standardized patient in a hospital bed describing his/her symptoms. Next, the fictional peer medical student compiled a patient history and performed a differential diagnosis in front of the senior physician.

The video-based cognitive modeling examples for the experimental groups were generated from real-life cases of the most common misdiagnosed illnesses causing dyspnea: pulmonary embolism, pneumonia, COPD, cardiac failure and myocardial infarct. Two clinicians selected ten common cases for which scripts were written. The ten scripts were cross-reviewed by the authors and five selected for filming. A pilot was conducted among twelve advanced medical students with these five video-based cognitive modeling examples (length ranged 5.43 to 8.37 minutes) who rated the videos on a scale of 0 (= not) to 5 (= very high) for each of the following criteria: difficulty, authenticity, distractors and knowledge contained. The three videos with the highest score were selected for the study. Next, two versions were created of each of the three examples: one, where the fictional peer medical student arrived at an erroneous diagnosis and one with a correct diagnosis.

A 15-page textbook chapter about differential diagnoses of dyspnea was provided to the control group (Füeßl & Middeke, 2010). In all conditions, a paper guidance sheet on diagnosing dyspnea was provided (see Appendix G).

6.4 Independent Variables

6.4.1 Erroneous cognitive modeling examples

In the erroneous cognitive modeling examples condition, the fictional peer medical student considered several possible diagnoses such as pulmonary embolism, pneumonia, COPD, cardiac insufficiency, and myocardial infarction, but set up an erroneous working hypothesis by weighting the wrong symptoms and not following up on possible other explanations. In the correct cognitive modeling examples condition the fictional peer medical student compiled the same patient history, but set up the correct working hypothesis by also including other possible explanations for the symptoms rather than the most prevalent one.

6.4.2 Providing peer feedback

In the peer feedback provision condition the participants were asked to provide feedback on the fictional medical peer's performance in the videos. Students who provided peer feedback received a note pad with two prompts asking them to write down at least two sentences on (a) "*what was good, what was bad, and where was the error*" and (b) "*what could be improved*" (see Appendix A). The prompts were based on Hattie and Timperley's (2007) criteria for learning-enhancing feedback. After the participants wrote their feedback in the text boxes, the CASUS system allowed them to proceed to the next video. In the condition where participants were not asked to provide peer feedback, they just observed the cognitive modeling examples but were allowed to take notes while watching the cognitive modeling examples.

6.5 Dependent Variables

6.5.1 Diagnostic competences – diagnostic knowledge

Diagnostic knowledge was assessed via 15 Multiple-choice items for conceptual knowledge (1 out of 5 answering options was correct; $\alpha = .51$), 15 open key-feature items for strategic knowledge (α = .58), and six problem-solving tasks for conditional knowledge (α = .59) (see Appendix B). The test format assesses a holistic concept of diagnostic competences with largely varying knowledge domains connected to dyspnea, which is reflected in a high variance of difficulty to discriminate high and low performing students, and which results in a corresponding moderate Cronbach's Alpha (Heitzmann et al., 2015).

Key-feature and problem-solving items were rated by one expert clinician from incorrect (0), mostly incorrect (1), partially correct (2), mostly correct (3) and fully correct (4) (M = 3.0, SD = 1.0). Interrater reliability was checked with Intra-Class Correlation for 30% of all items by a second expert clinician (ICC = .90). The three knowledge types were first z-transformed and then transformed to a percentage value ranging from 0 to 100. The three knowledge types were weighted equally (1/3) as part of a composite score for diagnostic knowledge.

6.5.2 Diagnostic competences – error detection skills

Error detection skills were assessed with two erroneous worked examples with 15 detectable reasoning errors (α = .61) in 28 possible locations where an error could occur. The two examples were created by an expert clinician and validated by another expert clinician (see Appendix B).

6.5.3 Mental effort

Mental effort was assessed via a self-report scale based on Paas and Kalyuga (2005) (See Appendix C). The scale consisted of eight items to assess difficulty during learning. Each item was rated on a 7 point Likert scale ranging from *very easy* (1) to *very difficult* (7). One example question is: '*How easy or hard was it for you to understand the content of the videos you watched?'* The scale was used to depict an overall mental effort without differentiating between types of effort. Cronbach's alpha was sufficient (α = .83).

6.5.4 Elaboration of provided peer feedback

The degree of elaboration of provided peer feedback was assessed with a coding scheme based on Strijbos, Van Goozen, and Prins (2012). Written peer feedback was segmented into propositional units. A unit was identified when a statement could be meaningful for itself. The units were categorized in three mutually exclusive dimensions: corrective feedback, elaborated feedback and off-topic. Corrective feedback units referred to propositional units that were based on proper medical knowledge without giving reasons or further explanations. When the propositional unit included reasons and further explanations, that unit was coded as an elaborated feedback unit. Off-topic units did not relate to medical content. A higher degree of elaboration of the overall feedback means that more propositional units were elaborated and included a rationale or reason. The number of elaborated propositional units divided by the total number of propositional units in the entire feedback served as measurement for the degree of elaboration. For the detailed coding scheme, see Appendix D. The intra-class-correlation of two independent coders who were blind to condition was .90.

6.6 Control Variables

6.6.1 Prior knowledge

Prior knowledge was assessed with six multiple-choice questions (1 out of 5 answering options was correct) and six open key-feature questions high in difficulty about the main differential diagnoses of dyspnea (Cronbach's α = .57). Due to the length of the intervention, the knowledge pre-test was not differentiated into the three different knowledge types. Multiple choice questions had a format of pick-N with six possible choices (see Appendix B). The multiple choice questions were created by two expert clinicians and evaluated by two other experts. The open key-feature questions were evaluated and independently coded by two independent expert clinicians who were blind to condition yielding an intra-class-correlation coefficient of .90.

6.6.2 Learning time

Learning time was automatically assessed via the CASUS learning platform. It is the duration for which the participants worked on the video-based cognitive modeling examples in the four experimental groups or read the textbook chapter in the control group.

6.7 Procedure

Participants entered the computer room and were introduced in terms of confidentiality, terms of participation and procedure (5 min.). Then they received the pretest (20 min.), worked on the three cognitive modeling examples and in some research conditions provided peer feedback (60 min.), answered the posttest (35 min.). Overall duration ranged between 90 and 140 minutes. Participants were group tested in eleven sessions with ten to twelve students per session. Sessions contained an equal mix of conditions with two to three students each.

6.8 Statistical Analyses

An independent t-test with all parts of diagnostic competences as dependent variables was used to compare the cognitive modeling examples conditions and the textbook condition (*RQ1*). A multivariate ANCOVA with diagnostic knowledge and error detection as dependent variables and prior knowledge and learning time as covariates was conducted to compare the four experimental conditions that worked with cognitive modelling examples (*RQ2*). The assumptions for a multivariate ANCOVA were met. The relationship of degree of elaboration of provided peer feedback *(RQ3)* and mental effort *(RQ4)* with gained diagnostic competences was examined by Pearson correlations. The effect of providing peer feedback and erroneous cognitive modeling examples on mental effort was examined by univariate ANOVA (*RQ4*).

6.9 Results

6.9.1 RQ 1 Acquisition of diagnostic competences with cognitive modeling examples

Prior knowledge between all five groups did not differ significantly (see Table 2 for descriptives). An independent t-test showed a significant difference in conceptual knowledge acquisition between all cognitive modeling examples groups combined (*M* = 76.49, *SD* = 12.00) and the text book control group (*M* = 70.08, *SD* = 10.10), *t(40.4)* = 2.651 *p* = .011, *d* = .57. No significant differences were found for the other knowledge types.

Table 2

Mean and standard deviation for prior knowledge, diagnostic knowledge, knowledge types and error detection skills for all five conditions in percentage (0-100)

	EMF-	*CMF-*	*EMF+*	*CMF+*	*CG*
	M (SD)	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*
Prior Knowledge	80.20 (13.72)	83.30 (12.35)	78.51 (17.12)	80.80 (15.31)	80.49 (11.25)
Error Detection	49.22 (12.94)	47.00 (14.56)	44.89 (14.77)	35.68 (20.06)	42.45 (20.52)
Diagnostic Knowledge	73.26 (8.89)	75.83 (9.31)	67.75 (9.81)	69.37 (10.94)	70.91 (9.06)
Conceptual	77.41 (13.16)	81.30 (9.28)	73.25 (12.24)	73.25 (12.90)	70.08 (10.71)
Strategic	80.27 (10.89)	81.73 (8.41)	71.74 (10.62)	76.87 (11.91)	77.36 (11.56)
Conditional	62.17 (14.53)	64.17 (19.36)	58.15 (19.52)	57.99 (22.89)	61.11 (18.78)

Note. Abbreviations: EMF-: Erroneous modeling examples without feedback-provision; CMF-: Correct modeling examples without feedback-provision; EMF+: Erroneous Modeling Examples with Feedback-Provision; CMF+: Correct modeling examples with feedback-provision; CG: Control group

6.9.2 RQ 2 Effect of providing peer feedback and type of cognitive modeling examples on the acquisition of diagnostic competences

A MANCOVA showed a significant main effect of providing peer feedback (Wilks $\lambda = .856$, $p = .000$) (see Table 2 for descriptives) on overall diagnostic competences. Yet, there was no main effect for type of cognitive modeling examples (Wilks $\lambda = .933$, $p = .190$). Learning time as covariate was non-significant (Wilks $\lambda = .940$, $p = .254$). There was no multivariate interaction between provision of peer feedback and type of cognitive modeling examples (Wilks $\lambda = 0.965$, $p = .535$). Hence, in the remainder the univariate effects are reported.

Effects of providing peer feedback

Students who only watched the video without providing peer feedback acquired higher levels of conceptual knowledge, $F(1,95) = 6.546$, $p = .012$, $\eta_p^2 = .07$, and higher levels of strategic knowledge, $F(1,95) = 8.827$, $p = .004$, $\eta_p^2 = .09$, than students who were asked to provide peer feedback. For conditional knowledge there was no difference, $F(1,95) = 1.377$, $p = .244$. Providing peer feedback showed a negative effect on error detection skills, $F(1,95) = 5.066$, $p = .027$, $\eta_p^2 = .01$.

Effects of errors in cognitive modeling examples

The main effect for erroneous versus correct cognitive modeling examples was not significant for all three types of diagnostic knowledge (conceptual knowledge: $F(1,95) = .060$, $p = .807$; strategic knowledge: $F(1,95) = 1.236$, $p = .269$; conditional knowledge: $F(1,95) = .000$, $p = .995$). However, students watching erroneous modeling examples acquired higher levels of error detection skills, $F(1,95) = 4.290$, $p = .041$, $\eta_p^2 = .05$.

Interaction of providing peer feedback and erroneous modeling examples

There was no significant univariate interaction between peer feedback provision and type of cognitive modeling examples for conceptual knowledge, $F(1,95) = .663$, $p = .418$, strategic knowledge, $F(1,95) = .809$, $p = .371$, conditional knowledge, $F(1,95) = .064$, $p = .809$, and error detection skills, $F(1,95) = 1.248$, $p = .267$.

6.9.3 RQ 3 Effect of providing peer feedback and type of cognitive modeling examples on mental effort

A univariate two-way ANOVA revealed no interaction of peer feedback provision and cognitive modeling examples type on mental effort, $F(1,95) = .485$, $p =.48$. Yet, a main effect was observed for type of cognitive modeling examples, $F(1,95) = 6.636$, $p = .012$, $\eta_p^2 = .07$. Students learning with erroneous cognitive modeling examples ($M = 3.91$, $SD = 0.72$) reported more mental effort than students learning with correct cognitive modeling examples ($M = 3.37$, $SD = 0.68$). Likewise, a main effect was found for providing peer feedback, $F(1,95) = 10.141$, $p = .002$, $\eta_p^2 = .10$. Students providing peer feedback ($M = 3.79$, $SD = 0.75$) reported more mental effort than students just observing ($M = 3.50$, $SD = 0.64$). The control group reported mental effort with $M = 3.07$, $SD = 0.99$.

6.9.4 RQ 4 Relationship between quality of provided peer feedback and acquisition of diagnostic competences

The processes underlying peer feedback provision were further examined. Combined across both feedback provision conditions, the feedback was 79.3% on-topic and 20.7% off-topic. The on-topic feedback consisted of 86% *corrective feedback* without explanation or deeper elaboration and 14% was coded as *elaborated feedback* with reasons and explanations provided for the statement.

For the overall amount of propositional units of provided peer feedback ($M = 20.5$, $SD = 5.80$) the degree (percentage) of elaboration ($M = 10.77$, $SD = 11.51$) and off-topic percentage ($M = 20.35$, $SD = 21.32$) of these propositional units of peer feedback were calculated. Post-test diagnostic knowledge was not correlated with overall amount of provided peer feedback ($r = .18$, $p = .24$) and degree of elaboration of feedback ($r = .11$, $p = .475$).

There was a moderate negative correlation between off-topic statements and diagnostic knowledge, *r* = -.30, *p* = .041. Prior knowledge was uncorrelated with overall amount of feedback (*r* = .11, *p* = .46), off-topic statements (*r* = .03, *p* = .84) and degree of elaboration of feedback (*r* = .02, *p* = .91). In the group that worked on erroneous cognitive modeling examples, the amount of overall provided peer feedback revealed a medium positive correlation with posttest diagnostic knowledge, *r* = .43, *p* = .049, whereas this was not the case in the group with correct cognitive modeling examples (*r* = -.04, *p* = .86). Mental effort (*M* = 3.79, *SD* = .75) was correlated with degree of elaboration of provided peer feedback (*r* = .33, *p* = .027) but not with diagnostic knowledge (*r* = -.06, *p* = .55), overall amount of provided peer feedback (*r* = -.01, *p* = .98) and off-topic statements (*r* = -.23, *p* = .14).

6.10 Conclusion and Discussion

My aim was to examine the educational value of peer feedback provision on erroneous cognitive modeling examples to teach diagnostic competences. The first research question was if cognitive modeling examples are superior to textbook learning for acquiring diagnostic competences. The second research question aimed at the added value of providing peer feedback on cognitive modeling examples and incorporating errors in the cognitive modeling examples with a possible interaction of the factors. Research questions three and four were exploratory and closely examined the relationship of degree of elaboration of provided peer feedback and mental effort with learning gain.

Results show for research question one that cognitive modeling examples are a valid educational method with higher gains in conceptual knowledge for medical students learning with cognitive modeling examples. The first hypothesis of this study that cognitive modeling examples are superior for the acquisition of diagnostic competencies to learning without cognitive modeling examples could be partially accepted for the conceptual parts of diagnostic knowledge.

Research question two focused on the effects of erroneous versus correct cognitive modeling examples and provision of peer feedback as an intervention to enhance learning. Concerning hypothesis 2a about beneficial effects of erroneous examples on learning as stated by for example Renkl and Atkinson (2003), results were mixed for diagnostic knowledge acquisition and error detection skills: There was no positive influence of learning with erroneous cognitive modeling examples on diagnostic knowledge acquisition. An explanation might be that learners were already too advanced so that example based learning methods – with or without errors – made no difference. Stark et al. (2011) found that comparing low and high-performing students can reveal an influential interaction effect of prior knowledge on learning with erroneous or correct examples. Due to homogeneity of prior knowledge, a comparison of low and high-performing students when learning with erroneous versus correct cognitive modeling examples was not possible in this first empirical study. Contrary to no significant effect on acquiring diagnostic knowledge, learning with erroneous cognitive modeling examples had an effect on error detection skills. Erroneous cognitive modeling examples seem to foster distinct skills of error detection. Gartmeier et al. (2010) explain this with the construction and integration of negative knowledge to existing knowledge when someone is exposed to an error. The construction of negative knowledge might be a key component of building medical expertise to prevent diagnostic errors (Dror, 2011), which makes erroneous cognitive modeling examples a needed addition in medical education. Using erroneous examples in a long-term intervention with measurements to examine an effect of exposure to errors on long-term error detection skills might shed light on the question if the found significant effect of erroneous cognitive modeling examples on error detection skills was just a short time heightening of sensitivity for errors or if it persists longer. A persisting effect supports the idea that erroneous cognitive modeling examples help building negative knowledge. In sum, hypothesis 2a could partly be accepted for error detection skills but not for diagnostic knowledge acquisition.

Hypothesis 2b was that provision of peer feedback on cognitive modeling examples might lead to increased learning gain. This hypothesis is in line with studies by Li et al. (2010) and Rouhi and Azizian (2013) about beneficial effects of peer feedback provision. The results of this study indicate that students providing peer feedback performed worse on two types of diagnostic knowledge and error detection skills compared to students who just observed videos. Hypothesis 2b can be rejected. Also, an interaction between providing peer feedback and erroneous examples could not be found as assumed in hypothesis 2c.

To explore the results further, research question three and four aimed at possible explanations for the negative effect of providing peer feedback with analyzing mental effort and degree of elaboration of provided peer feedback. The feedback provided was mostly corrective feedback and not elaborated. Detrimental effects on learning when providing shallow peer feedback are in line with findings by Y. H. Cho and Cho (2011). Degree of elaboration of provided peer feedback was related to mental effort, which might have interfered with learning effort. Unfortunately, different types of mental effort could not be differentiated. A positive relationship between providing peer feedback and knowledge acquisition was found in the group of students learning with erroneous cognitive modeling examples. This supports findings from Braaksma et al. (2002) and Y. H. Cho and Cho (2011) that explaining weak models helps students’ own learning by becoming a critical learner through self-monitoring and self-regulation.

Even though no interaction of peer feedback provision on erroneous cognitive modeling examples was found, the amount of provided peer feedback on erroneous cognitive modeling examples was related to learning gain of diagnostic knowledge. This shows that under certain circumstances, providing peer feedback increases learning gain.

Overall, the results do not offer an explanation for the decreased learning gain when students provided peer feedback. Possibly, students might not be used to providing peer feedback on a peer’s diagnoses as depicted by the lack of elaboration. The novelty of the situation might have led to cognitive overload, which decreased their focus of attention and therefore their learning outcome. Further, the task to provide written peer feedback on a video-taped student might have not been authentically simulating real-life feedback situations. Hoogerheide, Deijkers, Loyens, Heijltjes, and van Gog (2016) suggest that incorporating speaking instead of writing explanation tasks for learning with videos leads to social presence and authenticity, which leads to more learning gain.

If providing peer feedback is incorporated in medical education practices, the educator should make sure that students are not distracted by the peer feedback task and have sufficient knowledge about effective (peer) feedback provision. Providing them just-in-time guidance on how to provide peer feedback during observation might have increased the mental effort to solve the task. Training effective peer feedback provision might counteract this, decrease novelty and reduce mental effort by making the task to provide peer feedback more natural with less effort (Prins et al., 2006; Sluijsmans, Brand-Gruwel, & van Merriënboer, 2002). Similarly, receiving elaborated expert feedback as proposed by Heitzmann et al. (2015); Stark et al. (2011) on the cognitive modeling examples might reduce mental effort and increase learning gain.

Further research is needed to entangle the beneficial and detrimental processes of providing peer feedback on observed performance for own learning gains. For example, the effects of feedback training on elaboration and mental effort could be examined for further advancing the knowledge about how to make providing peer feedback a learning opportunity. So far, under certain circumstances of prior knowledge, elaboration of peer feedback and characteristics of the observed performance, peer feedback provision might be a useful tool to foster learning. It is needed to find out more about the factors leading to positive and negative effects of peer feedback provision and consultation as it is part of everyday clinical practice. If used right, peer feedback could greatly contribute to clinical practice and therefore patient safety.

7 Study 2: The Voice is Mightier than the Pen

How Communication Medium and Expert Feedback Influence Peer Feedback Provision and Learning

Context: Teaching diagnostic competences is the key to diagnostic accuracy. Diagnostic competences rely on at least two individual-level capacities which are diagnostic knowledge and error detection skills. To teach these competences, observing erroneous cognitive modeling examples of peers performing a cognitive task like differential diagnosing has been shown to be effective. However, students' engagement in the learning task needs to be fostered with instructional methods such as prompting self-explanations or explanations to others while providing peer feedback on the observed performance. How this task needs to be designed is open for discussion. There is evidence that medium of communication plays a crucial role. Further, receiving expert feedback might support the student with providing feedback, leading to better acquisition of diagnostic competences.

Objectives: Due to negative effects of providing peer feedback as an engaging learning intervention in the first study, this study investigated (a) the influence of providing spoken or written peer feedback on observed examples on the providers' elaboration and acquisition of diagnostic competences, and (b) the influence of expert feedback on the providers' elaboration of the cognitive modeling examples and acquisition of diagnostic competences.

Methods and Sample: In the context of differential diagnoses, students were asked to provide feedback on erroneous cognitive modeling examples of a peer (fellow medical student) diagnosing a patient with acute dyspnea. After observing the modeling examples, half of the students either provided written or spoken peer feedback. Additionally, they received elaborated expert feedback on the observed modeling example or did not. A control group did not provide or receive any feedback on the modeling examples. The sample consisted of advanced medical students (N = 134) randomly assigned across the four experimental conditions and a control condition that did not provide or receive any feedback.

Results: Receiving expert feedback did not significantly influence the degree of elaboration or acquisition of diagnostic competencies when learning with erroneous cognitive modeling examples with Wilks $\lambda = .335$, $p = .854$. An indirect effect was found for spoken over written provision of peer feedback which was explained by the mediation of the degree of elaboration of provided peer feedback.

7.1 Theoretical Framework

7.1.1 Diagnostic competences – diagnostic knowledge and error detection skills

Successful clinical reasoning refers to a decision-making process that relies on the ability to reason correctly, based on usable knowledge monitored by meta-cognition (Higgs et al., 2008). Usable knowledge is stored in illness scripts: complex mental representations of diseases based on experienced past cases by the physician (Schmidt et al., 1990). When encountering a new patient case, the physician compares the new case to his or her acquired illness scripts. The most likely explanatory illness script is selected and a working hypothesis of the disease causing the symptoms is set (Boshuizen & Schmidt, 2008). The diagnostic knowledge incorporated in illness scripts can be classified as conceptual knowledge about facts, strategic knowledge about procedures and strategies to secure a diagnosis, and conditional knowledge about the rationale and goal of these procedures (Stark et al., 2011; Van Gog et al., 2004). The decision process by which observed symptoms are matched to one or more illness scripts is crucial for diagnostic success and seems to be both conscious and (automated) unconscious (Croskerry, 2009). Selecting the correct illness script is a crucial moment in diagnosing, which is prone to errors with error-rates between 5 to 15% caused by a flawed clinical reasoning process (Berner & Graber, 2008; Graber et al., 2005; Singh et al., 2014). To select the correct illness script, rich and complex knowledge is a necessity but not sufficient, since errors are not only based on knowledge gaps but also on shortcomings during the diagnostic process (Elstein, 2009). Also, meta-cognition that focusses on monitoring the clinical reasoning process might play a crucial role for diagnostic accuracy (Mamede et al., 2012). Meta-cognition consists of meta-cognitive knowledge and skills. Meta-cognitive knowledge is referred to as meta-cognitive declarative knowledge, whereas meta-cognitive skills are referred to as meta-cognitive procedural knowledge (Veenman, Van Hout-Wolters, & Afflerbach, 2006). Meta-cognition is driven by meta-cognitive knowledge, i.e. when to use meta-cognitive monitoring processes (Flavell, 1979; Pintrich, 2002). The use of meta-cognitive skills has been hypothesized to be connected to diagnostic accuracy (Croskerry, 2003; Sherbino et al., 2011). For example, a study by Sherbino et al. (2014) assumed that teaching cognitive forcing strategies was effective against errors. Their skill training included teaching medical students self-monitoring, knowledge and identification strategies to counter biases and promote deliberate reflection during diagnostic reasoning. Results showed that cognitive forcing strategies were not related to a reduction of diagnostic errors.

Meta-cognitive knowledge might play a key role in the diagnostic process. The knowledge about when and how to use meta-cognitive skills is important for diagnostic success, i.e. it is important to know when to be cautious because an error is likely to happen or has already happened (Dror, 2011). Studies on medical expertise reveal that experts are more accurate in diagnosing (Schmidt & Boshuizen, 1992). This is not due to experts making less errors but about their ability to detect and correct errors (Patel et al., 2011). Medical experts show a higher ability to detect and correct their errors in diagnoses than novices (Nyssen & Blavier, 2006; Wilkinson et al., 2011). In fact, Oser and Spychiger (2005) and Gartmeier et al. (2010) consider the acquisition of negative knowledge essential to detect and correct errors. Assessing the outcome of learning interventions in terms of acquired diagnostic knowledge and error detection skills as parts of diagnostic competences is a promising approach to reduce errors. How to design interventions to increase diagnostic competences is examined in this study.

7.1.2 Fostering diagnostic competences with erroneous cognitive modeling examples

To enable the construction of rich illness scripts, medical students need to acquire complex knowledge by exposure to clinical cases (Schmidt & Rikers, 2007). When expert physicians need to diagnose a new patient case, their clinical reasoning is based on past cases that they have been exposed to, which are compared to the new case (Aamodt & Plaza, 1994). Therefore, educational methods should focus on *case-based learning*. Cases can be presented effectively in written form as worked examples (Stark et al., 2011; Sweller, 2006) or video-based as *cognitive modeling examples*. Cognitive modeling examples are examples of a (peer) agent, performing a certain task while explaining the rationale behind the approach to solve the task (Hoogerheide et al., 2014). They offer advantages over written worked examples in terms of focus of attention and authenticity, leading to better learning (Große & Renkl, 2007; Valmont, 1995; Van Gog & Rummel, 2010). The agent can be a peer or an expert (Van Gog & Rummel, 2010). Peer models should be used when experts' and peers' way of performing varies significantly in automatization, such as in clinical reasoning (Boshuizen & Schmidt, 2008; Hoogerheide, van Wermeskerken, et al., 2016). Perceived similarity of competence between the observed peer model and the observer positively influences learning as well (Braaksma et al., 2002). Students observing a weak but coping peer model with erroneous performance show better reflection and elaboration of the learning content, higher learning gain and a decreased likelihood of making (or repeating) the observed error later-on if the error is recognized (Blandin & Proteau, 2000; Domuracki, Wong, Olivieri, & Grierson, 2015; Gartmeier et al., 2008; Große & Renkl, 2007; Siegler, 2002). To activate the students to observe the examples and engage in elaboration and recognition of the errors in the examples, (self-)explanation prompts were found to be useful (Renkl, 2002; Renkl & Atkinson, 2003).

7.1.3 Elaboration of cognitive modeling examples with peer feedback provision

Fiorella and Mayer (2013) found that students who explain learning material to others gain a deeper understanding compared to students who self-explained. Explaining the learning material to peers by providing peer feedback on the observed example was found to enhance students' reflection, stimulate metacognitive processes such as the detection of errors, elaboration and increase learning gain (K. Cho & MacArthur, 2011; Y. H. Cho & Cho, 2011; Topping, 2005). However, providing peer feedback does not automatically foster learning. The quality of provided peer feedback has a direct positive effect on learning gain (Y. H. Cho & Cho, 2011; Li et al., 2010). For example, in the study by Li et al. (2010) providing elaborated high-quality peer feedback had a positive effect on learning, while providing shallow corrective peer feedback had a negative effect. The presented Study 1 in Chapter 6 by Strobel, Heitzmann, Strijbos, Kollar, and Fischer (2016) found that when medical students provide peer feedback on observed diagnoses in a computer based learning environment, they may even learn less. A proposed explanation was that the task to provide peer feedback imposed too much mental effort on the students. Furthermore, providing shallow low-quality feedback without elaboration was found to have detrimental effects on their learning in other studies as well (Y. H. Cho & Cho, 2011; Li et al., 2010).

In sum, providing peer feedback can be a useful learning tool if designed in a way that triggers enable students to provide *elaborated* feedback. The literature suggests different approaches to how such feedback activities can be designed, which are described in the following sections.

7.1.4 Differences in providing written versus spoken peer feedback

Wegerif (1998) points out that social relatedness is a key point in computer-supported communication. Results from a study by Hebert and Vorauer (2003) were that face-to-face communication increases provision of content relevant peer feedback and increases motivation to give qualitatively higher feedback as compared to written and computer-based settings. Krych-Appelbaum and Musial (2007) found that students explain more when they provide spoken face-to-face peer feedback than when they write in a computer-mediated setting.

Even though computer-based learning environments offer strong advantages in terms of availability and productivity of learning material (Blomberg et al., 2013), they certainly also have disadvantages: Compared to face-to-face communication, social relatedness and interactivity falls short by missing social presence in computer-supported learning needs to be enhanced or simulated (Cyr, Hassanein, Head, & Ivanov, 2007; Litosseliti, Marttunen, Laurinen, & Salminen, 2005; Rice, 1993). To stimulate the beneficial effects of face-to-face communication on learning, oral communication and explanation tasks to fictitious others can be integrated in computer-supported learning environments to simulate social presence. Hoogerheide, Deijkers, et al. (2016) found that students learned complex content better when they explained it orally to a fictitious other than when they wrote it. An explanation on basis of these results stems from Fiorella and Mayer (2013, 2014) that explaining it orally simulates social presence of another person. This effect could also be valid for providing oral peer feedback on cognitive modeling examples. Nevertheless, providing oral peer feedback might yield positive effects on the degree of elaboration and it must be ensured that students recognize the mistakes in the observed performance and know how to provide elaborated feedback. Incorporating elaborated ‘perfect’ expert feedback on observed erroneous performances might be a way to do so (Stark et al., 2011).

7.1.5 Incorporating expert feedback on cognitive modeling examples for learning

Example-based learning with errors can be an effective tool – given the student detects the error and engages in elaboration (Renkl & Atkinson, 2003). Students need to have additional support when the learning material becomes complex (Stark, Gruber, Mandl, & Hinkofer, 2001) but they seem to struggle sometimes when learning with erroneous examples and prompts for activities like explanation. Berthold and Renkl (2009) found that self-explanation prompts fostered both correct and incorrect elaboration. To help students understand the content matter in the example and enable correct elaboration, instructional explanations can be given. However, interaction effects of active self-explanation and (teacher-based) expert instructional explanation are unclear (Berthold & Renkl, 2009; Heitzmann et al., 2015). For written worked examples, the integration of instructional explanation of expert feedback on the example has been found to have a positive effect on elaboration. For example, Stark et al. (2011) used elaborated expert feedback on erroneous worked examples to help students acquire diagnostic conceptual, strategic and conditional knowledge when learning with these examples. In their studies, a fictitious expert clinician provided elaborated feedback after each diagnostic step of the example about correct parts, erroneous parts, and what parts of the diagnosis could be improved. The content and structure of the feedback is important as it influences learning gain. A high degree of elaboration of feedback was found to lead to better learning gain compared to shallow corrective feedback (Gielen et al., 2010; Hattie & Timperley, 2007; Stark et al., 2011). In general, however, results on the impact of quality of received (expert) feedback on learning outcome are mixed (Kollar & Fischer, 2010).

In summary, the findings of Stark et al. (2011) indicate that incorporating expert feedback on the observed performance to enhance elaboration, quality of own provided peer feedback and therefore learning might be transferred on to cognitive modeling examples. How this affects the quality of elaboration of own provided peer feedback is unclear and to my knowledge not yet examined.

7.2 Research Questions and Hypotheses

This study investigated the effects of different types of medium of communication (written versus spoken) through which peer feedback is provided and incorporating expert feedback on erroneous cognitive modeling examples on the acquisition of diagnostic competence.

The first research question investigates the influence of the communication medium of provided peer feedback and received expert feedback on degree of elaboration of peer feedback (*RQ1*). It was hypothesized that students providing spoken peer feedback on erroneous cognitive modeling examples have a higher degree of elaboration of peer feedback than students providing written peer feedback (*hypothesis 1a*). Furthermore, students that were presented with expert feedback after providing peer feedback on an erroneous cognitive modeling example are expected to elaborate more on the observed than the students who did not have expert feedback (*hypothesis 1b*).

The second research question examines the influence of the communication medium of peer feedback provision and the reception of expert feedback on erroneous cognitive modeling examples on the acquisition of diagnostic competences (*RQ2*). It was hypothesized that students acquire more diagnostic competence when they provide spoken peer feedback on the observed cognitive modeling examples (*hypothesis 2a*). Moreover, students presented with expert feedback after providing peer feedback on erroneous cognitive modeling examples are expected to have higher learning outcomes for diagnostic competence (*hypothesis 2b*).

A third research question aims to inspect the relationship between degree of elaboration of provided peer feedback and learning gain (*RQ3*). I assumed that the degree of elaboration of provided peer feedback is related to learning gain (*hypothesis 3a*). This relationship might mediate the effects of the medium of communication of peer feedback provision, reception of expert feedback on diagnostic competence acquisition via the degree of elaboration of provided peer feedback (*hypothesis 3b*).

7.3 Method

7.3.1 Participants and design

The conduction of this study was approved by the ethics committee of the LMU Medical Center in 2014. In all, 134 medical students participated voluntarily in the study. Three students were excluded as they did not have any prior diagnostic knowledge. Overall, 131 valid cases could be used for this study of which 74 were female and 57 male, with a mean age of $M_{age} = 24.19$ ($SD = 3.03$). The mean study duration (in semesters studied) was 7.48 ($SD = 1.55$), with a minimum of five and a maximum of eleven semesters. They were randomly assigned to one of five groups of a 2x2 factorial design with a control group. The four experimental groups (a) provided spoken peer feedback ($n = 29$) or (b) provided written peer feedback ($n = 27$) on cognitive modeling examples, and (c) provided spoken peer feedback while receiving expert feedback ($n = 24$) or (d) or provided written peer feedback while receiving expert feedback ($n = 25$). A control group ($n = 26$) did not provide peer feedback but just observed the cognitive modeling examples and received no expert feedback.

7.3.2 Learning environment

The learning environment for this study consisted of a pre-test for diagnostic knowledge, a short feedback training, three erroneous cognitive modeling examples followed by expert feedback in two of the four experimental groups and a post-test for mental effort, diagnostic knowledge and error detection skills. It was created as a complete learning unit about dyspnea in the CASUS online learning environment (M. Fischer, 2000), which is commonly used in medical education as an effective tool to create, monitor and evaluate example-based learning interventions. The learning environment allows for various assessment methods, step-by-step case presentation and prompts.

7.3.3 Materials

The erroneous cognitive modeling examples were short video clips, between 5.43 and 8.37 minutes in duration, and recorded with patient actors and physicians. The clips were based on the most common real-life cases of erroneous diagnoses of dyspnea in the emergency department of the Medical Center of the LMU Munich. Two expert clinicians created scripts for ten clips of which the five best were chosen by the non-author expert to be filmed. The five clips were used in a pilot study with twelve advanced medical students who rated them on a scale from 0 (*not*) to 5 (*very high*) for authenticity, difficulty, distractors and knowledge contained. The three best-scoring videos were used for this study. The content of the cognitive modeling examples included a fictional peer student who diagnosed dyspnea with special emphasis on leading causes of symptoms by pulmonary embolism, COPD, pneumonia, myocardial infarction and cardiac insufficiency. The error occurred during setting up a working hypothesis of the most likely diseases by wrongly weighting the symptoms and choosing the most obvious explanation.

The feedback training material consisted of a short introduction about the scientific evidence that elaborated feedback is important for learning, followed by a written worked example of high-quality feedback on erroneous diagnoses explained step-by-step (see Appendix E).

The written expert feedback was also presented in CASUS but only to two of the experimental groups after each erroneous cognitive modeling example. The expert feedback was created by two experts, and based on a correct diagnosis and the suggestions for elaborated feedback by Strijbos et al. (2010). See Appendix F for the example used.

7.4 Independent Variables

7.4.1 Written versus spoken peer feedback

The communication medium of provision of peer feedback was operationalized as written versus spoken. Overall, all experimental groups watched three erroneous cognitive modeling examples and provided peer feedback on the erroneous diagnostic reasoning of the observed (fictitious) peer student in the cognitive modeling example immediately after watching the videos. The students in the spoken peer feedback groups were presented with an instruction how to use a dictating machine to record their feedback for the fictitious peer student in the cognitive modeling example. The students in the experimental written peer feedback groups were presented with an instruction to write their peer feedback for the fictitious peer student in the cognitive modeling example in CASUS. All groups received a short reminder of the content of the feedback training.

7.4.2 Incorporating expert feedback

The expert feedback was provided after each erroneous diagnosis in the CASUS learning platform after the students provided peer feedback on the cognitive modeling examples. It follows the guideline structure of Hattie and Timperley (2007) (See Appendix A and F) used in the feedback training students received before working on the learning unit. It focusses on good aspects of the diagnosis, wrong aspects of the diagnosis and offers suggestions for improvement. Every aspect of good, bad and erroneous parts separately and elaborated on which is concluded in suggestions. See Appendix F for the example.

7.5 Dependent Variables

The dependent variable of diagnostic competences was operationalized as diagnostic knowledge and error detection skills. Both were standardized on a scale from 0 to 100 and combined with equal weight (1/2 knowledge, 1/2 error detection skills) to reflect diagnostic competences.

7.5.1 Diagnostic competences – diagnostic knowledge

Diagnostic knowledge was tested as three distinct knowledge types as proposed by Stark et al. (2011) with conceptual, strategic and conditional knowledge. Conceptual knowledge was measured with seven multiple-choice questions with one out of five answers being correct ($\alpha = .50$). The multiple-choice questions were designed to assess clinical interpretation, reasoning and justification. The answers were mutually exclusive. Twelve key-feature questions with three questions about four central problems tested for strategic knowledge ($\alpha = .67$). Six open problem-solving tasks measured conditional knowledge ($\alpha = .52$). Cronbach's alpha across the knowledge types was $\alpha = .76$. They were z-transformed into a scale from 0 to 100 to depict percentage of correct answers. The three knowledge parts were then combined with equal weight of 1/3 to measure diagnostic knowledge as a whole. The rather low Cronbach's alpha is due to the fact that several knowledge areas of dyspnea are tested across questions to depict a holistic concept of diagnostic knowledge and was found in other studies using this format as well (Heitzmann et al., 2015). Due to a high overall alpha, the alphas for the sub-scales could be accepted. For all questions, see Appendix B.

7.5.2 Diagnostic competences – error detection skills

Error detection skills were measured via two erroneous written worked examples of complex differential diagnoses of dyspnea (see Appendix B). The examples were created by two expert clinicians and validated in a pilot study. The method is based on the assessment of error detection of dialysis nurses by Wilkinson et al. (2011). The text-based examples depicted two patients with unspecific symptoms of dyspnea followed by an explained differential diagnostic rationale of a physician. In these rationales, 30 possible erroneous steps were highlighted. Out of the 30 steps, 15 were erroneous and 15 correct. The students had to recognize and remark on the erroneous steps. The examples were included in CASUS learning environment and student underlined the erroneous steps by clicking on them. The reliability for all 15 items was sufficient with α = .62.

7.5.3 Degree of elaboration of provided peer feedback

Degree of elaboration of provided peer feedback was analyzed in written form for both written and spoken peer feedback. Spoken peer feedback was first transcribed by trained student helpers. The coding of degree of elaboration of provided peer feedback was based on Strijbos et al. (2012). First, units of meaning were coded. These units of meaning were categorized in three distinct, mutually exclusive categories. The first category was *corrective feedback*, that is: statements aimed at the content but without an explanation, further elaboration or rationale, for example '*The working diagnosis of pulmonary embolism is incorrect.'* The second category refers to *elaborated feedback*, which means units of meaning which targeted the content but offered further explanation, elaboration, further thoughts or a rationale. For example '*The working diagnosis of pulmonary embolism is incorrect because Troponin is increased, which is a typical symptom for myocardial infarct'* The third category '*off-topic'* was coded if a unit of meaning did not target the content such as *'She did not wear a pony tail'.* See Appendix D for the coding scheme.

7.5.4 Mental effort

Mental effort was measured using a questionnaire with 8 items on basis of Paas and Kalyuga (2005), which were answered on a 5 point-Likert-scale (see Appendix C). The questions cover about how easy or hard the learning intervention was perceived ranging from

1(*very easy*) to 5 (*very hard*). Different types of mental effort could not be differentiated with this scale. The overall reliability was good (α = .80).

7.5.5 Prior knowledge

Prior knowledge was assessed with a pre-test that consisted of basic parts of the post-test for diagnostic knowledge. Types of knowledge were not differentiated in the pre-test, since the aim was to depict an overall picture of diagnostic knowledge without imposing too much effort on the students or using too much time of the intervention. Prior knowledge was measured with six multiple-choice items, six key-feature items and two problem-solving items. Across all 14 items Cronbach's alpha was .65.

7.6 Procedure

Students were randomly assigned to one of two equally sized computer rooms. They were shortly introduced to the study and completed a demographic questionnaire (5 min.) followed by a pre-test for diagnostic competences (20 min.). After completion, they received a short feedback training based on a written worked example (5 min.) and studied a guidance sheet on diagnosing dyspnea (5 min.) (see Appendix G). The intervention part (60 min.) consisted of watching three erroneous cognitive modeling examples and providing feedback on them. Afterwards, mental effort and diagnostic competences were assessed (45 min.). Overall, the study was designed to take 150 minutes in total and all students completed the unit within this time frame.

7.7 Statistical Analysis

A treatment check was performed by comparing the experimental groups to the control group with an independent sample t-test. A MANCOVA with prior knowledge as covariate was used to assess the effect of spoken (vs. written) and expert feedback (vs. no expert feedback) on the acquisition of diagnostic competences. For the effect of both factors on the degree of elaboration of provided peer feedback, a MANOVA analysis was performed. To determine the relationship between elaboration of provided peer feedback and diagnostic competence, spearman correlations between prior knowledge, elaboration of peer feedback, and all parts of diagnostic competences were computed. Based on these correlations, a MANCOVA was conducted for diagnostic competences with degree of elaboration of provided peer feedback. For further analysis, mediation analysis of the effects of written vs. spoken and expert feedback vs. no expert feedback on diagnostic competence – with elaboration of peer feedback as a mediator – was conducted. Hayes (2013) PROCESS macro 2.16 for mediation, moderation and conditional process analysis was used to compute the mediation analyses in SPSS.

7.8 Results

7.8.1 Comparing experimental groups with a control group

Table 3

Descriptives for feedback versus no feedback group comparison

	Feedback groups *N* = 105	Control group *N* = 26
	M (*SD*)	*M* (*SD*)
Prior Knowledge	72.20 (13.61)	70.41 (13.86)
Diagnostic Competences	57.11 (12.47)	56.05 (13.69)
Diagnostic Knowledge	68.38 (12.33)	67.49 (13.09)
Conceptual Knowledge	76.82 (14.89)	72.97 (17.52)
Strategic Knowledge	70.38 (14.07)	75.16 (14.12)
Conditional Knowledge	57.94 (17.46)	54.33 (15.07)
Error Detection Skills	45.84 (17.97)	44.62 (18.64)

Note. ‘Feedback groups’ refers to all four experimental groups that were compared to the control group which neither received expert feedback nor provided peer feedback on the erroneous cognitive modeling examples

The comparison of the four experimental groups providing peer feedback versus a control group that did not provide peer feedback on cognitive modeling examples did not reveal a significant difference for prior knowledge *t*(37.87) = -.594, *p* = .556, diagnostic competences, *t*(35.96) = -.360, *p* = .721, or any of the subcategories of diagnostic knowledge, *t*(36.77) = -.316, *p* = 754, or error detection skills, *t*(37.35) = -.302, *p* = .764, of diagnostic competences. See Table 3 for descriptives.

7.8.2 RQ1 - Effects of medium of communication and expert feedback on elaboration

To answer hypotheses 1a and 1b (effects of the varied factors on elaboration of provided peer feedback), a MANCOVA was performed which revealed the following results:

Table 4

Standardized degree of elaboration of feedback provided (percentage)

	SF	SFEx	WF	WFEx	Total
N	24	22	27	24	97
	M (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
Elaborated Feedback	48.10 (21.72)	53.93 (16.19)	39.11 (16.70)	39.76 (21.28)	44.92 (19.81)
Corrective Feedback	50.67 (19.9)	45.82 (16.50)	59.73 (16.69)	59.02 (21.16)	54.10 (19.27)
Off-topic	1.23 (5.15)	0.24 (.80)	1.12 (3.04)	1.23 (6.00)	0.98 (4.22)

Note. SF: Spoken Feedback, SFEx: Spoken Feedback plus Expert Feedback, WF: Written Feedback, WFEx: Written Feedback plus Expert Feedback

Prior knowledge was not significantly correlated with degree of corrective feedback (r_s = -.053, *p* = .611) degree of elaboration of feedback (r_s = .079, *p* = .444) or off-topic (r_s = -.07, *p* = .496) and was excluded as covariate. MANOVA analysis showed (a) a multivariate significant main effect of medium of communication of provision of peer feedback with Wilks λ = .913, *p* = .016, η_p^2 = .09 (b) no significant multivariate main effect of whether students received expert feedback or not, Wilks λ = .401, *p* = .671 and (c) no multivariate interaction main effect, Wilks λ = .322, *p* = .725. Between-subject effects showed a univariate main effect of medium of peer feedback provision for degree of corrective feedback, *F*(1,95) = 8.49, *p* =.004, η_p^2 = .08 and degree of elaborated feedback, *F*(1,95) = 8.75, *p* = .004, η_p^2 = .09. Off-topic was not influenced by the medium, *F*(1,95) = .27, *p* = .603.

7.8.3 RQ2 - Effects of medium and expert feedback on learning gain

To answer hypothesis 2a and 2b (effects of the varied factors in diagnostic competences acquisition), a conducted MANCOVA analysis revealed the following results:

Table 5

Descriptives across all experimental groups of Study 2

	Control	SF	SFEx	WF	WFEx	Total
N	26	29	24	27	25	131
	M (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
Diagnostic competences	56.02 (15.28)	55.38 (12.60)	56.78 (12.06)	60.39 (8.71)	57.11 (12.47)	56.90 (12.67)
Diagnostic knowledge	69.75 (14.68)	67.43 (13.37)	67.15 (11.00)	69.04 (10.00)	68.20 (12.44)	68.20 (12.44)
Conceptual Knowledge	72.97 (17.52)	77.64 (14.88)	76.15 (14.96)	76.37 (17.01)	77.03 (13.14)	76.06 (15.45)
Strategic Knowledge	75.16 (14.12)	70.69 (17.72)	69.36 (12.06)	69.83 (14.05)	71.58 (11.61)	71.33 (14.15)
Conditional Knowledge	54.33 (15.07)	60.92 (18.14)	56.77 (20.29)	55.25 (16.08)	58.50 (15.57)	57.22 (17.02)
Error Detection Skills	44.62 (18.64)	42.30 (21.40)	43.33 (18.44)	46.42 (16.89)	51.73 (13.09)	45.60 (18.04)

Note. SF: Spoken feedback, SFEx: Spoken feedback plus expert feedback, WF: written feedback, WFEx: Written feedback plus expert feedback, Control: Control group

The MANCOVA showed (a) a multivariate main effect of prior knowledge, Wilks λ = .508, *p* = .000, but (b) not for medium of provided peer feedback, Wilks λ = .947, *p* = .259 or (c) received expert peer feedback, Wilks λ = .335, *p* = .854. There was also no multivariate interaction main effect, Wilks λ = .289, p = .885 on any parts of diagnostic competences. No significant between-subjects effects were found.

7.8.4 RQ3 Relationship between medium of communication, expert feedback, elaboration of feedback, and acquisition of diagnostic competences

First, correlations between degree of elaboration of provided peer feedback, mental effort and diagnostic competences, as well as prior knowledge were computed.

Table 6

Correlations between prior knowledge, post-test diagnostic competences, elaboration, and mental effort

	Prior Knowledge	Diagnostic Competences	Conceptual Knowledge	Strategic Knowledge	Conditional Knowledge	Error Detection Skills	Mental Effort
Corrective Feedback	-.05	-.26 *	-.17	-.10	-.16	-.21 *	-.03
Elaborated Feedback	.08	.28 *	.17	.12	.18	.24 *	.04
Off-topic Feedback	-.08	-.01	.00	-.13	-.05	.01	.07
Mental Effort	-.03	-.11	-.07	-.01	-.13	-.13	

** = p ≤ .05*

Based on the positive correlations between degree of elaboration of provided peer feedback and diagnostic competences acquisition, a MANCOVA with degree of elaboration and prior knowledge as covariates was conducted:

Prior knowledge, Wilks $\lambda = .545$, $p = .000$, $\eta_p^2 = .47$, and degree of elaboration, Wilks $\lambda = .863$, $p = .011$, $\eta_p^2 = .14$, showed multivariate effects on the acquisition of diagnostic competences. Medium of feedback provision, Wilks $\lambda = .914$, $p = .094$, and reception of expert feedback, Wilks $\lambda = .227$, $p = .923$, were not significant. There was also no main interaction effect of medium of feedback provision and receiving expert feedback, Wilks $\lambda = .310$, $p = .871$. Univariate between-subjects effects showed a significant influence of degree of elaboration on overall diagnostic competences, $F(1,95) = 13.42$ $p = .000$, $\eta_p^2 = .13$, and error detection skills, $F(1,95) = 9.98$, $p = .002$, $\eta_p^2 = .10$, but not on conceptual, $F(1,95) = 2.83$, $p = .096$, strategic, $F(1,95) = .61$, $p = .436$, or conditional knowledge, $F(1,95) = 3.91$, $p = .051$. Although the multivariate main effect was not significant, medium of feedback provision showed significant univariate between-subject effects for overall diagnostic competences, $F(1,95) = 8.04$, $p = .006$, $\eta_p^2 = .08$ and error detection skills, $F(1,95) = 7.36$, $p = .008$, $\eta_p^2 = .08$. No significant effect was found for conceptual, $F(1,95) = .515$, $p = .475$, strategic, $F(1,95) = 1.22$, $p = .272$, or conditional, $F(1,95) = .683$, $p = .411$) knowledge parts of diagnostic competences.

A mediation analysis was conducted to determine whether degree of elaboration of peer feedback mediates the relationship between medium of peer feedback provision and diagnostic competences. An indirect-only mediation effect was found with a completely standardized indirect effect of .09, with a 95% confidence interval ranging from .02, .19.

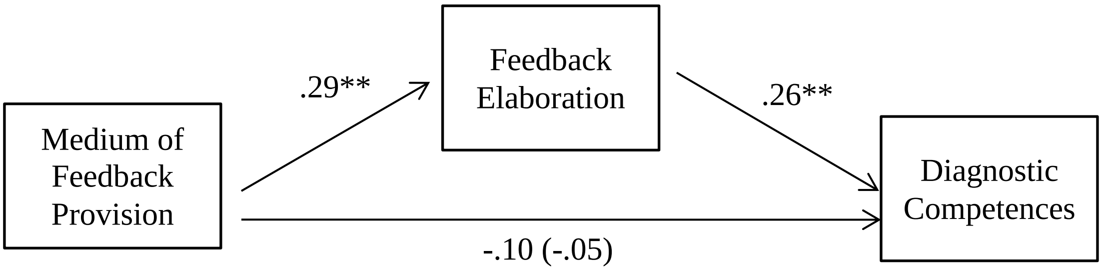

Figure 5. Model of mediation analysis between medium of communication, elaboration and diagnostic competences.

7.9 Discussion

The results depict that providing peer feedback on erroneous cognitive modeling examples does not per se enhance learning. This could be shown by the treatment check: having the task to provide peer feedback on observed performances alone does not foster learning.

The first research question investigated if there is a difference in the degree of elaboration of provided peer feedback on erroneous cognitive modeling examples if expert feedback is received during the task of providing peer feedback and if the variation of medium (spoken vs. written) influences the degree of elaboration as well. The medium of provided peer feedback influenced the degree of elaboration of peer feedback positively: the degree of elaboration of provided peer feedback was higher and corrective feedback was lower when spoken peer feedback was provided. Hypothesis 1a could be accepted. Hypothesis 1b about a positive influence of expert feedback on the degree of elaboration of provided peer feedback could be rejected, as no positive influence was found.

The second research question examined how medium of provided peer feedback and receiving expert feedback on erroneous cognitive modeling examples influence the acqusition of diagnostic competences. The results show that providing spoken or written peer feedback did not automatically lead to a higher learning outcome, which contradicts Hypothesis 2a about a positive influence of providing spoken peer feedback on erroneous cognitive modeling examples on acuqisition of diagnostic competences. Further, results show that the reception of expert feedback during learning with erroneous cognitive modeling examples did not influence the students' acqusition of diagnostic competences; hypothesis 2b could be rejected as well.

The third research question focussed on the relationship of medium of peer feedback, elaboration and learning gain. It was shown that elaboration of provided peer feedback was positively correlated with learning gain of diagnostic competences (and error detection skills in particular) while the percentage of corrective feedback was negatively correlated with diagnostic competences, especially error detection skills. There was no relationship between prior knowledge and degree elaboration of provided peer feedback or percentage of corrective feedback. Hypothesis 3a about a relationship between medium of providing peer feedback, degree of elaboration and acquisition of diagnostic competences could be accepted. A mediation analysis showed that there was an indirect-only mediation effect of medium of provided peer feedback on acqusition of diagnostic competences by the degree of elaboration of provided peer feedback as a mediator. Hypothesis 3b about the degree of elaboration of peer feedback mediating the effect of medium of peer feedback provision on acquisition of diagnostic competences could be partially accepted with an indirect-only mediation.

Overall, the reception of expert feedback did not positively influence the acquisition of diagnostic competences when learning with erroneous cognitive modeling examples. This is not in line with research highlighting the positive effect of elaborated expert feedback when learning with erroneous written worked examples as found by Heitzmann et al. (2015) and Stark et al. (2011). A possible explanation could be that the operationalization of the expert feedback was written in CASUS and not included in the cases. It might have been hard to properly connect the content of the erroneous cognitive modeling examples with the expert feedback due to the absence of such integration. Kolodner, Cox, and González-Calero (2005) argue that in order to be an effective means for learning from errors, feedback must be presented immediately after observing the error. Furthermore, as pointed out by Heitzmann et al. (2015), when reaching a certain point of expertise in a field, elaborated expert feedback might lead to an expertise reversal effect. This means that received expert elaboration may be redundant and students do not learn from it anymore (Kalyuga et al., 2003). Finally, students' high prior knowledge might be evidence that they were already too advanced to profit from the expert feedback in the present study.

The positive direct effect of spoken peer feedback compared to written peer feedback on the acquisition of diagnostic competences due to increased social presence when speaking – which was assumed based on the studies by Fiorella and Mayer (2013); Hoogerheide, Deijkers, et al. (2016); Hoogerheide, van Wermeskerken, et al. (2016) – was not found. However, by considering and analysing the mediating effect of the degree of elaboration of the provided peer feedback, the possible influence of medium of feedback provision could be demonstrated on acuqisition of diagnostic competences when feedback is elaborated. This is in line with studies on learning from providing elaborated peer feedback: Results show that only when students elaborate deeply on what they observed, they profit from providing peer feedback and acquire knowledge (Y. H. Cho & Cho, 2011; Li et al., 2010).

Hoogerheide, Deijkers, et al. (2016) argue that social presence plays a major role in the positive effect of spoken medium over written when explaining it to others. When engaged in spoken and face-to-face communication, students feel more personally involved and therefore engage more in elaboration of the learning material. Based on the findings, I also want to highlight another possibile explanation. Krych-Appelbaum and Musial (2007) suggested that students who write comments to fictitious others, estimate the recipient to be knowledgeable and have difficulties judging the (fictitious) recipient’s competence level – as results students tend to explain less. Apart from this maljudgement of recepient expertise, student reviewers were found to provide more spoken feedback (compared to written) because it is simply a more efficient, familiar way of communcation (Reynolds & Russell, 2008), which might lead to increased motivation to elaborate on the observed performance, leading to more learning gain due to less mental and physical effort involved in speaking than in writing.

Finally, the findings of this second study highlight the importance of examining the effect of the characteristics of the task to provide (and receive) peer feedback. It can be suggested that researchers and educators determine what works concerning medium, style and circumstances of feedback rounds as they play a crucial role. The present study already demonstrated that the communucation medium influences elaboration and learning. A thorough meta-analysis or empricial literature review of existing peer feedback intervention studies across domains might uncover additional factors that mediate and moderate possible benefical or detrimental effects of feedback provision and reception.

8 Summary and Comparison of the two Empirical Studies

8.1 Summary of Study 1

In Study 1, theoretical assumptions were outlined for why a framework of diagnostic competences composed of diagnostic knowledge types of conceptual, strategic and conditional knowledge proposed by Stark et al. (2011) should be expanded by an inclusion of error detection skills. This proposed expansion was based on the nature of the clinical reasoning process as both deliberate and intuitive (Kahnemann, 2011). To teach diagnostic competences, the beneficial effects of cognitive modeling examples for skill acquisition in other domains might be transferred to medical education (Van Gog & Rummel, 2010). Based on the literature on cognitive modeling examples, research gaps were identified regarding how these examples may be effectively utilized for teaching diagnostic competences. It was unclear whether errors should be incorporated in the cognitive modeling examples and if the task to provide peer feedback can serve as an interactive learning component which triggers engagement, elaboration and ultimately foster the acquisition of diagnostic competences. Therefore, four research questions were formulated: The first research question was if cognitive modeling examples are a more effective approach of acquiring diagnostic competences compared to reading a textbook chapter for diagnosing dyspnea. A second research question targeted the beneficial effect of providing peer feedback (versus just observing) on erroneous cognitive modeling examples (versus correct modeling examples) on the acquisition of diagnostic competences. A third research question examined the effect of both factors on mental effort. A forth research question focused on the relationship between the quality of provided peer feedback on acquisition of diagnostic competences.

The results indicated that cognitive modeling examples can be used as a possible approach to teach diagnostic competences and are better for teaching conceptual knowledge than a textbook. Positive effects of errors in cognitive modeling examples on error detection skills were found. However, the assumed beneficial effects of errors in cognitive modeling examples on the three types of diagnostic knowledge could not be found. In sum, the results by Stark et al. (2011) about beneficial effects of errors in written worked examples on the acquisition of diagnostic knowledge could not completely be transferred to erroneous cognitive modeling examples. Contrary to the assumptions of positive effects on learning gain of the task to provide peer feedback as reported by Li et al. (2010) or Y. H. Cho and Cho (2011) a negative effect was found. However, in the group providing peer feedback on erroneous cognitive modeling examples, degree elaboration of provided peer feedback was related to learning gain. Further exploration of the potential processes that might explain the negative effect did not yield clear results: Mental effort was positively correlated with elaboration of provided peer feedback while there was no significant correlation between the acquisition of diagnostic competences and neither elaboration nor mental effort.

In sum, the results did not offer clear explanations why providing peer feedback was negatively influencing the acquisition of diagnostic competences for the providing student. Given the degree of elaboration of provided peer feedback was correlated with mental effort, the positive correlation of mental effort, incorporating a feedback training and providing the students with expert feedback in the learning intervention may lead to a reduction of mental effort. Further, a change in medium of peer feedback provision from written to spoken was suggested which may increase social presence, elaboration and therefore learning gain. Due to these inconclusive results, it was important to examine these possible influential variables of the task to provide peer feedback, which might influence the effects on learning in a consecutive study (Study 2). Using feedback training, incorporating expert feedback and identifying positive influential factors for providing peer feedback were suggested for a future research design to ensure elaboration and decrease mental effort.

8.2 Summary of Study 2

Based on the results of Study 1 and additional review of literature, incorporating a feedback training, providing students with expert feedback and comparison of medium of feedback provision were identified as factors that could improve student learning from providing peer feedback on cognitive modeling examples. First of all, according to Y. H. Cho and Cho (2011) and Li et al. (2010), providing peer feedback needs to lead to elaboration in order to be beneficial for learning. The assumption that stimulating social presence, which in turn increases elaboration and ultimately learning outcome when providing spoken peer feedback, was based on studies by Hoogerheide, Deijkers, et al. (2016) and Fiorella and Mayer (2013, 2014), who found that spoken explanations for another fictitious person of learning materials are better for learning compared to writing. Further, incorporating expert feedback was assumed to help elaboration on the content of the examples and increase learning gain as demonstrated by Stark et al. (2011). Due to the importance of degree of elaboration of peer feedback for learning and these findings on beneficial effects of spoken over written peer feedback and fruitful incorporation of expert feedback on erroneous examples, two main research questions were formulated: The first research question targeted how (a) written versus spoken peer feedback provision and (b) expert feedback versus no expert feedback as well as (c) their different combinations influence the degree of elaboration of provided peer feedback. The second research question focused on the influence of both varied factors on the acquisition of diagnostic competences. A third one examined the interplay of communication medium and expert feedback with a possible mediation by the degree of elaboration and acquisition of diagnostic competences. Results indicate that while there was no effect of medium of peer feedback provision or reception of expert peer feedback on the acquisition of diagnostic competences, there was an effect of the medium of peer feedback provision on degree of elaboration. More specifically, a mediation analysis showed a positive indirect mediating effect of spoken peer feedback provision on degree of elaboration, which influenced the acquisition of diagnostic competences.

In conclusion, the beneficial effect of the task to provide peer feedback is strongly influenced by the degree of elaboration of the provided peer feedback. It could be shown that this elaboration is influenced by the medium of communication: Students providing spoken peer feedback elaborated more on the observed cognitive modeling examples, which led to more acquisition of diagnostic competences. It was assumed that mainly perceived social presence induced by speaking – similar to the findings of Fiorella and Mayer (2013, 2014) or Hoogerheide, Deijkers, et al. (2016) – was responsible for this positive effect. An explanation for the non-significant effects of expert feedback on elaboration and acquisition of diagnostic competences was given based on assumed incorrect timing: The incorporated expert feedback was not placed right after the error in the modeling example happened, but after providing peer feedback. Kolodner, Owensby, and Guzdial (2004) propose immediate expert feedback on errors in case-based learning methods so that the student can relate the feedback to the error in order to influence learning positively.

The design of Study 2 was based on that of Study 1. Due to similarities in measured variables, samples and design, an empirical comparison is possible to some extent. Therefore, the next sub-chapter aims to compare both studies.

8.3 Comparison of Study 1 and Study 2

Table 7

Age and Semester for the samples of Study 1 (N = 121) and Study 2 (N = 131)

	Study 1 *M(SD*)	Study 2 *M(SD)*
Age	24.38 (2.80).	24.19 (3.03)
Semester	7.00 (0.00)	7.48 (1.55)

Note. Samples depicted include the experimental and control groups

In terms of demographics, the samples were similar concerning age or sex ratio (56% female in Study 1 versus 58% female in Study 2). However, the sample of the first study consisted of medical students of the LMU Medical Center in the seventh semester only, whereas the sample of the second study was more diverse in terms of study semester ranging from a minimum of 5 to a maximum of 11 semesters, making the sample more heterogeneous in terms of prior knowledge. The sample size of Study 2 was also larger with ten more participants.

8.3.1 Diagnostic competences – differences in operationalization and empirical comparison

The knowledge test used for prior knowledge was the same for both studies in terms of conceptual (7 multiple choice items) and strategic knowledge (6 key-feature questions). For prior knowledge in Study 2, two problem-solving tasks were integrated to also assess conditional knowledge. The post-test for diagnostic competences was shortened from Study 1 to Study 2: The conceptual knowledge test was reduced by five multiple-choice items (from 15 to 10), and the test for strategic knowledge by three key-feature question items (from 15 to 12) to make the intervention more compact based on criticism by the students about the length. Conditional knowledge remained unchanged with six open-ended question items. The error detection test was also not changed. Since the control group in Study 1 only read a textbook chapter, 97 participants (out of 121) inform the experimental groups could be included. Study 2 remained unchanged in terms of participants that qualify for direct comparison (*N* = 131). Based on these limitations, the studies can be compared regarding (a) prior knowledge consisting of conceptual and strategic knowledge and (b) identical items of the post-test for diagnostic competences (*N* = 226). A t-test for independent samples was conducted for these comparisons. Requirements for this were met.

There was a significant difference in prior knowledge between the experimental groups of Study 1 (*M* = 54.1, *SD* = 9.3) performing lower than those ofStudy 2 (*M* = 70.7, *SD* = 11.4) in the pre-test, *t*(226) = 12.049, *p* = .00, d = 1.60.

There was also a significant difference in conceptual knowledge gain between the experimental groups of Study 1 (*M* = 79.2, *SD* = 13.8) and Study 2 (*M* = 74.3, *SD* = 13.8), *t*(226) = -2.637, *p* = .009, d = .35. No significant difference was found for strategic knowledge between Study 1 (*M* = 73.5, *SD* = 11.5) and Study 2 (*M* = 71.3, *SD* = 14.1), *t*(226) = -1.275, *p* = .204, or for conditional knowledge between Study 1 (*M* = 60.9, *SD* = 19.1) and Study 2 (*M* = 57.2, *SD* = 17.0), *t*(226) = -1.420, *p* = .157. Comparing error detection skills between Study 1 (*M* = 44.2, *SD* = 18.0) and Study 2 (*M* = 44.2, *SD* = 16.4) revealed no significant difference, *t*(226) = .602, *p* = .548. Overall diagnostic competence acquisition did not differ significantly, *t*(226) = -1.723, *p* = .086, between Study 1 (*M* = 64.6, *SD* = 9.9) and Study 2 (*M* = 62.1, *SD* = 11.4).

In sum, participants in Study 1 had less prior knowledge than participants in Study 2. An explanation might be that study semesters were more heterogeneous in Study 2. Participants in Study 1 were only in the seventh semester. Participants in Study 2 were mixed, ranging from fifth to eleventh semester. The prior knowledge test consisted of easy, medium and hard questions of different complexity. It might be that even though the mean of semesters was similar in Study 1 and 2, knowledge from seventh to eleventh semester does not increase just linear but rather exponentially, leading to an overall higher score in prior knowledge in Study 2 even though both means of semesters are similar (Tomic, Martins, Lotufo, & Benseñor, 2005). A spearman correlation revealed a medium correlation between semester and prior knowledge in Study 2 with $r_s = .389$, $p = .00$.

8.3.2 Differences in degree of elaboration of provided peer feedback

Concerning the degree of elaboration of provided peer feedback, participants in Study 1 provided significantly less elaborated feedback (*M* = 11.0, *SD* = 10.3) (and therefore more (general) corrective feedback) than participants in Study 2 (*M* = 45.8, , *SD* = 19.8) , *t*(140) = 15.309, *p* = .001:

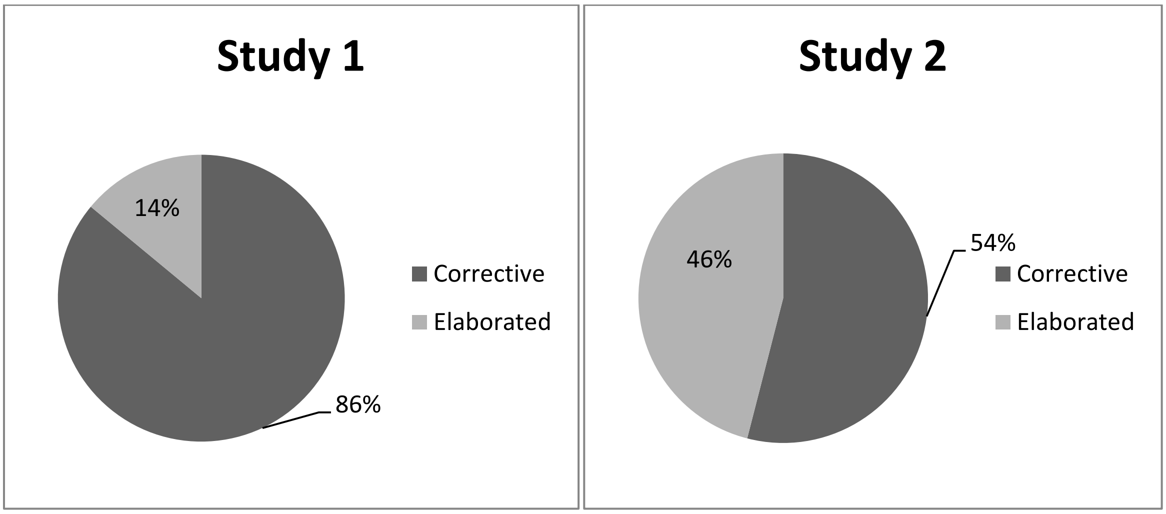

Figure 6. Comparison of degree of elaboration between Study 1 and Study 2

One explanation might be that just erroneous examples were used in Study 2, leading to a higher possibility to elaborate on the observed cognitive modeling examples: This is supported by the finding that participants in Study 1 provided significantly more elaborated peer feedback on erroneous cognitive modeling examples (*M* = 15.6, *SD* = 13.4) than correct cognitive modeling examples (*M* = 6.4, *SD* = 7.2), *t*(44) = 2.86, *p* = .005.

However, there was still a significant increase in degree of elaboration when comparing providing peer feedback on erroneous examples from Study 1 (*M* = 15.6, *SD* =

13.4) and Study 2 (M = 45.8, SD = 19.8), $t(48.5) = 8.68$, $p = .000$. Therefore, the increased degree of elaboration could not only be explained by the errors in cognitive modeling examples. Prior knowledge could not have been responsible for the difference in degree of elaboration as no correlation was found between prior knowledge and elaboration in Study 2 ($r_s = .08$, p = n.s.). Therefore, as another explanation for the difference in degree of elaboration it seems plausible that the increase in elaboration might be due to the short feedback training integrated in Study 2. Training medical students to provide effective peer feedback for learning was advised by Prins et al. (2006) after finding that general practitioners in training lack the ability to provide adequate feedback in other domains such as teacher training. Training peer revision skills has been found to be effective to teach feedback provision skills, that is after short trainings, students' quality of provided peer feedback significantly improved (Min, 2005; Rahimi, 2013; Sluijsmans et al., 2002).

Overall, the comparison of Study 1 and Study 2 highlights two important added findings: First, more prior knowledge does not lead to providing more elaborated peer feedback. Second, students who received a short feedback provision training prior to the task provided more elaborated peer feedback than students who did not.

9 General Discussion

In the following general discussion, section 9.1 provides interpretations, a discussion and implications of the findings. Section 9.2 focusses on conclusive answers to the main research questions of this thesis. Limitations and a future outlook are outlined in section 9.3.

9.1 Interpretation, Discussion and Implications of the Findings

9.1.1 Acquiring parts of diagnostic competences with cognitive modeling examples

In Study 1, cognitive modeling examples were found to be significantly superior for learning conceptual knowledge compared to textbook learning. No significant difference for strategic or conditional knowledge or error detection skills was found. While cognitive modeling examples were assumed to be effective for acquiring conceptual and procedural (strategic and conditional) knowledge parts of diagnostic competences (Braaksma et al., 2002; Custers, Regehr, McCulloch, Peniston, & Reznick, 1999; Van Gog & Rummel, 2010), Nokes-Malach, VanLehn, Belenky, Lichtenstein, and Cox (2013) found that written worked examples focus learners’ attention on conceptual knowledge acquisition rather than fostering problem-solving skills. Their results showed that students acquired more conceptual knowledge when learning with worked examples compared to learning with textbooks. No difference was found for problem-solving skills. The results of Nokes-Malach et al. (2013) are in line with the results of Study 1, supporting the hypothesis of Hoogerheide et al. (2014) that the effects of written worked examples on learning could be similar and transferred to learning settings with cognitive modeling examples. Overall, the non-significant results of cognitive modeling examples for other parts of diagnostic competences found in Study 1 question the effectiveness of cognitive modeling examples for learning diagnostic competences because of high effort involved in creating, piloting and incorporating such examples into learning environments for a limited advantage over learning with a textbook.

However, the learning material chosen for the control group was a text book chapter medical students normally use to prepare for their exams (Füeßl & Middeke, 2010). This book was also used to create and validate several questions of the diagnostic knowledge post-test. It covered a holistic and detailed knowledge base about the illnesses that were target of the knowledge test questions. This might have led to an advantage for the control group concerning correctly answering the post-test knowledge questions, outweighing possible advantageous effects on learning with cognitive modeling examples in comparison. Furthermore, the students are familiar with learning with such textbooks for exams. Even though students were familiar with CASUS, the learning environment was specifically designed and differed significantly to other environments in CASUS. A familiarization task for the specifically designed learning environment may have been useful to counter the possible influential difference of familiarity to learning materials (Wegerif, 1998). Both the detailed content and the familiarity of the learning material of the control group might have led to the rather high results of the control group in the post-test, which might have led to small non-significant differences in strategic and conditional knowledge parts compared to the experimental groups.

The findings that cognitive modeling examples are suitable for learning conceptual knowledge parts of diagnostic competences add to the theoretical foundations. Cognitive modeling examples have been proven effective in different, less structured domains such as writing (Braaksma et al., 2002) or more across-domain skills like collaboration, communication of self-regulative skills (Baldwin, 1992; Rummel & Spada, 2005; Zimmerman & Kitsantas, 2002). They were also found to be effective in medical education, but for procedural (not cognitive) skills like surgical skills or digital rectal examination (Custers et al., 1999; Stegmann, Pilz, Siebeck, & Fischer, 2012). The findings of Study 1 however advance this body of literature by finding benefits of cognitive modeling examples for conceptual diagnostic knowledge.

The presented results highlight that cognitive modeling examples should be incorporated into medical education practice for teaching medical students conceptual parts of diagnostic competences. One way to overcome the high effort to create cognitive modeling example would be to use OSCE settings, which are objective structured clinical examination tests in simulations where students’ performance in a simulation is rated, to create cognitive modeling examples. OSCE are reliable and valid but very expensive in terms of manpower, time and costs (Carraccio & Englander, 2000; Petrusa, Blackwell, & Ainsworth, 1990; Zayyan, 2011). Creating and validating (erroneous) cognitive modeling examples of students’ performances during the OSCE simulations should increase the accessibility of the performance to other students and therefore make the high expenses worthwhile.

9.1.2 Effects of errors in cognitive modeling examples on error detection skills

In Study 1, the effect of erroneous versus correct cognitive modeling examples was examined on acquiring diagnostic competences. To my knowledge, erroneous cognitive modeling examples have been used either in studies where they were presented simultaneously with correct ones for comparison as analogic reasoning task for psycho-/motor learning (Domuracki et al., 2015; Renkl, 2014) or alone in relation to increased attention and motivation (Renkl & Atkinson, 2003). The specific added benefits of only erroneous modeling examples on diagnostic competences and especially error detection skills – to my knowledge –has not been examined so far. In this respect, the first study of this thesis strengthens the suggestion to integrate errors into cognitive modeling examples to teach error related knowledge and skills that might lead to greater accuracy: The suggestion by Gartmeier et al. (2008, 2010) to expose learners to erroneous performances to teach error detection and error avoidance is applicable for erroneous cognitive modeling examples, even in a short learning intervention. The added benefit of using erroneous examples to teach error avoidance instead of real-life situations with actual patients can increase patient safety. However, results indicated that while erroneous cognitive modeling examples were superior for acquiring error detection skills when compared to correct cognitive modeling examples, there was no significant difference in conceptual, strategic or conditional knowledge acquisition. First and foremost, this might have been due to the content of the erroneous and correct examples being largely identical in all three knowledge parts included. The errors incorporated were not knowledge-based, but based on decision making biases. Therefore, they might have been engaging and motivating as proposed by Van Gog and Rummel (2010), but might have led the students' to focus on where and how the error during the diagnoses occurred, which led to increased error detection skills, but not increased knowledge.

The beneficial effect of erroneous cognitive modeling on error detection skills could be transferred into medical education practice. Critics of use of errors in medical education might claim that erroneous learning material is detrimental for learning when the error is left uncommented. However, as highlighted by Renkl and Atkinson (2003), erroneous examples are fruitful for intermediate and advanced learners when the error is highlighted and therefore found. The results indicate that errors in medical education settings are beneficial for error detection, which is, according to Gartmeier et al. (2010) and Dror (2011), an important part of medical expertise. So far, the approach in medical education to handle errors leans towards the attitude to "deny the mistakes that happened and vigorously defend against malpractice claims" (Rocke & Lee, 2013, p. 550). However, Study 1 showed that being exposed to errors in diagnosing seems to be beneficial for a distinct set of skills connected to mistakes. This adds to the findings from Domuracki et al. (2015), Gartmeier et al. (2010) and Ziv et al. (2005) about the importance of being exposed to errors in medical education.

9.1.3 Effects of (elaborated) peer feedback provision on the acquisition of diagnostic competences

So far, isolated effects of peer feedback provision on the provider's knowledge acquisition have been examined in a limited number of studies in different domains such as writing (K. Cho & MacArthur, 2011; Y. H. Cho & Cho, 2011) or technology education (Li et al., 2010). To my knowledge, both of the empirical studies of this thesis are the first to examine how providing peer feedback on videotaped peer diagnostic reasoning via different communication media affects the acquisition of diagnostic competences.

In Study 1, detrimental effects of providing peer feedback on cognitive modeling examples were found. A possible explanation for the negative effect of the task to provide peer feedback was that students were unfamiliar with providing peer feedback and mental overload led to negative effects on the acquisition of diagnostic competences. Analysis of quality of provided peer feedback in Study 1 showed that it was mostly corrective and not elaborated, which is in line with literature about the inability of students to provide ad-hoc adequate peer feedback without training (Min, 2005; Prins et al., 2006; Rahimi, 2013). This training for elaborated feedback was implemented to familiarize and enable students to provide elaborated peer feedback, which in turn should lead to a positive effect of peer feedback provision on learning as proposed by Y. H. Cho and Cho (2011) and Li et al. (2010). Therefore, a short training was included that emphasized why it is important and how to provide elaborated peer feedback at the beginning of the intervention of Study 2 to increase elaboration and decrease mental effort.

In comparison to Study 1, students in Study 2 provided more elaborated peer feedback on the diagnoses of the cognitive modeling examples than students in Study 1. The advanced prior knowledge of students in Study 2 did not reveal a significant effect on the degree of elaboration. Therefore, the increase of degree of elaboration of provided peer feedback may be caused by the feedback training as suggested by Prins et al. (2006) and Sluijsmans et al.

(2002). However, no significant differences in acquisition of diagnostic competences were found in direct comparison of Study 1 and Study 2. A relationship of degree of elaboration with acquisition of diagnostic competences however can be assumed due to the negative effect of corrective feedback in Study 1 and the positive effect of elaborated peer feedback in Study 2 on the acquisition of diagnostic competences.

First, the depicted results advance the literature on the impact of providing peer feedback on the provider for learning as depicted by Y. H. Cho and Cho (2011), K. Cho and MacArthur (2011) and Li et al. (2010). Peer feedback provision can be used as interactive learning method when learning with cognitive modeling examples, given the students elaborate on the observed performances. The presented results could be translated into medical education practice.

When medical educators use peer feedback tasks for learning, they should be aware of the difficulties that arise with this. Medical students seem to have problems with providing elaborated peer feedback, which might be detrimental for learning. Therefore, training students how to effectively provide peer feedback should be considered before peer feedback for learning is utilized. Furthermore, integrating guidance in form of feedback prompts (Baker & Lund, 1997; Gielen et al., 2010) or feedback training (Sluijsmans et al., 2002) should be provided. Just providing peer feedback without any guidance might be detrimental for learning. In reference to the example of OSCE settings earlier, observing students can be either trained beforehand or instructed during the task to provide elaborated peer feedback on the peer in the simulation to increase acquisition if diagnostic competences.

9.1.4 Medium of peer feedback provision on elaboration and acquisition of diagnostic competences

In Study 2, the difference between spoken versus written peer feedback provision on elaboration and acquisition of diagnostic competences was examined. Providing spoken peer feedback increased the degree of elaboration, which led to increased acquisition of diagnostic competences. These results advance the results of Hoogerheide, Deijkers, et al. (2016) of the advantage of the task to provide spoken over written explanations to a fictitious other when learning syllogistic reasoning tasks with examples: The results revealed that their findings can be transferred to the task of providing spoken peer feedback on diagnostic reasoning of cognitive modeling examples in medicine. A possible explanation for the positive effect of providing spoken peer feedback on the degree of elaboration and therefore learning gain were given: First, social presence was assumed to be influential when speaking versus writing. Social presence was assessed with a short form of the temple inventory for presence (Lombard, Ditton, & Weinstein, 2009), which was translated into German but was too short to yield reliable results. Hoogerheide, Deijkers, et al. (2016) measured social presence by coding the content of the given explanations. The presence and absence of pronouns was used as indicator for perceived social presence. A similar approach could have been used in my study as well. This method is based on suggestions and assumptions that pronouns might be seen as a rather weak indicator for social presence (Rourke, Anderson, Garrison, & Archer, 2007) unlike humor or emotions – which were not coded in the case of Hoogerheide, Deijkers, et al. (2016), and not reliably coded in the study by Rourke et al. (2007). Because I assumed that the learning setting (critical patient cases) was not an appropriate place for humor and emotions, this was neither expected nor yielded in the results. Thus, this was not pursued in the assessment. For practice use, this means that educators must choose the medium with which students provide each other feedback adequately and with care. Influential indicators could be the familiarity of the students or possibly the social presence this medium provides.

9.1.5 Ineffectiveness of expert feedback for acquisition of diagnostic competences

Unlike the beneficial effects of incorporating expert feedback when learning with written worked examples as found by Stark et al. (2011), no such benefits were found for cognitive modeling examples. A possible explanation for this might be that the expert feedback was not presented directly after watching the erroneous cognitive modeling examples. In order to have a beneficial effect on learning outcome, expert feedback needs to be presented shortly after the error was observed so that students can relate the feedback to the error and make sense of it (Kolodner et al., 2004). A timely delay might have led to a non-significant effect of incorporating expert feedback due to the students not being able to properly relate the expert feedback to the prior observed performance.

Furthermore, students in Study 2 were on an advanced level of prior diagnostic knowledge, which could have led to the expertise reversal effect: Presented expert feedback was conflicting with an already existing constructed explanation of the errors based on the students' prior knowledge, making it ineffective for learning (Heitzmann et al., 2015; Kalyuga et al., 2003). To counter this effect, incorporating adaptable feedback that could be adjusted to the learners' needs was proven to be effective when learning with worked examples (Heitzmann et al., 2015).

Contrary to other findings of worked example research (i.e. about feasibility for learning conceptual knowledge, see section 9.1.1), results of studies on expert feedback on written worked examples (Stark et al., 2011) could not be transferred to learning with expert feedback on cognitive modeling examples as the findings in Study 2 indicate. Therefore, it is to question if students profit from expert feedback under all circumstances. Adequate and well-timed expert feedback that is tailored on the students' needs is suggested for practice use (Heitzmann et al., 2015; Kolodner et al., 2005; Kolodner et al., 2004).

9.2 Conclusions

The first main research question focused on how medical students acquire diagnostic competences from providing peer feedback on erroneous or correct cognitive modeling examples of clinical reasoning: Study 1 showed that cognitive modeling examples are a valid alternative to learning with textbooks for acquiring diagnostic competences as they were superior to textbook learning for conceptual parts of diagnostic knowledge. Study 1 also highlights the importance of errors integrated into them. Even though erroneous cognitive modeling examples were not per se more beneficial for learning diagnostic knowledge when compared to correct ones, they offered advantages for students by increasing error detection skills. This is in line with my expectations that being exposed to erroneous behavior may increase meta-cognitive knowledge about where an error might happen or has happened in consecutive performance (Dror, 2011; Gartmeier et al., 2008). Providing corrective peer feedback was found to be detrimental for learning. Students should be familiarized and trained how to give elaborated peer feedback, which is beneficial for learning.

Concerning the effect of providing peer feedback on the cognitive modeling examples on learning, the results of Study 1 were at first contrary to the hypothesis: Students acquired less diagnostic competences when they provided peer feedback on cognitive modeling examples compared to just observing. This was due to the superficial nature of the corrective peer feedback they provided. Studies highlight the importance of the degree of elaboration of students' peer feedback in order for it to be fruitful for learning (K. Cho & MacArthur, 2011; Y. H. Cho & Cho, 2011). In response to the results of Study 1 which emphasized the importance of the degree of elaboration when providing peer feedback for learning, a training why and how to provide elaborated peer feedback was implemented in Study 2, which increased elaboration. In Study 2, there was a positive effect of degree of elaboration of peer feedback on learning gain. The assumed negative effect of corrective feedback in Study 1 found support in these results of Study 2. The novelty of the task to provide peer feedback was hypothesized to be responsible for a higher mental effort and less elaboration, which was assumed to be countered by the training.

The second main question focused on the influence of expert feedback and the medium of communication of the task to provide peer feedback on erroneous cognitive modeling on elaboration and acquisition of diagnostic competences. The communication medium for providing peer feedback influenced the degree of elaboration further: Students who provided spoken peer feedback elaborated more than students providing written peer feedback, and this elaboration led to better learning of diagnostic competences. This is in line with Hoogerheide, Deijkers, et al. (2016) and Fiorella and Mayer (2013, 2014) examining the positive effect of spoken communication medium on learning gain when explaining. Contrary to the beneficial effect of communication medium, receiving expert feedback did not enhance elaboration or learning as assumed according to studies like Stark et al. (2011).

To summarize, the results show that using erroneous cognitive modeling examples in medical education has beneficial effects on parts of diagnostic competences, which are conceptual knowledge and error detection skills. Students providing spoken, elaborated peer feedback learn most from erroneous cognitive modeling examples. This elaboration could be further enhanced with a short training about why and how to effectively provide elaborated peer feedback before the task to provide peer feedback. However, the results of both presented studies offer possibilities for criticism and additional future directions. The following section discusses possible limitations and an outlook for future research

9.3 Limitations and Outlook for Future Research

The presented empirical studies can be critically discussed in terms of methodological limitations which are outlined in the following concerning instruments in section 9.3.1 and learning material in section 9.3.2. After this, a possible outlook in section 9.3.3 for future research based on this thesis is described.

9.3.1 Instrument limitations

A first limitation is the way diagnostic competences were assessed in both studies. Epstein and Hundert (2002) define professional competence in medicine as “the habitual and judicious use of communication, knowledge, technical skills, clinical reasoning, emotions, values, and reflection in daily practice for the benefit of the individual and community being served” (Epstein & Hundert, 2002, p. 226). Despite consisting of several knowledge component questions based on Stark et al. (2011) as well as error detection skill tasks, the pre- and post-test used might not fully depict the performance of the students in real diagnostic settings. This might decrease external validity, similar to for example objective structured clinical examination tests in simulations (OSCE) that simulate such a daily situation (Carraccio & Englander, 2000; Petrusa et al., 1990). A possible alternative that shows sufficient predictors for later clinical performance might be the script-concordance test (SCT) (Brailovsky, Charlin, Beausoleil, Cote, & Van der Vleuten, 2001; Monnier, Bédard, Gagnon, & Charlin, 2011). However, it can be questioned if the SCT offers more validity and reliability than the test proposed by Stark et al. (2011), because it is similar to the test used in the empirical studies of this thesis in terms of creating, validation and rating by independent experts. Future research about how to assess diagnostic competences might tap into the question about a gold standard for the assessment of diagnostic competences by relating and comparing several assessment methods with each other by a multi-trait multi-method approach (Campbell & Fiske, 1959; Monnier et al., 2011).

Another limitation is the additional part of error detection skills to the three diagnostic knowledge parts of three components testing (Stark et al., 2011; Van Gog et al., 2004). While the error detection test has been used before in other studies (Nyssen & Blavier, 2006), and proven reliable in both studies of this thesis, the fit of the error-detection part relative to the knowledge parts constituting diagnostic competences was not empirically examined. One way to clarify the relationship between error detection skills with other performance measures would be to relate it to actual diagnostic reasoning accuracy and performance in a real life setting in similar and transfer cases. For both studies, a three-component concept of diagnostic competences was extended with a forth component of error detection skills. First, further research should address the concept of diagnostic competences as a three or four component concept by assessing conceptual, strategic and conditional knowledge and error detection skills with for example an exploratory factor analysis. Therefore, it is important to relate error detection skills to performance measurements and look for predictive validity of it for diagnostic accuracy and patient care. Such an approach was already made by Wilkinson et al. (2011) in exploring how expertise of dialysis technicians correlates with their error detection skills. To correlate error detection skills and the knowledge parts of the concept of diagnostic competences used for this thesis with the diagnostic accuracy of physicians in real-life or simulation settings could give further insight into how error detection skills, conceptual, strategic and conditional knowledge are related or even predicting to ‘actual competence’ and performance.

Another limitation in terms of measurements of covariates is that overall mental effort was measured in both studies instead of two (intrinsic and extraneous) or three components of germane, intrinsic and extraneous cognitive load but as an overall score obtained with the 8-item inventory by Paas and Kalyuga (2005). Even though the results of Study 1 showed a correlation of mental effort and providing peer feedback that indicate that different types of cognitive load might be related to providing peer feedback, the instrument to assess mental effort was reused in Study 2. In order to see the relationship between provide peer feedback and different types of cognitive load, future studies could use the questionnaire proposed by Leppink et al. (2013).

9.3.2 Learning material limitations

Concerning external validity, a limitation arises from the use of video-based models in a computer-based learning environment. This approach was chosen as it is superior to real-life settings in terms of standardization of observed scripted performances and allowed for more control to investigate the isolated effects of feedback provision. It was ensured that every student observed the same performance by a fictional peer. However, the task to provide peer feedback on video-based cognitive modeling examples in a computer-based learning environment like CASUS might not be comparable to a peer feedback situation in real life or simulations because it might be perceived as unauthentic, which could have led to different behavior or changed the attitude of students compared to real-life settings. In my study designs, the model's performance in the video was used as target for the task to provide peer feedback. The model could not react to the comments made by the providing student. This reaction and interaction between provider and recipient of peer feedback might lead to different processes and dynamics of providing peer feedback, influencing the effects on learning. For example, Liu and Carless (2006) highlight the importance of dialogue in peer feedback to be beneficial for learning. A study contrasting the effects of providing peer feedback on performance of cognitive modeling examples versus providing peer feedback on performance of a student in a simulation with and without the possibility of dialogue could give insights into the ecological validity of the task to provide peer feedback on cognitive modeling examples. The effects of peer feedback provision on learning as found in both studies might then give further insight on a possible generalizability to other settings of domains.

9.3.3 Future outlook

The empirical studies presented in Chapters 6 and 7 shed light on the positive effects of using erroneous cognitive modeling examples of a peer’s diagnostic process to acquire

Renkl (2014) highlights the beneficial effects of mixing erroneous and correct examples in *analogic reasoning*. Students exposed to both correct and erroneous examples need to compare and find out differences that could lead to errors. Future studies examining the effect on analogic reasoning (compared to erroneous and correct examples only) on the acquisition of complex cognitive skills like diagnostic competences might give insight into whether a combination of erroneous and correct examples, which is analogical reasoning, can be used in medical education and if it is advantageous over showing just correct or erroneous examples. Also, the external validity of cognitive modeling examples could be the target of future research, for example by contrasting the effects of cognitive modeling examples on learning with the effects of observing real-life diagnostic reasoning and diagnostic reasoning in simulations.

The ‘hidden’ cognitive processes of the students during providing peer feedback were not explicitly examined in the empirical studies of this thesis. Future research could tackle unveiling these processes. Verbalizing the reasoning processes during providing peer feedback or just observing cognitive modeling examples with think-aloud method (Cotton & Gresty, 2006; Fonteyn & Fisher, 1995) might interfere too much. Therefore, retrospective judgement and analysis as proposed by Van Den Haak, De Jong, and Jan Schellens (2003) of the peer feedback recorded with the dictation machine by the student may be starting point for further insight into the reasoning of students during providing peer feedback and what factors influenced their provision process.

Furthermore, the feedback provision training in Study 2 was short but effective. As results of Study 1 indicate, exposing students to a guidance sheet of which parts are important to keep in mind (see Appendix G) while providing peer feedback on examples was not effective and did not lead to elaboration of the provided peer feedback. Contrasting different degrees of elaboration, length and detail of the feedback training might be a next step and clarify what leads to beneficial effects of feedback provision training on elaboration of provided peer feedback.

Another possibility for future research regarding the effectiveness of learning with cognitive modeling examples may be to incorporate practice phases after watching cognitive modeling examples. In the empirical studies of this thesis, providing peer feedback was used as interactive learning activity. However, a body of research on vicarious (or observational) learning highlights the importance of a rehearsal of the observed with a practice phase after observing (Bandura & Jeffrey, 1973; Fryling, Johnston, & Hayes, 2011; Stegmann et al., 2012). Exploring if, how and when incorporating training phases after (or before) observing cognitive modeling examples and relating and comparing this effect to providing peer feedback could be object of investigations for further studies.

9.4 In closing

This thesis advances the knowledge about positive effects of providing peer feedback on peers' performances on the provider and how watching erroneous examples positively affects learning.

When peer feedback provision tasks are used for learning, educators need to carefully design these tasks to make them beneficial for both the recipient and the provider of the feedback. Especially the degree of elaboration of the provided peer feedback influences the learning gain and must be ensured. Further research on the influential factors that determine elaboration is highly needed.

In response to a body of research that suggests that errors can be used for learning diagnostic competences – with often just theoretical assumptions how this might be possible – the results of this thesis highlight that using erroneous examples in medical education helps acquiring error-detection skills. The distinct value of learning from errors for students and educators should gain more attention in empirical research. Exposing medical students to errors during learning might create a more holistic diagnostic skill and knowledge repertoire of young physicians, leading to better patient care with better diagnostic accuracy. Maybe, accepting and learning from our errors is essential for experiencing the full picture.

References

Aamodt, A., & Plaza, E. (1994). Case-based reasoning: Foundational issues, methodological variations, and system approaches. *AI communications, 7*(1), 39-59.

Althauser, R., & Darnall, K. (2001). Enhancing critical reading and writing through peer reviews: An exploration of assisted performance. *Teaching Sociology*, 23-35.

Anderson, J. R. (1987). Skill acquisition: Compilation of weak-method problem situations. *Psychological review, 94*(2), 192-210.

Anderson, J. R. (1990). *Cognitive psychology and its implications*. New York, NY: WH Freeman/Times Books/Henry Holt & Co.

Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. (2000). Learning from examples: Instructional principles from the worked examples research. *Review of educational research, 70*(2), 181-214.

Atkinson, R. K., Renkl, A., & Merrill, M. M. (2003). Transitioning From Studying Examples to Solving Problems: Effects of Self-Explanation Prompts and Fading Worked-Out Steps. *Journal of educational psychology, 95*(4), 774-783.

Badets, A., & Blandin, Y. (2004). The role of knowledge of results frequency in learning through observation. *Journal of Motor Behavior, 36*(1), 62-70.

Badets, A., & Blandin, Y. (2005). Observational learning: Effects of bandwidth knowledge of results. *Journal of Motor Behavior, 37*(3), 211-216.

Badets, A., Blandin, Y., Wright, D. L., & Shea, C. H. (2006). Error detection processes during observational learning. *Research quarterly for exercise and sport, 77*(2), 177-184.

Baker, M., & Lund, K. (1997). Promoting reflective interactions in a CSCL environment. *Journal of Computer Assisted Learning, 13*(3), 175-193.

Baldwin, T. T. (1992). Effects of alternative modeling strategies on outcomes of interpersonal-skills training. *Journal of Applied Psychology, 77*(2), 147-154.

Bandura, A. (1981). Self-referent thought: A developmental analysis of self-efficacy. *Social cognitive development: Frontiers and possible futures*, 200-239.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Bandura, A., & Jeffrey, R. W. (1973). Role of symbolic coding and rehearsal processes in observational learning. *Journal of personality and social psychology, 26*(1), 122-130.

Bao, L., Cai, T., Koenig, K., Fang, K., Han, J., Wang, J., . . . Luo, Y. (2009). Learning and scientific reasoning. *Science, 323*(5914), 586-587.

Berner, E. S., & Graber, M. (2008). Overconfidence as a cause of diagnostic error in medicine. *The American journal of medicine, 121*(5), 2-23.

Berthold, K., & Renkl, A. (2009). Instructional aids to support a conceptual understanding of multiple representations. *Journal of educational psychology, 101*(1), 70-87.

Blandin, Y., & Proteau, L. (2000). On the cognitive basis of observational learning: Development of mechanisms for the detection and correction of errors. *The Quarterly Journal of Experimental Psychology: Section A, 53*(3), 846-867.

Blomberg, G., Renkl, A., Sherin, M. G., Borko, H., & Seidel, T. (2013). Five research-based heuristics for using video in pre-service teacher education. *Journal for educational research online, 5*(1), 90-114.

Boekhout, P., Gog, T., Wiel, M. W., Gerards-Last, D., & Geraets, J. (2010). Example-based learning: Effects of model expertise in relation to student expertise. *British Journal of Educational Psychology, 80*(4), 557-566.

Boshuizen, H. P., & Schmidt, H. G. (2008). The development of clinical reasoning expertise. *Clinical reasoning in the health professions, 3*, 113-121.

Bowen, J. L. (2006). Educational strategies to promote clinical diagnostic reasoning. *New England Journal of Medicine, 355*(21), 2217-2225.

Braaksma, M. A., Rijlaarsdam, G., & Van den Bergh, H. (2002). Observational learning and the effects of model-observer similarity. *Journal of educational psychology, 94*(2), 405-415.

Brailovsky, C., Charlin, B., Beausoleil, S., Cote, S., & Van der Vleuten, C. (2001). Measurement of clinical reflective capacity early in training as a predictor of clinical reasoning performance at the end of residency: an experimental study on the script concordance test. *Medical education, 35*(5), 430-436.

Branch, W. T. J., & Paranjape, A. (2002). Feedback and Reflection: Teaching Methods for Clinical Settings. *Academic Medicine, 77*(12, Part 1), 1185-1188.

Brennan, T. A. (2000). The Institute of Medicine report on medical errors—could it do harm? *New England Journal of Medicine, 342*(15), 1123-1125.

Brünken, R., Moreno, R., & Plass, J. L. (2010). *Cognitive load theory*: Cambridge University Press.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological bulletin, 56*(2), 81-105.

Carraccio, C., & Englander, R. (2000). The objective structured clinical examination: a step in the direction of competency-based evaluation. *Archives of pediatrics & adolescent medicine, 154*(7), 736-741.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*(4), 293-332.

Charlin, B., Lubarsky, S., Millette, B., Crevier, F., Audétat, M. C., Charbonneau, A., . . . Bourdy, C. (2012). Clinical reasoning processes: unravelling complexity through graphical representation. *Medical education, 46*(5), 454-463.

Charlin, B., Tardif, J., & Boshuizen, H. P. (2000). Scripts and medical diagnostic knowledge: theory and applications for clinical reasoning instruction and research. *Academic Medicine, 75*(2), 182-190.

Chi, M. T. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science, 1*(1), 73-105.

Chi, M. T., Glaser, R., & Farr, M. J. (2014). *The nature of expertise*: Psychology Press.

Chi, M. T., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist, 49*(4), 219-243.

Cho, K., & MacArthur, C. (2011). Learning by reviewing. *Journal of educational psychology, 103*(1), 73-84.

Cho, Y. H., & Cho, K. (2011). Peer reviewers learn from giving comments. *Instructional Science, 39*(5), 629-643.

Clarke, J. R., Spejewski, B., Gertner, A. S., Webber, B. L., Hayward, C. Z., Santora, T. A., . . . Fabian, T. C. (2000). An objective analysis of process errors in trauma resuscitations. *Academic Emergency Medicine, 7*(11), 1303-1310.

Coderre, S., Mandin, H., Harasym, P., & Fick, G. (2003). Diagnostic reasoning strategies and diagnostic success. *Medical education, 37*(8), 695-703.

Cooper, G., & Sweller, J. (1987). Effects of schema acquisition and rule automation on mathematical problem-solving transfer. *Journal of educational psychology, 79*(4), 347-362.

Cotton, D., & Gresty, K. (2006). Reflecting on the think-aloud method for evaluating e-learning. *British Journal of Educational Technology, 37*(1), 45-54.

Craig, S. D., Chi, M. T., & VanLehn, K. (2009). Improving classroom learning by collaboratively observing human tutoring videos while problem solving. *Journal of educational psychology, 101*(4), 779-789.

Croskerry, P. (2003). The importance of cognitive errors in diagnosis and strategies to minimize them. *Academic medicine, 78*(8), 775-780.

Croskerry, P. (2009). Clinical cognition and diagnostic error: applications of a dual process model of reasoning. *Advances in health sciences education, 14*(1), 27-35.

Croskerry, P., Singhal, G., & Mamede, S. (2013a). Cognitive debiasing 1: origins of bias and theory of debiasing. *BMJ quality & safety, 22*(Suppl 2), 58-64.

Croskerry, P., Singhal, G., & Mamede, S. (2013b). Cognitive debiasing 2: impediments to and strategies for change. *BMJ quality & safety,* 65-72.

Custers, E. J., Regehr, G., McCulloch, W., Peniston, C., & Reznick, R. (1999). The effects of modeling on learning a simple surgical procedure: see one, do one or see many, do one? *Advances in Health Sciences Education, 4*(2), 123-143.

Cyr, D., Hassanein, K., Head, M., & Ivanov, A. (2007). The role of social presence in establishing loyalty in e-service environments. *Interacting with computers, 19*(1), 43-56.

Davies, P. (2000). Computerized peer assessment. *Innovations in Education and Teaching International, 37*(4), 346-355.

Davis, F. D. (1997). Phronesis, Clinical Reasoning and Pellegrino's Philosophy of Medicine. *Theoretical Medicine, 18*(1), 173-195. doi:10.1023/a:1005709722944

De Jong, T. (2010). Cognitive load theory, educational research, and instructional design: some food for thought. *Instructional Science, 38*(2), 105-134.

Domuracki, K., Wong, A., Olivieri, L., & Grierson, L. E. (2015). The impacts of observing flawed and flawless demonstrations on clinical skill learning. *Medical education, 49*(2), 186-192.

Dror, I. (2011). A novel approach to minimize error in the medical domain: cognitive neuroscientific insights into training. *Medical teacher, 33*(1), 34-38.

Elstein, A. S. (2009). Thinking about diagnostic thinking: a 30-year perspective. *Advances in health sciences education, 14*(1), 7-18.

Elstein, A. S., & Schwarz, A. (2002). Clinical problem solving and diagnostic decision making: selective review of the cognitive literature. *British Medical Journal, 324*(7339), 729-732.

Elstein, A. S., Shulman, L. S., & Sprafka, S. A. (1978). Medical problem solving an analysis of clinical reasoning.

Ely, J. W., & Graber, M. A. (2015). Checklists to prevent diagnostic errors: a pilot randomized controlled trial. *Diagnosis, 2,* 163-169.

Epstein, R. M., & Hundert, E. M. (2002). Defining and assessing professional competence. *Jama, 287*(2), 226-235.

Ericsson, K. A. (2004). Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Academic Medicine, 79*(10), S70-S81.

Eva, K. W. (2005). What every teacher needs to know about clinical reasoning. *Medical education, 39*(1), 98-106.

Eva, K. W., Link, C. L., Lutfey, K. E., & McKinlay, J. B. (2010). Swapping horses midstream: factors related to physicians’ changing their minds about a diagnosis. *Academic medicine: journal of the Association of American Medical Colleges, 85*(7), 1112-1117.

Falchikov, N. (2001). *Learning together: Peer tutoring in higher education*. New York, NY: Routledge Psychology Press.

Feltovich, P. J., Coulson, R. L., Spiro, R. J., & Dawson-Saunders, B. K. (1992). Knowledge application and transfer for complex tasks in ill-structured domains: Implications for instruction and testing in biomedicine *Advanced models of cognition for medical training and practice* (pp. 213-244): Springer.

Fiorella, L., & Mayer, R. E. (2013). The relative benefits of learning by teaching and teaching expectancy. *Contemporary Educational Psychology, 38*(4), 281-288.

Fiorella, L., & Mayer, R. E. (2014). Role of expectations and explanations in learning by teaching. *Contemporary Educational Psychology, 39*(2), 75-85.

Fischer, F., Kollar, I., Ufer, S., Sodian, B., Hussmann, H., Pekrun, R., . . . Fischer, M. (2014). Scientific reasoning and argumentation: Advancing an interdisciplinary research agenda in education. *Frontline Learning Research, 2*(3), 28-45.

Fischer, M. (2000). CASUS-An authoring and learning tool supporting diagnostic reasoning. *Zeitschrift für Hochschuldidaktik, 1*(1), 87-98.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American psychologist, 34*(10), 906-911.

Fonteyn, M., & Fisher, A. (1995). Use of think aloud method to study nurses' reasoning and decision making in clinical practice settings. *Journal of Neuroscience Nursing, 27*(2), 124-128.

Fryling, M. J., Johnston, C., & Hayes, L. J. (2011). Understanding observational learning: An interbehavioral approach. *The Analysis of verbal behavior, 27*(1), 191-203.

Füeßl, H., & Middeke, M. A. (2010). *Klinische Untersuchung*: Stuttgart: Thieme.

Gartmeier, M., Bauer, J., Gruber, H., & Heid, H. (2008). Negative knowledge: Understanding professional learning and expertise. *Vocations and Learning, 1*(2), 87-103.

Gartmeier, M., Bauer, J., Gruber, H., & Heid, H. (2010). Workplace errors and negative knowledge in elder care nursing. *Human Resource Development International, 13*(1), 5-25.

Gielen, S., Peeters, E., Dochy, F., Onghena, P., & Struyven, K. (2010). Improving the effectiveness of peer feedback for learning. *Learning and instruction, 20*(4), 304-315.

Gilhooly, K. J. (1990). Cognitive psychology and medical diagnosis. *Applied cognitive psychology, 4*(4), 261-272.

Gill, J. B., Miller, D. D., Boucher, C. A., & Strauss, H. W. (1986). Clinical Decision Making. *Journal of Nuclear Medicine, 27*, 132-137.

Graber, M. (2005). Diagnostic errors in medicine: a case of neglect. *The Joint Commission Journal on Quality and Patient Safety, 31*(2), 106-113.

Graber, M. (2009). Educational strategies to reduce diagnostic error: can you teach this stuff? *Advances in health sciences education, 14*(1), 63-69.

Graber, M., Franklin, N., & Gordon, R. (2005). Diagnostic error in internal medicine. *Archives of internal medicine, 165*(13), 1493-1499.

Graber, M., Gordon, R., & Franklin, N. (2002). Reducing diagnostic errors in medicine: what's the goal? *Academic Medicine, 77*(10), 981-992.

Groen, G. J., & Patel, V. L. (1988). The relationship between comprehension and reasoning in medical expertise. In M. T. Chi, R. Glaser, & M. J. Farr (Eds.), *The nature of expertise* (pp. 287-310). Hillsdale, NJ: Erlbaum.

Große, C. S., & Renkl, A. (2007). Finding and fixing errors in worked examples: Can this foster learning outcomes? *Learning and instruction, 17*(6), 612-634.

Groves, M., O'rourke, P., & Alexander, H. (2003). Clinical reasoning: the relative contribution of identification, interpretation and hypothesis errors to misdiagnosis. *Medical teacher, 25*(6), 621-625.

Gruber, H., Mandl, H., & Renkl, A. (2000). Was lernen wir in Schule und Hochschule: Träges Wissen? In H. Mandl & J. Gerstenmeier (Eds.), *Die Kluft zwischen Wissen und Handeln : empirische und theoretische Lösungsansätze* (pp. 139-156). Goettingen: Hogrefe.

Hadie, S. N., & Yusoff, M. S. (2016). Assessing the validity of the cognitive load scale in a problem-based learning setting. *Journal of Taibah University Medical Sciences, 11*(3), 194-202.

Hammond, K. (1990). Intuitive and analytic cognition: Information models *Concise encyclopedia of information processing in systems and organizations* (pp. 306-312): Pergamon Press, England.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research, 77*(1), 81-112.

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*: Guilford Press.

Hebert, B. G., & Vorauer, J. D. (2003). Seeing through the screen: is evaluative feedback communicated more effectively in face-to-face or computer-mediated exchanges? *Computers in Human Behavior, 19*(1), 25-38.

Heitzmann, N., Fischer, F., Kühne-Eversmann, L., & Fischer, M. R. (2015). Enhancing diagnostic competence with self-explanation prompts and adaptable feedback. *Medical education, 49*(10), 993-1003.

Hersh, R. H. (2005). What does college teach? *Atlantic Monthly, 296*(4), 140-143.

Hesse, I., & Latzko, B. (2009). *Diagnostik für Lehrkräfte*. Obladen: Budrich.

Higgs, J. (2006). *The complexity of clinical reasoning: exploring the dimensions of clinical reasoning expertise as a situated, lived phenomenon*. Paper presented at the Faculty of Health Sciences, University of Sydney, Australia.

Higgs, J., Jones, M. A., Loftus, S., & Christiensen, N. (2008). *Clinical reasoning in the health professions*: Elsevier Health Sciences.

Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist, 42*(2), 99-107.

Hoogerheide, V., Deijkers, L., Loyens, S. M., Heijltjes, A., & van Gog, T. (2016). Gaining from explaining: Learning improves from explaining to fictitious others on video, not from writing to them. *Contemporary Educational Psychology, 44*, 95-106.

Hoogerheide, V., Loyens, S. M., Jadi, F., Vrins, A., & van Gog, T. (2015). Testing the model-observer similarity hypothesis with text-based worked examples. *Educational psychology*, 1-16.

Hoogerheide, V., Loyens, S. M., & Van Gog, T. (2014). Comparing the effects of worked examples and modeling examples on learning. *Computers in Human Behavior, 41*, 80-91.

Hoogerheide, V., Loyens, S. M., & Van Gog, T. (2016). Learning from video modeling examples: does gender matter? *Instructional Science, 44*(1), 69-86.

Hoogerheide, V., van Wermeskerken, M., Loyens, S. M., & van Gog, T. (2016). Learning from video modeling examples: Content kept equal, adults are more effective models than peers. *Learning and instruction, 44*, 22-30.

Ilgen, J. S., Humbert, A. J., Kuhn, G., Hansen, M. L., Norman, G. R., Eva, K. W., . . . Sherbino, J. (2012). Assessing diagnostic reasoning: a consensus statement summarizing theory, practice, and future needs. *Academic Emergency Medicine, 19*(12), 1454-1461.

Kahneman, D. (2011). *Thinking, fast and slow*. New York: Macmillan.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, *D.* Griffin & *D. Kahneman* (Eds.), *Heuristics* and *biases*: The *psychology* of *intuitive judgment* (pp.*49*-81*).* New York: Cambridge University Press

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the econometric society*, 263-291.

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review, 19*(4), 509-539.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist, 38*(1), 23-31.

Kapur, M. (2008). Productive failure. *Cognition and Instruction, 26*(3), 379-424.

Kassirer, J. P., & Gorry, G. A. (1978). Clinical problem solving: a behavioral analysis. *Annals of Internal Medicine, 89*(2), 245-255.

Kirschner, P. A., & Kirschner, F. (2012). Mental Effort. In N. M. Seel (Ed.), *Encyclopedia of the Sciences of Learning* (pp. 2182-2184). Boston: Springer.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*(2), 75-86.

Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive science, 12*(1), 1-48.

KMK, B. d. K. (2004). *Bildungsstandards im Fach Mathematik für den Primarbereich*. Neuwied: Luchterhand.

Kohn, L. T., Corrigan, J. M., & Donaldson, M. S. (2000). *Errors in health care: a leading cause of death and injury*. Washington, DC: National Academy Press.

Kollar, I., & Fischer, F. (2010). Peer assessment as collaborative learning: A cognitive perspective. *Learning and instruction, 20*(4), 344-348.

Kollar, I., Fischer, F., & Slotta, J. D. (2007). Internal and external scripts in computer-supported collaborative inquiry learning. *Learning and instruction, 17*(6), 708-721.

Kolodner, J. L. (2014). *Case-based reasoning*. San Mateo, CA: Morgan Kaufmann.

Kolodner, J. L., Cox, M., & González-Calero, P. (2005). Case-based reasoning-inspired approaches to education. *The Knowledge Engineering Review, 20*(03), 299-303.

Kolodner, J. L., Owensby, J. N., & Guzdial, M. (2004). Case-based learning aids. *Handbook of research on educational communications and technology, 2*, 829-861.

Kopp, V., Stark, R., & Fischer, M. R. (2008). Fostering diagnostic knowledge through computer-supported, case-based worked examples: effects of erroneous examples and feedback. *Medical education, 42*(8), 823-829.

Kopp, V., Stark, R., Kühne-Eversmann, L., & Fischer, M. R. (2009). Do worked examples foster medical students' diagnostic knowledge of hyperthyroidism? *Medical education, 43*(12), 1210-1217.

Kostons, D., Van Gog, T., & Paas, F. (2012). Training self-assessment and task-selection skills: A cognitive approach to improving self-regulated learning. *Learning and instruction, 22*(2), 121-132.

Krasne, S., & Stevens, C. D. (2010). Does basic science knowledge correlate with clinical reasoning in assessments of first-year medical students? *Procedia-Social and Behavioral Sciences, 2*(2), 1287-1294.

Kruglanski, A. W., & Gigerenzer, G. (2011). Intuitive and deliberate judgments are based on common principles. *Psychological review, 118*(1), 97-109.

Krych-Appelbaum, M., & Musial, J. (2007). Students' perception of value of interactive oral communication as part of writing course papers. *Journal of Instructional Psychology, 34*(3), 131-137.

Kuhn, D. (2002). What is scientific thinking and how does it develop? *The Wiley-Blackwell handbook of childhood cognitive development*. Malden, MA: Blackwell Publishers Ltd.

Kuhn, G. J. (2002). Diagnostic errors. *Academic Emergency Medicine, 9*(7), 740-750.

Lawson, A. E., Clark, B., Cramer-Meldrum, E., Falconer, K. A., Sequist, J. M., & Kwon, Y. J. (2000). Development of Scientific Reasoning in College Biology: Do Two Levels of General Hypothesis-Testing Skills Exist? *Journal of research in Science Teaching, 37*(1), 81-101.

Lawson, A. E., & Daniel, E. S. (2011). Inferences of clinical diagnostic reasoning and diagnostic error. *Journal of biomedical informatics, 44*(3), 402-412.

Leahy, W., & Sweller, J. (2011). Cognitive load theory, modality of presentation and the transient information effect. *Applied cognitive psychology, 25*(6), 943-951.

Leake, D. B. (1996). *Case-Based Reasoning: Experiences, lessons and future directions*: MIT press.

Leppink, J., Paas, F., Van der Vleuten, C. P., Van Gog, T., & Van Merriënboer, J. J. (2013). Development of an instrument for measuring different types of cognitive load. *Behavior research methods, 45*(4), 1058-1072.

Leppink, J., Paas, F., Van Gog, T., van Der Vleuten, C. P., & Van Merrienboer, J. J. (2014). Effects of pairs of problems and examples on task performance and different types of cognitive load. *Learning and instruction, 30*, 32-42.

Li, L., Liu, X., & Steckelberg, A. L. (2010). Assessor or assessee: How student learning improves by giving and receiving peer feedback. *British Journal of Educational Technology, 41*(3), 525-536.

Lieberman, M. D., Gaunt, R., Gilbert, D. T., & Trope, Y. (2002). Reflection and reflexion: A social cognitive neuroscience approach to attributional inference. *Advances in experimental social psychology, 34*, 199-249.

Lin, S. S., Liu, E. Z.-F., & Yuan, S.-M. (2001). Web-based peer assessment: feedback for students with various thinking-styles. *Journal of Computer Assisted Learning, 17*(4), 420-432.

Litosseliti, L., Marttunen, M., Laurinen, L., & Salminen, T. (2005). Computer-based and Face-to-face Collaborative Argumentation in Secondary Schools in England and Finland. *Education, Communication & Information, 5*(2), 131-146.

Liu, N.-F., & Carless, D. (2006). Peer feedback: the learning element of peer assessment. *Teaching in Higher education, 11*(3), 279-290.

Lombard, M., Ditton, T. B., & Weinstein, L. (2009). *Temple presence inventory (TPI) questionnaire items and factor structures.* Paper presented at the the 12th Annual International Workshop on Presence in Los Angeles, CA.

Lundstrom, K., & Baker, W. (2009). To give is better than to receive: The benefits of peer review to the reviewer's own writing. *Journal of second language writing, 18*(1), 30-43.

Mamede, S., van Gog, T., Moura, A. S., de Faria, R., Peixoto, J. M., Rikers, R. M., & Schmidt, H. G. (2012). Reflection as a strategy to foster medical students' acquisition of diagnostic competence. *Medical education, 46*(5), 464-472.

Mamede, S., Van Gog, T., Van den Berge, K., Rikers, R. M., Van Saase, J. L., Van Guldener, C., & Schmidt, H. G. (2010). Effect of availability bias and reflective reasoning on diagnostic accuracy among internal medicine residents. *Jama, 304*(11), 1198-1203.

Mayer, R. E. (2002). Multimedia learning. *Psychology of learning and motivation, 41*, 85-139.

McGaghie, W. C., Issenberg, S. B., Petrusa, E. R., & Scalese, R. J. (2010). A critical review of simulation-based medical education research: 2003–2009. *Medical education, 44*(1), 50-63.

Min, H.-T. (2005). Training students to become successful peer reviewers. *System, 33*(2), 293-308.

Monnier, P., Bédard, M.-J., Gagnon, R., & Charlin, B. (2011). The relationship between script concordance test scores in an obste rics-gynecology rotation and global performance assessments in the curriculum. *International journal of medical education, 2*, 3-6.

Morris, B. J., Masnick, A. M., Zimmerman, C., & Croker, S. (2012). The emergence of scientific reasoning. In H. Kloos, B. J. Morris, & J. Amaral (Eds.), *Trends in Cognitive Development* (pp. 61-83). Rijcka, Croatia: InTech Publisher.

Moulton, C.-a., Regehr, G., Mylopoulos, M., & MacRae, H. M. (2007). Slowing down when you should: a new model of expert judgment. *Academic medicine, 82*(10), S109-S116.

Narciss, S. (2008). Feedback strategies for interactive learning tasks. *Handbook of research on educational communications and technology, 3*, 125-144.

Nelson, K., & Gruendel, J. (1986). *Event knowledge: Structure and function in development*. Hillsdale NJ: Lawrence Erlbaum Associates.

Nendaz, M. R., & Bordage, G. (2002). Promoting diagnostic problem representation. *Medical education, 36*(8), 760-766.

Neufeld, V., Norman, G. R., Feightner, J., & Barrows, H. (1981). Clinical problem-solving by medical students: a cross-sectional and longitudinal analysis. *Medical education, 15*(5), 315-322.

Nicol, D., Thomson, A., & Breslin, C. (2014). Rethinking feedback practices in higher education: a peer review perspective. *Assessment & Evaluation in Higher Education, 39*(1), 102-122.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological review, 84*(3), 231-259.

Nokes-Malach, T. J., VanLehn, K., Belenky, D. M., Lichtenstein, M., & Cox, G. (2013). Coordinating principles and examples through analogy and self-explanation. *European Journal of Psychology of Education, 28*(4), 1237-1263.

Norman, G. R. (2005). Research in clinical reasoning: past history and current trends. *Medical education, 39*(4), 418-427.

Norman, G. R., & Eva, K. W. (2010). Diagnostic error and clinical reasoning. *Medical education, 44*(1), 94-100.

Nyssen, A.-S., & Blavier, A. (2006). Error detection: a study in anaesthesia. *Ergonomics, 49*(5-6), 517-525.

Oser, F., & Spychiger, M. (2005). *Lernen ist schmerzhaft: Zur Theorie des negativen Wissens und zur Praxis der Fehlerkultur*. Weinheim: Beltz.

Paas, F., & Kalyuga, S. (2005). *Cognitive measurements to design effective learning environments.* Paper presented at the International Workshop and Mini-conference on Extending Cognitive Load Theory and Instructional Design to the Development of Expert Performance, Heerlen, Netherlands.

Paas, F., & Sweller, J. (2014). Implications of cognitive load theory for multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 27-42). Cambridge, MA: Cambridge University Press.

Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist, 38*(1), 63-71.

Paas, F., & Van Merriënboer, J. J. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *Journal of educational psychology, 86*(1), 122-133.

Page, G., Bordage, G., & Allen, T. (1995). Developing key-feature problems and examinations to assess clinical decision-making skills. *Academic Medicine, 70*(3), 194-201.

Paris, S. G., Lipson, M. Y., & Wixson, K. K. (1983). Becoming a strategic reader. *Contemporary Educational Psychology, 8*(3), 293-316.

Patchan, M. M., & Schunn, C. D. (2015). Understanding the benefits of providing peer feedback: how students respond to peers' texts of varying quality. *Instructional Science, 43*(5), 591-614.

Patel, V. L., Arocha, J. F., & Zhang, J. (2005). Thinking and reasoning in medicine. *The Cambridge handbook of thinking and reasoning, 14*, 727-750.

Patel, V. L., & Cohen, T. (2008). New perspectives on error in critical care. *Current opinion in critical care, 14*(4), 456-459.

Patel, V. L., Cohen, T., Murarka, T., Olsen, J., Kagita, S., Myneni, S., . . . Ghaemmaghami, V. (2011). Recovery at the edge of error: debunking the myth of the infallible expert. *Journal of Biomedical Informatics, 44*(3), 413-424.

Patel, V. L., & Groen, G. J. (1991). The general and specific nature of medical expertise: A critical look. *Toward a general theory of expertise: Prospects and limits*, 93-125.

Petrusa, E. R., Blackwell, T. A., & Ainsworth, M. A. (1990). Reliability and validity of an objective structured clinical examination for assessing the clinical performance of residents. *Archives of internal medicine, 150*(3), 573-577.

Pintrich, P. R. (2002). The role of metacognitive knowledge in learning, teaching, and assessing. *Theory into practice, 41*(4), 219-225.

Popper, K. R. (1959). *The logic of scientific discovery*. London: Hutchinson.

Prins, F. J., Sluijsmans, D. M., & Kirschner, P. A. (2006). Feedback for general practitioners in training: quality, styles, and preferences. *Advances in health sciences education, 11*(3), 289-303.

Quinn, H., Schweingruber, H., & Keller, T. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*: National Academies Press.

Rahimi, M. (2013). Is training student reviewers worth its while? A study of how training influences the quality of students' feedback and writing. *Language Teaching Research, 17*(1), 67-89.

Rajkomar, A., & Dhaliwal, G. (2011). Improving diagnostic reasoning to improve patient safety. *Permanente J, 15*, 68-73.

Reason, J. T., & Zapf, D. (1994). *Errors, error detection, and error recovery*. Hillsdale NJ: Lawrence Erlbaum Associates.

Renkl, A. (2002). Worked-out examples: Instructional explanations support learning by self-explanations. *Learning and instruction, 12*(5), 529-556.

Renkl, A. (2014). Toward an instructionally oriented theory of example-based learning. *Cognitive science, 38*(1), 1-37.

Renkl, A., & Atkinson, R. K. (2003). Structuring the transition from example study to problem solving in cognitive skill acquisition: A cognitive load perspective. *Educational Psychologist, 38*(1), 15-22.

Renkl, A., Stark, R., Gruber, H., & Mandl, H. (1998). Learning from worked-out examples: The effects of example variability and elicited self-explanations. *Contemporary Educational Psychology, 23*(1), 90-108.

Reynolds, J., & Russell, V. (2008). Can you hear us now?: A comparison of peer review quality when students give audio versus written feedback. *The WAC Journal, 19*(1), 29-44.

Rice, R. E. (1993). Media appropriateness. *Human communication research, 19*(4), 451-484.

Riesbeck, C. K., & Schank, R. C. (2013). *Inside case-based reasoning*: Psychology Press.

Rikers, R. M., Schmidt, H. G., & Boshuizen, H. P. (2000). Knowledge encapsulation and the intermediate effect. *Contemporary Educational Psychology, 25*(2), 150-166.

Rocke, D., & Lee, W. T. (2013). Medical Errors: Teachable Moments in Doing the Right Thing. *Journal of Graduate Medical Education, 5*(4), 550-552. doi:doi:10.4300/JGME-D-13-00110.1

Rouhi, A., & Azizian, E. (2013). Peer review: Is Giving Corrective Feedback Better than Receiving it in L2 Writing? *Procedia-Social and Behavioral Sciences, 93*, 1349-1354.

Rourke, L., Anderson, T., Garrison, D. R., & Archer, W. (2007). Assessing social presence in asynchronous text-based computer conferencing. *International Journal of E-Learning & Distance Education, 14*(2), 50-71.

Rummel, N., & Spada, H. (2005). Learning to collaborate: An instructional approach to promoting collaborative problem solving in computer-mediated settings. *The Journal of the Learning Sciences, 14*(2), 201-241.

Sackett, D. L. (1997). *Evidence-based medicine.* Paper presented at the Paper presented at the Seminars in perinatology 21.

Schank, R. C. (1999). *Dynamic memory revisited*. Cambridge: Cambridge University Press.

Schank, R. C., & Abelson, R. (1977). *Scripts, goals, plans, and understanding*: Hillsdale, NJ: Erlbaum.

Schauble, L. (1996). The development of scientific reasoning in knowledge-rich contexts. *Developmental Psychology, 32*(1), 102-119.

Scheiter, K., Gerjets, P., Huk, T., Imhof, B., & Kammerer, Y. (2009). The effects of realism in learning with dynamic visualizations. *Learning and instruction, 19*(6), 481-494.

Schiff, G. D., Hasan, O., Kim, S., Abrams, R., Cosby, K., Lambert, B. L., . . . Krosnjar, N. (2009). Diagnostic error in medicine: analysis of 583 physician-reported errors. *Archives of internal medicine, 169*(20), 1881-1887.

Schmidmaier, R., Eiber, S., Ebersbach, R., Schiller, M., Hege, I., Holzer, M., & Fischer, M. R. (2013). Learning the facts in medical school is not enough: which factors predict successful application of procedural knowledge in a laboratory setting? *BMC medical education, 13*(1), 13-28.

Schmidt, H. G., & Boshuizen, H. P. (1992). Encapsulation of biomedical knowledge *Advanced models of cognition for medical training and practice* (pp. 265-282): Springer.

Schmidt, H. G., & Boshuizen, H. P. (1993). On acquiring expertise in medicine. *Educational Psychology Review, 5*(3), 205-221.

Schmidt, H. G., Loyens, S. M., Van Gog, T., & Paas, F. (2007). Problem-based learning is compatible with human cognitive architecture: Commentary on Kirschner, Sweller, and Clark (2006). *Educational Psychologist, 42*(2), 91-97.

Schmidt, H. G., & Mamede, S. (2015). How to improve the teaching of clinical reasoning: a narrative review and a proposal. *Medical education, 49*(10), 961-973.

Schmidt, H. G., Norman, G. R., & Boshuizen, H. P. (1990). A cognitive perspective on medical expertise: theory and implication [published erratum appears in Acad Med 1992 Apr; 67 (4): 287]. *Academic Medicine, 65*(10), 611-621.

Schmidt, H. G., & Rikers, R. M. (2007). How expertise develops in medicine: knowledge encapsulation and illness script formation. *Medical education, 41*(12), 1133-1139.

Schnotz, W., & Kürschner, C. (2007). A reconsideration of cognitive load theory. *Educational Psychology Review, 19*(4), 469-508.

Schunk, D. H. (1987). Peer models and children’s behavioral change. *Review of educational research, 57*(2), 149-174.

Schunk, D. H., Hanson, A. R., & Cox, P. D. (1987). Peer-model attributes and children's achievement behaviors. *Journal of educational psychology, 79*(1), 54-61.

Schunn, C. D., & Anderson, J. R. (1999). The generality/specificity of expertise in scientific reasoning. *Cognitive science, 23*(3), 337-370.

Sherbino, J., Dore, K. L., Siu, E., & Norman, G. R. (2011). The effectiveness of cognitive forcing strategies to decrease diagnostic error: an exploratory study. *Teaching and learning in medicine, 23*(1), 78-84.

Sherbino, J., Kulasegaram, K., Howey, E., & Norman, G. R. (2014). Ineffectiveness of cognitive forcing strategies to reduce biases in diagnostic reasoning: a controlled trial. *Canadian journal of emergency medicine, 16*(01), 34-40.

Shulman, L. S. (1992). Toward a pedagogy of cases. *Case methods in teacher education*, 1-30.

Siegler, R. S. (2002). Microgenetic studies of self-explanation. In N. Granott & J. Parziale (Eds.), *Microdevelopment: Transition processes in development and learning* (pp. 31-58). Cambridge: Cambridge University Press.

Simmons, B. (2010). Clinical reasoning: concept analysis. *Journal of Advanced Nursing, 66*(5), 1151-1158.

Singh, H., Meyer, A. N., & Thomas, E. J. (2014). The frequency of diagnostic errors in outpatient care: estimations from three large observational studies involving US adult populations. *BMJ quality & safety*, 727-731.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological bulletin, 119*(1), 3.

Sluijsmans, D. M., Brand-Gruwel, S., & van Merriënboer, J. J. (2002). Peer assessment training in teacher education: Effects on performance and perceptions. *Assessment & Evaluation in Higher Education, 27*(5), 443-454.

Stark, R., Gruber, H., Mandl, H., & Hinkofer, L. (2001). Wege zur Optimierung eines beispielbasierten Instruktionsansatzes: Der Einfluss multipler Perspektiven und instruktionaler Erklärungen auf den Erwerb von Handlungskompetenz. *Unterrichtswissenschaft, 29*(1), 26-40.

Stark, R., Kopp, V., & Fischer, M. R. (2011). Case-based learning with worked examples in complex domains: Two experimental studies in undergraduate medical education. *Learning and instruction, 21*(1), 22-33.

Stark, R., Mandl, H., Gruber, H., & Renkl, A. (1999). Instructional means to overcome transfer problems in the domain of economics: Empirical studies. *International Journal of Educational Research, 31*(7), 591-609.

Stegmann, K., Pilz, F., Siebeck, M., & Fischer, F. (2012). Vicarious learning during simulations: is it more effective than hands-on training? *Medical education, 46*(10), 1001-1008.

Stokes, D. E. (1997). *Pasteur's quadrant: Basic science and technological innovation*. Washington, DC: The Brookings Institution.

Strijbos, J.-W., Narciss, S., & Dünnebier, K. (2010). Peer feedback content and sender's competence level in academic writing revision tasks: Are they critical for feedback perceptions and efficiency? *Learning and instruction, 20*(4), 291-303.

Strijbos, J.-W., & Sluijsmans, D. (2010). Unravelling peer assessment: Methodological, functional, and conceptual developments. *Learning and instruction, 20*(4), 265-269.

Strijbos, J.-W., Van Goozen, B., & Prins, F. (2012). *Developing a coding scheme for analysing peer feedback messages.* Paper presented at the EARLI SIG 1: Assessment and Evaluation Conference.

Strobel, C., Heitzmann, N., Strijbos, J.-W., Kollar, I., & Fischer, M. R. (2016). *Watching People Fail: Improving Diagnostic Competence by Providing Peer Feedback on Erroneous Diagnoses.* . Paper presented at the AERA 2016, Washington.

Sweller, J. (2003). Evolution of human cognitive architecture. In B. Ross (Ed.), *The psychology of learning and motivation* (Vol. 43, pp. 215–266). San Diego, CA.

Sweller, J. (2006). The worked example effect and human cognition. *Learning and instruction, 16*(2), 165-169.

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review, 22*(2), 123-138.

Sweller, J., & Cooper, G. A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction, 2*(1), 59-89.

Sweller, J., Van Merrienboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251-296.

Thistlethwaite, J. E., Davies, D., Ekeocha, S., Kidd, J. M., MacDougall, C., Matthews, P., . . . Clay, D. (2012). The effectiveness of case-based learning in health professional education. A BEME systematic review: BEME Guide No. 23. *Medical teacher, 34*(6), 421-444.

Tomic, E. R., Martins, M. A., Lotufo, P. A., & Benseñor, I. M. (2005). Progress testing: evaluation of four years of application in the school of medicine, University of São Paulo. *Clinics, 60*(5), 389-396.

Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of educational research, 68*(3), 249-276.

Topping, K. (2005). Trends in peer learning. *Educational psychology, 25*(6), 631-645.

Topping, K. (2009). Peer assessment. *Theory into practice, 48*(1), 20-27.

Topping, K. (2010). Methodological quandaries in studying process and outcomes in peer assessment. *Learning and instruction, 20*(4), 339-343.

Trilling, B., & Fadel, C. (2009). *21st century skills: Learning for life in our times*. San Francisco, CA: John Wiley & Sons.

Trowbridge, R. L. (2008). Twelve tips for teaching avoidance of diagnostic errors. *Medical teacher, 30*(5), 496-500.

Valmont, W. J. (1995). *Creating videos for school use*. Boston: Allyn & Bacon.

Van Den Haak, M., De Jong, M., & Jan Schellens, P. (2003). Retrospective vs. concurrent think-aloud protocols: testing the usability of an online library catalogue. *Behaviour & information technology, 22*(5), 339-351.

Van Gog, T. (2015). Learning from erroneous examples in medical education. *Medical education, 49*(2), 142-144.

Van Gog, T., Paas, F., & Van Merriënboer, J. J. (2004). Process-oriented worked examples: Improving transfer performance through enhanced understanding. *Instructional Science, 32*(1-2), 83-98.

Van Gog, T., & Rummel, N. (2010). Example-based learning: Integrating cognitive and social-cognitive research perspectives. *Educational Psychology Review, 22*(2), 155-174.

Van Gog, T., Verveer, I., & Verveer, L. (2014). Learning from video modeling examples: Effects of seeing the human model's face. *Computers & Education, 72*, 323-327.

Van Zundert, M., Sluijsmans, D., & Van Merriënboer, J. (2010). Effective peer assessment processes: Research findings and future directions. *Learning and instruction, 20*(4), 270-279.

VanLehn, K. (1996). Cognitive skill acquisition. *Annual review of psychology, 47*(1), 513-539.

VanLehn, K. (1999). Rule-learning events in the acquisition of a complex skill: An evaluation of CASCADE. *The Journal of the Learning Sciences, 8*(1), 71-125.

Veenman, M. V., Van Hout-Wolters, B. H., & Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations. *Metacognition and learning, 1*(1), 3-14.

Vollmeyer, R., & Rheinberg, F. (2005). A surprising effect of feedback on learning. *Learning and instruction, 15*(6), 589-602.

Vygotsky, L. (1987). *Zone of proximal development*. Cambridge, MA: Harvard University Press.

Wegerif, R. (1998). The social dimension of asynchronous learning networks. *Journal of asynchronous learning networks, 2*(1), 34-49.

Weinberger, A., Ertl, B., Fischer, F., & Mandl, H. (2005). Epistemic and social scripts in computer–supported collaborative learning. *Instructional Science, 33*(1), 1-30.

Weinberger, A., Stegmann, K., Fischer, F., & Mandl, H. (2007). Scripting argumentative knowledge construction in computer-supported learning environments *Scripting computer-supported collaborative learning* (pp. 191-211): Springer.

Wilkinson, W. E., Cauble, L. A., & Patel, V. L. (2011). Error detection and recovery in dialysis nursing. *Journal of patient safety, 7*(4), 213-223.

Williams, B. (2005). Case based learning—a review of the literature: is there scope for this educational paradigm in prehospital education? *Emergency Medicine Journal, 22*(8), 577-581.

Wimmers, P. F., Schmidt, H. G., & Splinter, T. A. (2006). Influence of clerkship experiences on clinical competence. *Medical education*, *40*(5), 450-458.

Wiswell, J., Tsao, K., Bellolio, M. F., Hess, E. P., & Cabrera, D. (2013). "Sick" or "not-sick": accuracy of System 1 diagnostic reasoning for the prediction of disposition and acuity in patients presenting to an academic ED. *The American journal of emergency medicine, 31*(10), 1448-1452.

Woods, N. N. (2007). Science is fundamental: the role of biomedical knowledge in clinical reasoning. *Medical education*, *41*(12), 1173-1177.

Zapf, D., & Reason, J. T. (1994). Introduction: Human errors and error handling. *Applied Psychology*, *43*(4), 427-432.

Zayyan, M. (2011). Objective structured clinical examination: the assessment of choice. *Oman Med J, 26*(4), 219-222.

Zimmerman, B. J., & Kitsantas, A. (2002). Acquiring writing revision and self-regulatory skill through observation and emulation. *Journal of educational psychology*, *94*(4), 660-668.

Ziv, A., Ben-David, S., & Ziv, M. (2005). Simulation based medical education: an opportunity to learn from errors. *Medical teacher*, *27*(3), 193-199.

Zwaan, L., & Singh, H. (2015). The challenges in defining and measuring diagnostic error. *Diagnosis, 2*(2), 97-103.

List of Tables

List of Figures

Appendix

APPENDIX A

Feedback Guidelines

Bitte notiere dir hier Notizen zu folgenden Punkten:

Was war gut am diagnostischen Prozess des Medizinstudenten im Video?

Was würdest du am Diagnoseprozess des Medizinstudenten im Video kritisieren?

Hast du Fehler im Diagnoseprozess bemerkt? Wenn ja, wo war(en) diese?

Was würdest du gegebenenfalls anders machen?

APPENDIX B

Diagnostic Competences Tests

Example Knowledge Test –Conceptual Knowledge

MC1

Ein 68 jähriger Patient leidet seit 15 Tagen unter zunehmender Dyspnoe.

Vorerkrankungen: langjährige arterielle Hypertonie und Zigarettenabusus.

Untersuchungsbefund: Deutliche Fußrücken- und Unterschenkelödeme und vergrößerte Leber.

Lunge: feuchte inspiratorische Rasselgeräusche beidseits. Blutdruck 164/92 mmHg;

Labor: Serumelektrolyte und -lipide normal. Serumkreatinin 2,2 mg/dl.

Welches ist die wahrscheinlichste Diagnose?

A: Akutes Nierenversagen
B: Dekompensierte Herzinsuffizienz
C: Infektexazerbierte Bronchitis
D: Dekompensierte Nierenarterienstenose
E: Nephrotisches Syndrom

MC2

Eine 40-jährige Raucherin stellt sich mit einem geschwollenen linken Bein und plötzlichem Thoraxschmerz mit Husten in der Notaufnahme vor.

Welche Diagnose ist am wahrscheinlichsten?

A: Dekompensierte Herzinsuffizienz
B: Infektexazerbierte COPD
C: Lungenarterienembolie
D: Embolie der Arteria femoralis communis
E: Pneumothorax

Example Knowledge Test – Strategic Knowledge

KF 1-1

In die Notaufnahme wird ein 56 jähriger Patient von Zuhause mit Notarztbegleitung transportiert. Er berichtet über seit 3 Stunden bestehende und zunehmende Atemnot und Druck auf der Brust.

Im Rahmen Ihrer differentialdiagnostischen Überlegungen möchten Sie ein akutes Koronarsyndrom ausschließen.

Welche Schritte führen Sie hierfür als erstes durch? Nennen Sie mind. 2.

KF 1-2

Sie denken bei Ihrem Patienten auch an muskuloskelettale Beschwerden

Frage 2: Welche Patienten haben hierfür ein besonderes Risiko? Nennen Sie mind. 2.

KF 1-3

Sie wollen außerdem ausschließen, dass bei dem Patienten eine Lungenembolie vorliegt.

Welche weiteren diagnostischen Schritte (mind. 2) sowie akute und langfristige therapeutische Maßnahmen (mind. 2) führen Sie nun durch?

Example Knowledge *Test - Conditional Knowledge*

PL 1-1

Herr Meier bekommt seit einigen Stunden immer schlechter Luft und berichtet zusätzlich über ein Stechen auf der Brust; die Beschwerden nahmen nun so zu, dass er den Notarzt gerufen hat. Sie veranlassen eine Röntgen Untersuchung des Thorax.

Welche Verdachtsdiagnosen können Sie bei diesem Patienten mit einer Röntgen Untersuchung des Thorax ausschließen? Bitte begründen Sie Ihre Aussage

PL 1-2

Warum ist die Frage nach dem Nikotinkonsum für beides von Bedeutung?

Example Error Detection Test

ED1

Hier ein Beispiel einer Situation aus dem klinischen Alltag mit einigen Fehlern. Bitte markieren Sie, wenn Ihnen ein Fehler in der Diagnose des Arztes auffällt **die Zahl vor der Stelle** durch Unterstreichen:

Beispiel: Die Erde (1) hat Kontinente und ***(2)*** *ist eine Scheibe*

Frau Anton, 45 jährige Patientin mit metastasiertem Lungenkarzinom, wird mit Chemotherapie behandelt. Vor einem Tag hat sie Luftnot entwickelt, die unter Belastung stärker wird. Vitalparameter: Tachykard – Herzfrequenz: 111/min, Sauerstoffsättigung bei 4 l Sauerstoff bei 91%, Temperatur: 37,3°C EKG: Sinustachykardie BGA: pO2: 54 mmHg, pCO2: 26 mmHg, pH: 7,50 Labor: Anämie mit Hb 8,5 g/dl, LDH-Erhöhung und CRP-Erhöhung auf 6 mg/dl. D-Dimere deutlich erhöht. Bei der körperlichen Untersuchung ergaben die Auskultation der Lunge sowie des Herzens keine Auffälligkeiten. Am Bein wurden beidseits deutlich eindrückbare Unterschenkelödeme festgestellt. Die Jugularvene war prominent. Die diensthabende Ärztin hält eine (1) Pneumonie für die wahrscheinlichste Diagnose, da (2) Entzündungszeichen sowie (3) subfebrile Temperaturen in Verbindung mit der (4) plötzlich aufgetretenen Atemnot dafürsprechen. Auch der (5) unauffällige Auskultationsbefund der Lunge passt hierzu. Eine Lungenembolie hält sie für (6) weniger wahrscheinlich, da zwar die (7) erhöhten D-Dimere bei dieser Patientin einen Hinweis darauf geben, aber keine für eine Lungenembolie typische (8) ST-Hebung und ein (9) S1-Q3 Typ im EKG zu sehen waren. Bei der körperlichen Untersuchung war auch kein (10) prominenter Jugularvenenpuls zu fühlen oder eine (11) einseitige Schwellung oder (12) Schmerzen am Bein zu sehen. Um eine Pneumonie nachzuweisen, veranlasst die Ärztin (13) ein CT sowie eine (14) Bestimmung von Troponin und CK. Durch das CT kann so das (15) Lungenparenchym bez. entzündlicher oder tumoröser Infiltrationen beurteilt werden, um so eine Pneumonie zu bestätigen oder auszuschließen.

APPENDIX C

Example: Mental Effort

Bitte kreuzen Sie die für Sie zutreffende Antwortalternative an (von „sehr leicht“ bis „sehr schwer“)

	sehr leicht	**leicht**	**eher leicht**	**weder leicht noch schwer**	**eher schwer**	**schwer**	**sehr schwer**
Wie leicht oder schwer finden Sie das Thema „Dyspnoe“?	☐	☐	☐	☐	☐	☐	☐
Wie leicht oder schwer fällt es Ihnen, mit dieser Lernumgebung zu arbeiten?	☐	☐	☐	☐	☐	☐	☐

APPENDIX D

Feedback Coding Scheme

Level	Erklärung	Beispiel	Kürzel
General Feedback	Allgemeinwissen, allgemeines Fakten- oder Lehrbuchwissen, unspezifische Bewertungen, Feedback ohne jegliche Angabe von Gründen, Beschränkungen, Erklärungen etc. -> *Beschreibungen* (statt Begründungen)	*“Hat es gut gemacht.”* *"Ist strukturiert vorgegangen."* *"Troponin Erhöhung spricht für Myokardinfarkt."*	G_F
Elaborated Feedback	Feedback auf *spezifische* Fakten unter Angabe von Gründen, Beschränkungen, Alternativen, Erklärungen o.ä. Prozedurales Wissen, *konkretes* Fehlerwissen Hinweiswörter können sein: obwohl, statt dessen, aber, weil, trotz, sondern	*„Lungenembolie ist es nicht, weil Troponin erhöht, was eher für Myokardinfarkt spricht."*	E_F
Off-Topic	Feedback auf all die Dinge, die nichts mit den Krankheitsbildern oder der Diagnostik im Video zu tun haben, z.B.: Feedback auf die Durchführung der Anamnese und körperlichen Untersuchung oder hinsichtlich der (weiteren) Therapie, da *nicht* auf die Diagnostik bezogen; Feedback bezüglich der Patientenvorstellung	*"Gutes Eingehen auf Patient, gute Fragen gestellt."* *"Sie hatte die Haare offen."*	Off
Fehler-erkennung	Ja / Nein: Fehler erkannt, wenn *Premature Closure* (= vorzeitiges Abschließen einer Diagnose bzw. Festlegen auf eine Diagnose) erkannt	*„Evtl weniger auf eine möglicherweise sichere Verdachtsdiagnose versteifen.“* *„Nicht abgewartet.“* *“Keine weiteren Differenzialdiagnosen in Betracht gezogen.”* *“Weitere Diagnostik nicht beachtet, weitere Tests nötig um eine Arbeitsdiagnose zu verfolgen.”*	ERR

APPENDIX E

Feedback Training

Hier ein Beispiel, wie gutes Feedback aussehen könnte. Bitte lies es dir ganz kurz durch.

Herr Hartl, 43-jähriger Patient, der angibt sich seit einer Erkältung vor 5 Wochen müde und erschöpft zu fühlen. Er bekäme Atemnot schon nach dem Steigen von ca. 10 Stufen. Weder Angina noch Vorerkrankungen. Er ist Raucher seitdem er 16 ist (15 Packyears).

AZ akut reduziert
RR 170/130mmHG rechts, 173/126mmHG links, P 90/min, Größe 1,84m, Gewicht 95kg
Corrhythmisch, keine pathologischen Geräusche.

Pulmo mit leichter Spastik und basalen Rasselgeräuschen
Orthopnoe
Abdomen unauffällig
Diskrete Knöchelödeme bds.

Ein Famulant stellt aufgrund der unauffälligen Befunde die Diagnose auf, dass Herr Hartl durch das Rauchen eine leichte Atemnot und die Spastik entwickelt hat, ihm aber sonst nichts Gravierendes fehlt. Er rät dem Patienten zu mehr Sport, gesünderer Ernährung und mit dem Rauchen aufzuhören. Er veranlasst zur Sicherheit noch ein Röntgenbild der Lunge und plant bereits die Einleitung einer bronchodilatatorischen Therapie

Gutes Feedback könnte so aussehen:

Der Famulant hat durchaus richtige Teilbefunde erhoben und richtige Konsequenzen gezogen **(Guter Punkt),** da die Befunde größtenteils unauffällig sind **(Erklärung)**

Die Diagnose, dass die Symptome durch das langjährige Rauchen ausgelöst werden, ist als endgültige Entscheidung aber zu früh **(schlechter Punkt),** da andere Krankheiten wie COPD oder Asthma bei solchen Patienten durchaus eine hohe Prävalenz haben können **(Erklärung).**

Er sollte aber die Beinödeme, die Orthopnoe und die Blutdruckwerte berücksichtigen **(was könnte man besser machen),** da Beinödeme und Atemnot bei Belastung Anzeichen für eine Herzerkrankung wie die Herzinsuffizienz sein können **(Erklärung).**

APPENDIX F

Model Expert Feedback

Das folgende, sehr elaborierte Feedback hat ein Mitstudent auf die Diagnose gegeben. Bitte lies es dir einmal sorgfältig durch.

Was war gut? Der Peer liefert relativ viele und z.T. auch seltene Differentialdiagnosen. Dies lässt breite differential-diagnostische Überlegungen zu.

Was war schlecht? Der Famulus äußert im ersten Abschnitt nicht eine Differentialdiagnose sondern formuliert mit einiger Gewissheit: " Ich gehe angesichts ....von... aus....", somit priorisiert er eine mögliche Diagnose zu stark. Die Bewertung, dass eine LE weniger wahrscheinlich sei, ist angesichts der uneindeutigen Befunde und ausstehender Diagnostik zu früh. Die LE ist das Chamäleon der Inneren..... Ebenso scheint die Schlussfolgerung der Herzinsuffizienz nach MI durch die erhöhte AF und erniedrigtem PO2 nachvollziehbar und doch unvollständig, da in der körperlichen Untersuchung die typischen Zeichen (Rasselgeräusche usw., gestaute Halsvenen) fehlen. Gleichzeitig wird das geschwollene Bein nicht erwähnt, was zum einen den MI unwahrscheinlicher macht und gleichzeitig in der Argumentation für die LE fehlt. Ebenso ist der weitere diagnostische Schritt übereilt. Zuerst sollte man versuchen nicht-invasiv die Diagnose zu stellen oder zu präzisieren. Außerdem ist die vorgeschlagene Diagnostik zu limitiert, um Anhaltspunkte für oder gegen die anderen Differentialdiagnosen zu bekommen.

Wo war der Fehler? Der Peer hat z.T. Fakten außer Acht gelassen und hat sich scheinbar zu schnell festgelegt. Die weitere Diagnostik ist zu eingeschränkt.

Was könnte man besser machen? Der Student könnte eine offenere Darstellung der Möglichkeiten wählen. Dies würde das differential-diagnostische Denken eher befördern, als einschränken. Gleichzeitig könnte er alle relevanten Fakten, die für oder gegen eine Differentialdiagnose sprechen nennen, um größere Argumentationsketten aufzubauen und eine bessere Abschätzung der Wahrscheinlichkeiten zu ermöglichen.

APPENDIX G

Guidance Sheet on Diagnosing Dyspnea

Diagnose	COPD (obstruktion)	Pneumonie (Diffusion)	Lungenembolie (Perfusion)	Kardiale Dekompensation (Perfusion/Diffusion)	Akuter Myokardinfarkt (Perfusion)	Muskoskeletal (Ventilation)
Anamnese	• Raucher • Belastungsdyspnoe • Exazerbation bei Infekten • Chronisch-progredienter Verlauf • Staubexposition	• Neu aufgetretener Husten • reduzierter Allgemeinzustand • kein bis purulenter/blutiger Auswurf • Neu aufgetretene Somnolenz • Alkoholabusus • Immunsuppression • Atemabhängiger stechender Thoraxschmerz	• Akute Dyspnoe • Atemabhängiger Thoraxschmerz • Vorausgegangene OP • bekannte Thrombophile Diathese (Immobilisation, Malignome, Schwangerschaft, Hormontherapie) • tiefe Beinvenenthrombose • Vernichtungsschmerz • Herzstolpern • Synkope	• Kurzatmigkeit in Ruhe oder Belastung • Rasche Ermüdbarkeit • Knöchelödeme • Appetitlosigkeit • Herzstolpern • Tachykardie • Panik • Synkope	• Risikofaktoren • Ängstlich • Übelkeit • Schweißausbrüche • Kollaps • Bewusstlosigkeit • Thoraxschmerz	• Kurzatmigkeit unter Belastung, • Rasche Ermüdbarkeit, • Trauma • Malignomen

Vitalparameter	• O2-Sättigung erniedrigt, • AF kann erhöht sein • RR erhöht • Puls erhöht • Temp. normal	• O2-Sättigung erniedrigt, • AF erhöht • RR erniedrigt • Puls erhöht • Temp. erhöht	• O2-Sättigung erniedrigt, • AF erhöht • RR erniedrigt • Puls erhöht • Temp. normal	• O2-Sättigung erniedrigt, • AF erhöht • RR erniedrigt oder erhöht • Puls erhöht oder erniedrigt • Temp. normal	• O2-Sättigung normal, • AF normal • RR erniedrigt oder erhöht • Puls erhöht oder erniedrigt • Temp. normal	• O2-Sättigung normal, • AF normal oder erhöht • RR normal oder erhöht • Puls erhöht • Temp. normal
• Körperliche Untersuchung	• Emphysemzeichen • hypersonorer Klopfschall • Jugularvenenstauung • Atemgrenzen tiefstehend, leises Atemgeräusch • Giemen/Pfeifen • zentrale Zynose • Radiologisch: Hypertransparenz, tiefstehende abgeflachte Zwerchfelle, • peribronchiale Zeichnungsvermehrung	• Feuchte Rasselgeräusche klingend • Fieber • Pleurareiben • Abgeschwächte Atemgeräusche • Bronchialatmen	• Einseitige Beinschwellung • Adipositas • Tachykardie • Jugularvenenstauung	• Tachykardie • Bradykardie • Tachypnoe • Feuchte Rasselgeräusche • Prominenter Jugularvenenpuls, • periphere Ödeme • Hepatomegalie • 3. Herzschlag, • auffällige Herzgeräusche • Kachexie	• Blass • Tachykardie • Bradykardie	• Druckschmerz • Zentrale Hypoxie • Knistern bei Pneumothorax • Einseitig abgeschwächt. Atemgeräusch