Facilitating Collaborative Diagnostic Reasoning

Effects of Collaboration Scripts in Agent-Based Medical Simulations

Inaugural-Dissertation zur Erlangung des Doktorgrades der Philosophie

an der Ludwig-Maximilians-Universität München

vorgelegt von

Anika Radkowitsch

München, Juni 2020

Erstgutachter: Prof. Dr. Frank Fischer

Zweitgutachter: Prof. Dr. Martin Fischer

Tag der mündlichen Prüfung: 27.07.2020

Acknowledgements

I would like to express my sincere gratitude to Frank Fischer for supervising my thesis and for giving me the chance to participate in an inspiring scientific community. I experienced you as a fantastic leader with a good instinct for when guidance is needed and when not. Martin Fischer, thank you for always having an open ear for my concerns and my needs. I am impressed how you manage to – almost magically – come up with further resources and easy solutions for any problems. I am also glad and grateful for the opportunity to work with Ralf Schmidmaier. Thank you for being patient with my medical unknowing, for being a competent consultant and for sharing your expertise on both, medicine, and medical education.

Moreover, I would like to express my deepest thanks to the DFG research group COSIMA. Getting the chance to be part of this group enabled me not only to better understand crucial scientific research processes, but also to understand the importance of a scientific community. Particularly, I would like to thank Ansgar Opitz, Nicole Heitzmann, and Olga Chernikova, who smoothly paved the path for us doctoral students. Thank you for being open for discussions and questions and for making our voice heard in COSIMA. I am also much obliged to Amadeus Pickal, Elisabeth Bauer, Astrid Wirth, Max Fink, Angelika Wildgans, Christians Schons, Elias Codreanu, Maria Kramer, Stephanie Kron, and Sarah Bichler for sharing the experience of being a doctoral student including the hilarious journeys to Frauenchiemsee and elsewhere, amusing days and evenings, but also for sharing the often experienced confusion or fatigue after moments of tremendous workload.

Moreover, this thesis could not have been realized without the help of Hannah Gerstenkorn, Victoria Reitmeier, Felix Behrmann, Georg Kuschel, Martin Adler, Rosa Haas, Simone Steiger, Laura Brandl, Andreas Wildner, and Julia Haluszczynski. Thank you for your immense support during the development of the learning material, organization and implementation of the data collections, and data preparation. Freydis Vogel, Elisabeth Bauer, Michael Sailer, and Matthias Stadler, thank you for critical discussions and valuable feedback throughout the last years. I am glad for the fruitful collaborations of the last three years and am looking forwards to those that yet need to become fruitful.

I also want to give my thanks to P. Fischer, who was an important part of my non-academic life during the past 2.5 years. Thank you for showing me how to be kinder to and more patient with myself. You were a very benevolent and encouraging companion on this very special journey.

Beyond that, I am sincerely thankful to my parents, who supported me emotionally and financially during my entire academic life. Their support enabled me to focus my attention on my work. Without them, the last nine years would not have been as carefree as they were. I am also grateful to my family, Silja, Florian, Tina, Lina, and Alma, for reminding me on what is truly important in life.

Last but not least, I am grateful beyond all measures to Claudius who sometimes has lost his patience with, but never faith in myself. Thank you for being by my side all these years and for your emotional and academic support, especially throughout the last weeks of this thesis. Even if our future may still be unclear, the knowledge that we will walk the path together illuminates my days.

Abstract

Being a competent and efficient diagnostician in situations that require collaboration is essential in many professions. For instance, physicians regularly not only diagnose a patients' illness individually, but in collaboration with other physicians of the same or another professional background. Collaborative diagnostic reasoning is the competence to generate and evaluate evidence and hypotheses, and to share, elicit, and negotiate them with others in order to reduce uncertainty with respect to a professional decision. In the medical context this means, for example, that a physician shares and discusses symptoms and differential diagnoses about the disease of a patient with another physician in order to identify the most likely diagnosis. For developing collaborative diagnostic reasoning competences, the application of theoretical knowledge to case scenarios in collaborative situations is crucial. Implementing opportunities for practice in higher education is, therefore, essential. Simulations offer the opportunity to apply knowledge to authentic situations, thereby facilitating the learning of complex competences such as collaborative diagnostic reasoning. However, for the learning of complex competences instructional support beyond mere problem solving is often important. For the learning of collaborative competences, collaboration scripts provide socio-cognitive support and have been found to be effective for facilitating the learning of collaboration skills. Therefore, the goal of this thesis was to deepen the understanding of collaborative diagnostic reasoning and to identify conditions under which collaborative diagnostic reasoning of students can be effectively facilitated when learning with simulations and with collaboration scripts. To address these goals, a model for collaborative diagnostic reasoning (CDR model) was suggested, an agent-based simulation representing a common collaborative diagnostic situation in the medical context was developed, and three empirical studies were conducted. In contrast to human-to-human interaction, agent-based simulations offer a high standardization and a high control over collaborative processes and, therefore, offer ideal conditions for the investigation of the effectiveness of instructional support.

In all, the thesis comprises three studies. Study 1 examined the validity of the developed agent-based simulation. Study 2 examined the general effectiveness of collaboration scripts in the context of computer-supported collaborative learning. Study 3 examined the effectiveness of collaboration scripts when learning with agent-based simulations. Thus, while Study 1 examined collaborative diagnostic reasoning from a measurement perspective, Study 2 and 3 investigated its facilitation with collaboration scripts.

The first empirical study was meant to provide evidence for the validity of the agent-based simulation. In a quasi-experimental study, medical students with low and high prior knowledge were compared to internists with at least three years of working experience with respect to the quality of collaborative and diagnostic reasoning processes as well as their cognitive load when working with the simulation. Further, internists rated the authenticity of the simulation. Validity evidence was collected and interpreted based on Kane's (2006) validity framework. The study showed that the simulation allows one to differentiate between different levels of prior knowledge on most collaborative diagnostic reasoning measures, yet it also highlighted some weaknesses of the agent-based simulation that were addressed afterwards. For instance, the more general measure information sharing skill was split up into evidence sharing and evidence elicitation skills due to low internal consistency. Nevertheless, the results of this first study already indicated that in all the simulation is a valid instrument to assess collaborative diagnostic reasoning.

The second study addressed the question of whether collaboration scripts in general could be an effective instructional means for facilitating collaborative diagnostic reasoning. To this end, two comprehensive literature searches were conducted, resulting in the identification of 56 empirical studies on the effect of collaboration scripts on collaboration skills. Those studies were synthesized by means of a random effects meta-analysis. Moreover, the study addressed a widespread criticism according to which collaboration scripts undermine learners' agency and motivation. Such negative effects on motivation could – in the long run – affect learners' behavior and, thus, reduce their effectiveness. To assess the empirical validity of this argument, the meta-analysis also estimated the effect of collaboration scripts on motivation. The analyses showed that collaboration scripts have medium positive effects on collaboration skills, but no significant effect on motivation. Yet, only few studies examined the effect of collaboration scripts on motivation indicating that further evidence is necessary. Beyond that, the meta-analysis investigated moderating variables to identify conditions under which collaboration scripts are particularly effective. The findings indicated that collaboration scripts are descriptively more effective if collaboration scripts prompted only a single type of collaboration skills. Overall, the second study showed that collaboration scripts could be an effective instructional means to advance the learning of collaboration skills without undermining learners' agency.

The third study builds upon findings of the prior two studies and examined the effects of collaboration scripts on collaborative diagnostic reasoning when learning with agent-based simulations in an experimental intervention study. In this study, medical students worked on patient cases within the agent-based simulation used in Study 1. During the intervention, one third of the participants was supported with a static collaboration script and one third learnt with an adaptive collaboration script. The last group learnt without any instructional support. Both collaboration scripts consisted of meta-knowledge prompts about the simulated collaboration partners' tasks and responsibilities. The three groups were compared with respect to their performance in collaborative diagnostic reasoning, especially with respect to the subskills of evidence elicitation and evidence sharing, their extraneous cognitive load and the satisfaction of their basic psychological needs. For supporting evidence elicitation, static collaboration scripts seemed most effective. However, learners of all groups scored high on a post-test indicating that evidence elicitation was easier to learn and additional support in form of collaboration scripts was not necessary. For supporting evidence sharing, both collaboration scripts were effective during the intervention. Yet, only the adaptive collaboration script supported transfer to an unsupported post-test and, thus, internalization of the collaboration script. Moreover, the findings of the study suggest that adaptive collaboration scripts were more positive for learners' perceived competence and their extraneous cognitive load when compared to static collaboration scripts. However, both collaboration scripts did not significantly affect learners' perceived autonomy or perceived social relatedness. These findings suggest that for supporting learning with agent-based simulations, collaboration scripts consisting of meta-knowledge prompts are an effective instructional means. To exploit the full potential of collaboration scripts, they should be adapted to the learners' performance.

Overall, the results of this work emphasize the relevance of implementing the learning of collaborative diagnostic reasoning in higher education. The thesis provides further insights of how learning of collaborative diagnostic reasoning is effective. For instance, meta-knowledge about the collaboration partners' tasks, roles and responsibilities was identified as an important determinant of successful collaborative diagnostic reasoning. The three studies conducted further suggest, firstly, that agent-based simulations are suitable for modelling collaborative diagnostic reasoning processes. Secondly, that in general collaboration scripts are an effective instructional means for facilitating collaboration skills. Thirdly, that when implementing collaboration scripts in agent-based simulations to support collaborative diagnostic reasoning, adapting them to the learners' performance is beneficial since this supports the internalization of collaboration scripts and competence experience and reduces the extraneous cognitive load.

Deutsche Zusammenfassung

Für viele Berufe ist es wichtig in Situationen, die eine Zusammenarbeit erfordern, effizient diagnostizieren zu können. Beispielsweise diagnostizieren Ärzte die Krankheit eines Patienten regelmäßig nicht nur individuell, sondern auch in Zusammenarbeit mit Ärzten derselben oder einer anderen Fachrichtung. Unter kollaborativem Diagnostizieren versteht man die Kompetenz, Evidenzen und Hypothesen zu generieren und zu evaluieren und sie mit anderen zu teilen oder von anderen zu elizitieren und zu verhandeln, um die Unsicherheit in Bezug auf eine berufliche Entscheidung zu verringern. Zum Beispiel teilt und diskutiert ein Arzt Symptome und Differentialdiagnosen über die Krankheit eines Patienten mit einem anderen Arzt, um die wahrscheinlichste Diagnose zu ermitteln. Für die Entwicklung von kollaborativen Diagnosekompetenzen scheint die Anwendung von Wissen auf Fallszenarien in kollaborativen Situationen von entscheidender Bedeutung zu sein. Es ist daher unumgänglich, Möglichkeiten für das Üben solcher Anwendungen in die Hochschulausbildung zu implementieren, zum Beispiel durch den Einsatz von Simulationen. Simulationen bieten die Möglichkeit, das Wissen auf authentische Situationen anzuwenden, wodurch das Erlernen komplexer Kompetenzen wie zum Beispiel kollaboratives Diagnostizieren erleichtert wird. Für das Erlernen komplexer Kompetenzen scheint jedoch eine über die reine Problemlösung hinausgehende Unterstützung vorteilhaft zu sein. Für das Erlernen von kollaborativen Kompetenzen könnten Kollaborationsskripts lernwirksame sozio-kognitive Unterstützung zu bieten. Daher zielt diese Arbeit darauf ab, das Verständnis von kollaborativem Diagnostizieren zu vertiefen und Bedingungen zu identifizieren, unter denen kollaborative Diagnosekompetenzen von Studierenden beim Lernen mit Simulationen mit Kollaborationsskripts effektiv gefördert werden können. Um diese Ziele zu erreichen, wurde in dieser Arbeit ein Prozessmodell für kollaboratives Diagnostizieren (CDR-Modell) vorgeschlagen, eine agentenbasierte Simulation entwickelt, die eine übliche kollaborative diagnostische Situation im medizinischen Kontext darstellt. Im Gegensatz zur Mensch-zu-Mensch-Interaktion bieten agentenbasierte Simulationen eine hohe Standardisierung und eine hohe Kontrolle über kollaborative Prozesse und bieten daher ideale Voraussetzungen für die Untersuchung der Wirksamkeit instruktionaler Unterstützung. Außerdem wurden drei empirische Studien durchgeführt. Studie 1 untersuchte die Validität der entwickelten agentenbasierten Simulation. Studie 2 untersuchte die allgemeine Effektivität von Kollaborationsskripts beim computergestützten kollaborativem Lernen. Studie 3 untersuchte die Wirksamkeit von Kollaborationsskripts beim Lernen mit agentenbasierten Simulationen. Während Studie 1 kollaboratives Diagnostizieren aus einer methodischen Perspektive heraus

untersuchte, untersuchten Studie 2 und 3 die Förderung von Kollaborationsfähigkeiten bzw. kollaborativen Diagnostizierens durch Kollaborationsskripts.

Die erste empirische Studie zielte darauf ab, Evidenzen für die Validität der agentenbasierten Simulation zu erbringen. In einer quasi-experimentellen Studie wurden Medizinstudierende mit niedrigem und hohem Vorwissen mit Internisten mit mindestens drei Jahren Berufserfahrung hinsichtlich der Qualität kollaborativer und individueller Diagnoseprozesse sowie ihrer kognitiven Belastung beim Bearbeiten der Simulation verglichen. Ferner bewerteten die Internisten die Authentizität der Simulation. Evidenzen für die Validität wurden basierend auf Kane's (2006) Framework gesammelt und interpretiert. Die Studie zeigte, dass es mit der Simulation möglich ist, bezüglich der meisten kollaborativen Diagnosemaße zwischen verschiedenen Vorwissensgruppen zu unterscheiden. Die Studie zeigte aber auch einige Schwächen der agentenbasierten Simulation auf, die anschließend adressiert wurden. So wurde zum Beispiel die allgemeinere Fähigkeit zum Informationsaustausch aufgrund der geringen internen Konsistenz in Fähigkeiten zum Teilen und zum Elizitieren von Evidenzen aufgeteilt. Insgesamt deuten die Ergebnisse dieser ersten Studie darauf hin, dass die Simulation ein valides Instrument zur Messung kollaborativer Diagnoseprozesse ist.

Die zweite Studie befasste sich mit der Frage, ob Kollaborationsskripts grundsätzlich eine wirksame instruktionale Unterstützung zur Förderung kollaborativer Diagnosekompetenzen sein könnten. Zu diesem Zweck wurden zwei umfassende Literaturrecherchen durchgeführt, bei denen 56 empirische Studien über die Wirkung von Kollaborationsskripts auf Kollaborationsfähigkeiten im Kontext von computergestütztem kollaborativen Lernen identifiziert wurden. Diese Studien wurden mit Hilfe einer *random-effects* Meta-Analyse zusammengefasst. Darüber hinaus befasste sich die Studie mit einer weit verbreiteten Kritik, die besagt, dass Kollaborationsskripts die Autonomie und Motivation der Lernenden untergraben. Solche negativen Auswirkungen auf die Motivation könnten - auf lange Sicht - das Verhalten der Lernenden beeinflussen und damit die Wirksamkeit von Kollaborationsskripten verringern. Daher wurde in der Meta-Analyse auch die Wirkung von Kollaborationsskripts auf die Motivation analysiert. Die Ergebnisse der Analysen zeigten, dass Kollaborationsskripts mittlere positive Effekte auf die Kollaborationsfähigkeiten, aber keinen signifikanten Effekt auf die Motivation haben. Es untersuchten jedoch nur wenige Studien den Effekt von Kollaborationsskripts auf die Motivation, was darauf hindeutet, dass weitere Studien notwendig sind. Darüber hinaus untersuchte die Meta-Analyse moderierende Variablen, um Bedingungen zu identifizieren unter denen Kollaborationsskripts besonders effektiv sind. Die Ergebnisse zeigten, dass Kollaborationsskripte deskriptiv wirksamer sind, wenn diese nur eine einzige Kollaborationsfähigkeit unterstützen. Insgesamt zeigte die zweite Studie, dass Kollaborationsskripte eine wirksame instruktionale Unterstützung sein können, um das Erlernen von Kollaborationsfähigkeiten zu fördern, ohne die Motivation der Lernenden zu reduzieren.

Die dritte Studie baut auf den Ergebnissen der beiden vorangegangenen Studien auf, indem sie die Wirkung von Kollaborationsskripts auf kollaboratives Diagnostizieren beim Lernen mit agentenbasierten Simulationen in einer experimentellen Interventionsstudie untersuchte. In dieser Studie bearbeiteten Medizinstudenten Patientenfällen in der in Studie 1 verwendeten agentenbasierten Simulation. Während der Intervention wurde ein Drittel der Teilnehmenden mit einem statischen Kollaborationsskript unterstützt, ein Drittel lernte mit einem adaptiven Kollaborationsskript und die letzte Gruppe lernte ohne jegliche Unterstützung. Die Kollaborationsskripts bestanden aus Meta-Wissen Prompts über die Aufgaben und Verantwortungen des simulierten Kollaborationspartners. Die drei Gruppen wurden bezüglich ihrer Performanz beim kollaborativen Diagnostizieren, insbesondere bezüglich der Teilfähigkeiten Evidenz elizitieren und Evidenz teilen, ihrer extrinsischen kognitiven Belastung und der Erfüllung ihrer psychologischen Grundbedürfnisse verglichen. Zur Unterstützung beim Evidenzen elizitieren schienen statische Kollaborationsskripts am wirksamsten zu sein. Allerdings erzielten die Lernenden aller Gruppen in einem Post-Test hohe Punktzahlen, was darauf hinweist, dass Evidenzen elizitieren leichter zu erlernen ist und zusätzliche Unterstützung in Form von Kollaborationsskripts nicht zwingend notwendig ist. Zur Unterstützung von Evidenz sharing waren beide Kollaborationsskripts während der Intervention wirksam. Allerdings unterstützte nur das adaptive Kollaborationsskript den Transfer auf den Post-Test und damit die Internalisierung des Kollaborationsskripts. Darüber hinaus legen die Ergebnisse der Studie nahe, dass adaptive Kollaborationsskripts im Vergleich zu statischen Kollaborationsskripts positiver für das Kompetenzerleben der Lernenden und ihre extrinsische kognitive Belastung waren. Beide Kollaborationsskripts hatten jedoch keinen signifikanten Einfluss auf das Autonomieerleben oder die soziale Eingebundenheit der Lernenden. Diese Ergebnisse zeigen, dass Kollaborationsskripts zur Unterstützung des Lernens mit agentenbasierten Simulationen wirksam sind. Um das volle Potenzial von Kollaborationsskripts auszuschöpfen, sollten sie jedoch an die Leistung der Lernenden angepasst werden.

Insgesamt heben die Ergebnisse dieser Arbeit hervor, wie relevant die Implementierung des Lernens kollaborativen Diagnostizierens in der Hochschule ist. Die

Arbeit liefert außerdem Erkenntnisse darüber, unter welchen Bedingungen das Lernen kollaborativen Diagnostizierens effektiv ist. Beispielsweise wurde Meta-Wissen über die Aufgaben und Verantwortung des Kollaborationspartners als wichtig für erfolgreiches kollaboratives Diagnostizieren identifiziert. Des Weiteren legen die drei durchgeführten Studien nahe, dass, erstens, agentenbasierte Simulationen zur Modellierung kollaborativer Diagnoseprozesse geeignet sind. Zweitens sind Kollaborationsskripts im Allgemeinen eine wirksame instruktionale Unterstützung zur Förderung von Kollaborationsfähigkeiten. Drittens ist bei der Implementierung von Kollaborationsskripts in agentenbasierte Simulationen zur Unterstützung des kollaborativen Diagnostizierens eine Anpassung an die Leistung der Lernenden von Vorteil, um die Internalisierung von Kollaborationsskripts und das Kompetenzerleben zu unterstützen und die extrinsische kognitive Belastung zu reduzieren.

Contents

1 General Introduction

Anika Radkowitsch

This chapter partly draws upon ideas that were first presented in a manuscript accepted for publication as:

Radkowitsch, A., Sailer, M., Fischer, M.R., Schmidmaier, R., & Fischer, F. (accepted). Diagnosing collaboratively: A theoretical model and a simulation-based learning environment. In F. Fischer & A. Opitz (Eds.) *Learning to diagnose with simulations - examples from teacher education and medical education*. Series: Springer Briefs in Education. New York: Springer.

1.1 Aim and Structure of the Thesis

The ability to collaborate with others is a core competence of the 21st century (e.g., Griffin & Care, 2015; OECD, 2017). The importance of collaboration already becomes clear when reflecting on its influence on our daily routine. Most products that we use on a daily basis – be it the desk we sit at or the computer we conduct our analyses with – are invented and produced collaboratively (Hutchins, 1995), and even the overall standard of living of a country depends on its citizens' ability to collaborate (Gräbner, Heimberger, Kapeller, & Schütz, 2020; Hidalgo, Klinger, Barabási, & Hausmann, 2007). Collaboration frequently enhances the quality of solutions since it allows for the integration of multiple perspectives and sources of knowledge and expertise in the problem solving process (Graesser et al., 2018). Thus, it does not come as a surprise that the importance of collaboration as a core competence also has become visible in educational and psychological research. For instance, researchers have proposed frameworks on and instruments for collaboration in different contexts such as collaborative problem solving and collaborative learning (Fiore et al., 2010; e.g., Liu, Hao, von Davier, Kyllonen, & Zapata-Rivera, 2015; e.g., Sun et al., 2020), or collaboration at the workplace (e.g., Claus & Wiese, 2019).

The present thesis is concerned with collaboration in a context that until now - notwithstanding the actual relevance of collaboration - has been investigated mainly as an individual competence: diagnostic reasoning. Diagnostic reasoning is the "goal-oriented collection and interpretation of case-specific or problem-specific information to reduce uncertainty in order to make […] [professional] decisions" (Heitzmann et al., 2019, p. 4). It is relevant for many professions such as teaching (e.g., Wildgans-Lang, Scheuerer, Obersteiner, Fischer, & Reiss, 2020), medicine (e.g., Norman, 2005), or mechatronic engineering (e.g., Abele, 2018). Particularly in the medical context, collaborative diagnostic reasoning is part of many diagnosticians' daily routines, which makes medicine an attractive context for investigating collaborative diagnostic reasoning. For instance, physicians discuss diagnoses or treatment plans of patients in roundtable discussions such as tumor boards or consult more specialized physicians to identify the best diagnosis or treatment for a patient. An example for the latter case is physicians working in the emergency departments who often request radiological investigations to obtain further evidence about the patient's disease. For a correct interpretation of such radiological images, clinical details as well as precise clinical questions are crucial. Otherwise, severe errors during the radiologic reasoning process, such as interpreting a post-surgical change in the breast for malignancy or mistaking a thrombus for a retained cannula, could occur (Davies et al., 2018). Therefore, the requesting physicians must be able to identify information that is important for their collaboration partners and share it, or identify relevant information that the collaboration partner might have and elicit it accordingly. In practice, however, such collaborative diagnostic reasoning often functions deficiently since, for instance, inadequate clinical information has been identified as a major cause of errors in radiology (Brady, Laoide, McCarthy, & McDermott, 2012). Given the relevance of collaborative diagnostic reasoning in medicine and the severe implications of errors, a closer examination of collaborative diagnostic reasoning in the context of medicine seems warranted. The present thesis addresses this challenge. I assume, however, that the findings presented in this thesis transfer to different contexts as well, although this is an assumption that ultimately requires empirical testing, an issue that will be taken up in the discussion at the end of the thesis.

Lately, the relevance of collaboration in medical education has also been recognized by medical educators. For example, in a German national competence-based learning goal catalogue, the role of physicians as members of a team is emphasized by setting the goal that medical students should learn collaboration skills during medical education (MFT Medizinischer Fakultätentag der Bundesrepublik Deutschland e. V., 2015). Several simulation centers have also acknowledged the importance of team trainings. They offer full-scale trainings of different scenarios with simulated patients, ambulances, or emergency departments (e.g., Gardner & Ahmed, 2014). Such simulation-based trainings provide opportunities for practice in a controlled, ethically unobjectionable, and risk-free environment (Ziv, Wolpe, Small, & Glick, 2003). However, full-scale trainings are expensive and time consuming, which is why medical students and physicians rarely get the chance to actively participate in such trainings. Rather, they spent a lot of time observing peers acting in the simulation (Zottmann, Dieckmann, Taraszow, Rall, & Fischer, 2018). For the learning of competences as complex as collaborative diagnostic reasoning, however, it is necessary that learners practice actively and repeatedly while focusing on particular difficult subtasks (e.g., Ericsson, 2004; Kolodner, 1992).

During practicing, learners ideally receive further instructional support to avoid overwhelming learners (Kirschner, Sweller, & Clark, 2006). Thus, providing learners with the opportunity to learn with small-scale simulations complemented with further instructional support, such as socio-cognitive scaffolds, could offer effective conditions for learning collaborative diagnostic reasoning. Such socio-cognitive scaffolds are, for instance, collaboration scripts which support learners to develop internal cognitive schemata that guide collaborative practices or modify existing cognitive schemata that do not result in beneficial actions (Fischer, Kollar, Stegmann, & Wecker, 2013). This thesis aims to enhance our understanding of collaborative diagnostic reasoning and to investigate conditions under which learners effectively learn to diagnose collaboratively when learning with simulations.

The remainder of this thesis is structured in three main parts: the first part is dedicated to the theoretical underpinning. First and foremost, I define collaborative diagnostic reasoning. To this end, I introduce a model for collaborative diagnostic reasoning that distinguishes between *individual* and *collaborative* diagnostic activities and explicates their interaction. Next, I describe how collaborative diagnostic reasoning can be facilitated by two different means of instructional support – agent-based simulations and collaboration scripts – and describe the general research questions of the thesis. In the second part, I present three studies that were conducted to address the research questions. The first study validated an agent-based simulation that was developed with the goal to facilitate and measure collaborative diagnostic reasoning. The second and third study examined the effectiveness of collaboration scripts. While Study 2 took a more general approach in form of a meta-analysis on the effectiveness of collaboration scripts, Study 3 builds upon both prior studies and examined the effectiveness of collaboration scripts when it comes to facilitate collaborative diagnostic reasoning within agent-based simulations. In the third part of this thesis, I synthesize the results of the three studies and discuss their joint implications for my theoretical model and the research questions of this thesis.

1.2 Collaborative Diagnostic Reasoning

To date, collaborative diagnostic reasoning has not been defined thoroughly, which makes an explicit examination and definition of the construct necessary. For that, it is necessary to draw upon the rich body of literature on related constructs. First and foremost, diagnostic reasoning can be understood as a subtype of problem solving. Problem solving is the goal-directed and not routinely achieved transition from the current state of a system to a goal state (Jonassen, 2000). Problem solving has also been examined in collaborative contexts (i.e., collaborative problem solving; e.g., Herborn, Stadler, Mustafić, & Greiff, 2018). The literature on collaborative problem solving is comprehensive and provides some fundamental theoretical underpinnings necessary for defining collaborative diagnostic reasoning (for a review on collaborative problem solving see Graesser et al., 2018). For PISA 2015, the Organisation for Economic Co-operation and Development (OECD) brought together several outstanding researchers who defined the competence of collaborative problem solving as

> "the capacity of an individual to effectively engage in a process whereby two or more agents attempt to solve a problem by sharing the understanding and effort required to come to a solution and pooling their knowledge, skills and efforts to reach that solution" (OECD, 2017, p. 134).

This definition highlights three important aspects: first, the competence of collaborative diagnostic reasoning is an individual competence and, therefore, can be assessed at the level of individuals (OECD, 2017). This is a crucial assumption since this makes it possible to facilitate and to assess collaborative diagnostic reasoning in individual, standardized settings such as agent-based collaboration (see Section 1.3.1). Second, collaborative problem solving is the joint attempt to engage in problem solving processes to reach a common goal. Thus, the specific goal of the problem (i.e., identifying an accurate diagnosis for a patient) and the cognitive activities and knowledge necessary to reach that goal define problem solving processes occurring during collaborative diagnostic reasoning. Finally, the process of collaborative problem solving involves both, cognitive and collaborative activities. Whereas cognitive activities refer to processes related to individual problem solving, collaborative activities refer to processes related to the interaction among collaborators (Graesser et al., 2018; OECD, 2017). Hence, it can be assumed that collaborative diagnostic reasoning similarly consists of individual cognitive and collaborative activities. Below, I will describe what is known about individual diagnostic reasoning and more specifically diagnostic reasoning in the medical context and related cognitive activities. In a next step, by further describing collaborative processes, these ideas will be extended to the collaborative context.

1.2.1 Individual diagnostic activities

Individual diagnosing has the goal to identify a diagnosis that is accurate (Simmons, 2010) and enables taking appropriate treatment actions (Charlin et al., 2012). To reach this goal, diagnosticians engage in epistemic reasoning (Heitzmann et al., 2019). The epistemic reasoning model Scientific Discovery as Dual Search (SDDS, Klahr & Dunbar, 1988) conceptualizes (scientific) reasoning as the coordination between evidence and hypotheses by searching through two hypothetical spaces: a hypothesis space containing a universal set of hypotheses and an experiment space containing means to test hypotheses. The idea of the universal hypothesis space was later extended by Van Joolingen and De Jong (1997) by adding a learner hypotheses space consisting of those hypotheses considered during the reasoning process. To describe the coordination between those spaces (i.e., reasoning), Klahr and Dunbar (1988) propose a hierarchy of cognitive processes that comprises the activities of specifying hypotheses, deriving predictions from hypotheses, and testing and evaluating the hypotheses in the light of existing evidence. In a similar, yet more concise activity-oriented reasoning model, Fischer et al. (2014) suggest eight epistemic activities (problem identification, asking questions, generating evidence, evaluating evidence, generating hypotheses, constructing artefacts, drawing conclusions, scrutinizing results) that have been found useful for knowledge generation in different domains and already have been used to describe diagnostic reasoning processes (Heitzmann et al., 2019). In contrast to the SDDS model, Fischer et al. (2014) assume no hierarchical structure or order of epistemic activities. This makes their model more flexible. At the same time, Klahr and Dunbar's (1988) notion of two spaces allows for a more explicit description of the interdependencies between the suggested epistemic activities, which is useful for understanding diagnostic processes. On a more abstract level, reasoning processes further have been described with so called *dual-process theories*. These theories distinguish two types of reasoning processes, one quick, intuitive and non-analytic, and one slow, effortful and deliberate (Chaiken & Trope, 1999). For instance, Kahneman, Slovic, and Tversky (1982) distinguish two sets of processes, named System 1 and System 2, based on their speed and the content these processes act upon. More specifically, they assume that System 1 automatically and with high speed produces answers to a problem and System 2 monitors these processes and intervenes to correct or override the answers (Kahneman & Frederick, 2002). Although such models may provide some insight into how easily and accurate diagnosticians identify a diagnosis, they are less beneficial for explaining the content of diagnostic reasoning processes.

The presented general descriptions of reasoning processes align well with descriptions of diagnostic reasoning processes in the medical context. When diagnosing a patient's illness, experienced physicians often quickly come up with an initial hypothesis (Charlin, Boshuizen, Custers, & Feltovich, 2007). Such initial hypotheses are often generated effortlessly and non-analytically based on pattern recognition (Norman, Young, & Brooks, 2007) and are often surprisingly accurate (Elstein, Shulman, & Sprafka, 1978). Feltovich and Barrows (1984) introduced the *illness script theory* to explain these findings. According to this theory, medical knowledge necessary for identifying a diagnosis matching signs and symptoms of a patient is stored in so-called illness scripts. Illness scripts consist of knowledge about the underlying pathophysiological processes of a disease, its consequences in form of signs and symptoms, as well as enabling conditions (i.e., characteristics and circumstances that make a specific diagnosis more likely such as age, gender, or familial pre-existing conditions). The theory assumes that based on identification of initial relevant cues (particularly such describing enabling conditions, Van Schaik, Flynn, Van Wersch, Douglass, & Cann, 2005), specific illness scripts become activated. This allows the diagnosticians to make predictions about further symptoms (Barrows & Feltovich, 1987; Elstein et al., 1978). By that, illness scripts guide the further data gathering that allows physicians to confirm or refute the activated hypothesis (Charlin et al., 2007; Elstein & Schwarz, 2002). There is much empirical evidence endorsing the illness script theory (Custers, 2015). For instance, once a specific illness script is activated by naming a diagnosis, typical case information is processed faster than atypical case information (Custers, Boshuizen, & Schmidt, 1996).

According to the illness script theory, an important characteristic of medical expertise is, thus, the efficient organization of knowledge. Although it is assumed that novices and laypersons also have illness scripts (Custers, 2015), Schmidt and Boshuizen (1993) hypothesize that their structure fundamentally changes during the development of medical expertise. In early phases of medical training, students acquire much biological and pathophysiological knowledge that is stored in networks explaining causes and consequences of diseases. Medical students in this early phase of expertise development tend to use this pathophysiological knowledge to deliberately relate isolated signs and symptoms of a patient to each other (Boshuizen & Schmidt, 1992). These reasoning processes are considered deliberate, slow, and prone to error (Rikers, Schmidt, & Boshuizen, 2000). However, through repeated application, pathophysiological knowledge becomes *encapsulated* into simplified models that are more efficient but have the same explanatory power. Hence, experienced physicians base their diagnostic reasoning more on clustered knowledge about symptoms and patterns of symptoms than on isolated signs and pathophysiological knowledge (Schmidt & Rikers, 2007). Although experienced physicians access pathophysiological knowledge less frequently compared to medical students, they still have some access to this knowledge and use it when necessary, for example during patient communication or when diagnosing particularly difficult patient cases (Charlin et al., 2007; Patel, Groen, & Arocha, 1990).

Medical educators have outlined the educational implications of the presented theories (e.g., Eva, 2005). Most importantly, giving medical students an early opportunity to apply their knowledge to authentic patient cases seems vital for the restructuring of knowledge (Boshuizen, Gruber, & Strasser, 2020). During medical education, medical students should process a large number of patient cases with different diseases. These patient cases should represent authentic patient cases showing the range of different representations of specific diseases. Particularly complex and elaborated cases and cases of which medical students know the solution upfront seem less beneficial for the development of illness scripts (Custers, 2015; Eva, 2005; Lubarsky, Dory, Audétat, Custers, & Charlin, 2015).

To sum up, when engaging in diagnostic reasoning, physicians generate and evaluate evidence and hypothesis. These processes are particularly influenced by the physicians' medical knowledge, its structure and organization. The structure and organization of medical knowledge (i.e., whether pathophysiological knowledge is encapsulated and whether illness scripts are available) mainly affects how effortful and deliberate diagnostic reasoning processes are (Charlin et al., 2007). Medical expertise becomes particularly evident in the speed of coming up with an accurate hypothesis (Sherbino et al., 2012; Woods, Howey, Brooks, & Norman, 2006), the knowledge about enabling conditions (Schmidt & Rikers, 2007) but less in the quantity of diagnostic reasoning processes (Hodges, McNaughton, & Tiberius, 1999). The educational implications of the presented theoretical and empirical evidence are straightforward: for the restructuring and reorganization of medical knowledge, the early exposure to patient cases is considered essential (Eva, 2005; Eva, Hatala, LeBlanc, & Brooks, 2007; Lubarsky et al., 2015).

1.2.2 Collaborative diagnostic activities

For collaborative diagnostic reasoning *collaboration* is relevant in addition to the described individual diagnostic activities. According to the OECD's definition of collaborative problem solving introduced above, collaboration has the function to pool "knowledge, skills and efforts" (OECD, 2017, p. 134) distributed among collaborators and required to reach the solution of a problem. There is a rich body of literature on collaborative processes from different research perspectives describing conditions for and processes of successful collaboration. For instance, collaboration is examined in organizational psychology and management (e.g., DeChurch & Mesmer-Magnus, 2010; Kozlowski, 2018), cognitive psychology (e.g., Hinsz, Tindale, & Vollrath, 1997), educational psychology (e.g., Fischer et al., 2013; Kozlov & Große, 2016), or social psychology (e.g., Larson, Christensen, Franz, & Abbot, 1998). They agree that collaboration is a complex process but differ in the way and focus for investigating it. Two different strands of research can be distinguished: firstly, research investigating *collaborative activities* focuses on how collaborators interact with each other. This line of research closely observes behavioral processes of individuals to differentiate between successful and unsuccessful collaborative processes (e.g., Andrews-Todd & Forsyth, 2018; OECD, 2017; Sun et al., 2020). The second line of research focusses on the cognition related to collaboration (e.g., Cannon-Bowers, Salas, & Converse, 1993; Wegner, 1987). This research strand attempts to understand how *knowledge* relevant for collaboration and its distribution among team members is organized and how it affects team performance (Mohammed & Dumville, 2001). To understand collaborative processes during collaborative diagnostic reasoning, both research strands are important. Below I will shortly introduce them and describe how they help understanding collaborative diagnostic reasoning.

With respect to research on collaborative activities, several models that describe collaborative activities in the context of collaborative problem solving and that mainly differ in their granularity have been proposed recently (e.g., Hesse, Care, Buder, Sassenberg, & Griffin, 2015; Liu et al., 2015; Sun et al., 2020). For example, Liu et al. (2015) suggest four social skills (sharing, negotiating, regulating problem solving, and maintaining communication) and provide a coding scheme to categorize team talk (e.g., Andrews-Todd & Forsyth, 2018). Hesse et al. (2015) propose three skills (perspective taking, participation, and social regulation) with two to four subskills each. Differences in the granularity considered aside, these models overlap to large extents. They seem to concur on the following collaborative activities (although some use an alternative terminology): *Sharing* information in a way that takes the audience's background into account, *eliciting* information from collaboration partners to extend the own knowledge, *negotiating* conflicting ideas, *regulating* the collaborative processes by setting goals and monitoring progress. Sharing and elicitation, or, in other words, the pooling of information, have been considered crucial for processing information on a group level (Fischer, Bruhn, Gräsel, & Mandl, 2002; Hinsz et al., 1997). These collaborative activities allow collaborators to construct a shared mental representation of the problem and possible solutions, which is necessary for successful collaboration (Meier, Spada, & Rummel, 2007; Roschelle & Teasley, 1995). The need to negotiate conflicting ideas between collaborators is particularly important when conflict between collaborators arise (Hesse et al., 2015). Successful negotiation can prevent groups, for instance, from ignoring dissenting information or premature closure (Nickerson, 1988; Patel, Kaufman, & Arocha, 2002). Finally, regulation is relevant to coordinate goals and strategies to reach these goals (Järvelä & Hadwin, 2013).

These activities have received varying degrees of attention in different contexts. In collaborative contexts in which collaborators have roughly equally distributed knowledge, engaging in all these collaborative activities seems beneficial for team performance (Andrews-Todd & Forsyth, 2018). In contexts in which collaborators depend on each other's knowledge, research has emphasized the importance of sharing and eliciting information. Research on the collaborative activities sharing and elicitation started with a paradigm that became known as *hidden profile* paradigm (Stasser & Titus, 1985, 1987). In their prototypical studies, Stasser and Titus (1985) asked groups of participants to make a decision based on information that was distributed among participants. The information was distributed among team members in a way that some information was shared, and some was unique to specific team members. To make the best decision in these scenarios, pooling of unshared information was necessary. Their findings showed that information distributed between team members was significantly less likely to be shared with other team members and was also less likely to be incorporated into the final decision. Hence, teams often failed to make the best possible decision. The effect that unshared information is less likely to be shared with team members also has been replicated in the context of collaborative diagnostic reasoning (Christensen et al., 2000; Larson et al., 1998; Tschan et al., 2009). Larson et al. (1998), for instance, showed that even highly trained medical teams engaging in authentic problem solving are less likely to pool unshared information over shared information. The accuracy of the diagnoses was higher when more unshared information was pooled. These studies suggest that sharing and elicitation of information is difficult when group members depend on each other's knowledge, and that this difficulty threatens the quality of problem solutions (e.g., the accuracy of a diagnosis).

An general explanation for how and why collaborators fail to engage in beneficial sharing and elicitation processes can be found in the *script theory of guidance* that was developed in the context of computer-supported collaborative learning (Fischer et al., 2013). Fischer and colleagues (2013) argue that collaborative practices are dynamically shaped by internal collaboration scripts. Internal collaboration scripts consist of four hierarchically ordered types of components (play, scene, scriptlet, and role) that dynamically guide the collaborative process. The configuration of the internal collaboration script is influenced by the goals and perceived situational characteristics of the collaboration partners (Fischer et al., 2013). This implies that the internal collaboration scripts of diagnosticians – or more generally collaborators – determines how and in which collaborative activities they engage. Overall, this first strand of research suggests that when collaborating, diagnosticians engage in the collaborative activities of *sharing*, *elicitation*, *negotiation*, and *regulation*. In the medical context, it is particularly *sharing* and *elicitation* that have received particular attention (e.g., Tschan et al., 2009). The collaborative processes of diagnosticians are guided by so-called internal collaboration scripts.

The second strand of research on collaboration might complement the first by understanding why the sharing and elicitation of information is difficult. An important theory for that is transactive memory theory (Wegner, 1987). The *transactive memory theory* assumes a transactive memory system of a team that consists of several individual memory systems and a collective understanding of how knowledge is distributed across the team. Hence, team members can be perceived as an external memory resource. The transactive memory theory emphasizes that if information is distributed among collaborators, it is important to know *how* the information is distributed among the team (Peltokorpi, 2008; Wegner, 1987). This assumption is supported by a study of Stasser, Vaughan, and Stewart (2000) who found that the hidden profile effect introduced above diminishes when roles are distributed among collaborators. The authors argued that those roles raise awareness of the own knowledge and the collaborators knowledge. Similar effects were found in a study by Engelmann and Hesse (2010). They distributed information among dyads working on a problem solving task. Some of the dyads were additionally provided with meta-knowledge, that is information about how knowledge is distributed between the participants. Those who were made aware of the knowledge distribution tended to discuss unshared information earlier and were more likely to include such information in the problem solution.

Another theory that has been used to explain team effectiveness with the organization of knowledge is the *shared mental model theory* (Cannon-Bowers et al., 1993). The shared mental model theory assumes that teams whose members share a similar understanding of, inter alia, the task or the team perform better than teams that do not. In contrast to the transactive memory theory that focusses on how knowledge is distributed among the team, the shared mental model theory emphasizes knowledge that is similar for team members. Although both theories seem quite distinct at first, they are in fact complementary to each other. In that, the shared mental model theory also assumes that collaborators have a shared mental model of the team that includes the team members' knowledge, roles, and responsibilities (Cannon-Bowers et al., 1993). According to Cannon-Bowers, Tannenbaum, Salas, and Volpe (1995), such shared team mental models are important since they allow team members to anticipate the other team members' activities and to adapt the own activities accordingly. Empirical evidence for the importance of shared team mental models comes from a study by Mathieu, Heffner, Goodwin, Salas, and Cannon-Bowers (2000) who investigated the effect of shared team mental models on collaborative processes and outcomes of dyads during collaborative problem solving in a flight simulator. They found that similar mental models of the team enhanced the quality of collaborative processes, which in turn had positive effects on problem solutions. Thus, the transactive memory theory and the shared mental model theory emphasize the importance of knowledge about the team members' knowledge, roles, and responsibilities. Henceforth, I call this knowledge *meta-knowledge* in the style of group awareness literature (e.g., Engelmann & Hesse, 2011).

Transferring the presented theories to collaborative diagnostic reasoning I assume that when diagnosing collaboratively, physicians need to engage in collaborative activities such as sharing, elicitation or negotiation of information and knowledge (Liu et al., 2015). To do so effectively, physicians need functional corresponding internal collaboration scripts (Fischer et al., 2013) and, further, understand the knowledge, roles and responsibilities of their counterparts (Cannon-Bowers et al., 1993; Wegner, 1987). A shared conception of the task and the team seems particularly beneficial for anticipating the collaboration partners' activities.

1.2.3 Proposing a model for collaborative diagnostic reasoning

This section synthesizes the theories and empirical evidence presented above by, firstly, providing a general definition of collaborative diagnostic reasoning and, secondly, by proposing a model for collaborative diagnostic reasoning (CDR model, for a comprehensive description see Radkowitsch, Sailer, Fischer, Schmidmaier, & Fischer, accepted). Based on the literature presented above, collaborative diagnostic reasoning is defined as 'the coordinated process of diagnosing a patient's problem with at least one other diagnostician by generating and evaluating evidence and hypotheses and by sharing, eliciting, and negotiating evidence and hypotheses under consideration of the own and the collaboration partners' knowledge, roles, and responsibilities, with the goal to identify an accurate diagnosis and to reduce uncertainty to a degree that enables taking appropriate action' (see Cannon-Bowers et al., 1993; Charlin et al., 2007; Heitzmann et al., 2019; Hesse et al., 2015; Liu et al., 2015; Wegner, 1987).

The CDR model was developed and is described in the medical context. However, I[1] suppose that it is also valid for collaborative diagnostic reasoning in other domains. The CDR model describes collaborative diagnostic reasoning processes based on individual diagnostic activities (Fischer et al., 2014), collaborative diagnostic activities (e.g., Andrews-Todd & Forsyth, 2018; Liu et al., 2015), and their interaction (Klahr & Dunbar, 1988). The CDR model as depicted in Figure 1 describes collaborative diagnostic activities of two diagnosticians. I assume, however, that the same processes apply during collaborative diagnostic reasoning of more than two diagnosticians. Collaborative diagnostic reasoning is typically triggered by cues from a *system to be diagnosed* (STBD). The STBD is an external

[1] The CDR model was first described by Radkowitsch, Sailer, et al. (accepted). Although henceforth referring to *my* model, I acknowledge that the model was developed in a collaborative effort.

system that contains any information about the patient that could possibly be considered during the diagnostic process. For instance, the STBD can provide information about the social environment of the patient or laboratory test results. Thus, the conception of the STBD is a holistic perspective of the patient within a bio-psycho-social environment. A diagnostician starts the diagnostic process by perceiving these cues and evaluating and generating them as evidence. I consider any information about the STBD that could influence the diagnosis of the system's status as evidence, including findings (e.g., laboratory value), a symptom (e.g., headache), or enabling conditions (e.g., smoking). A diagnostician generates evidence by interpreting information as relevant or generating new information about the STBD by, for instance, asking a question. Ideally, evidence gets evaluated with respect to its validity (Fischer et al., 2014). Based on patterns of evidence, the diagnosticians generate one or more hypotheses that is a statement about an assumed status of a system (Fischer et al., 2014) in an effortful, deliberate and/or non-analytical reasoning process (Norman et al., 2007). The generated hypotheses allow diagnosticians to make predictions about the status of specific aspects of the STBD (e.g., increased inflammation values) that are tested in the evidence space.

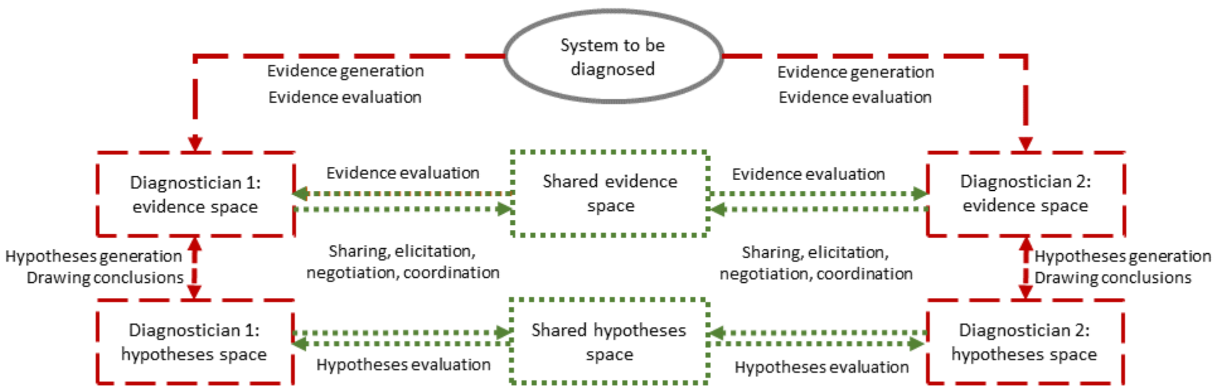

Figure 1. Collaborative Diagnostic Reasoning (CDR) Model. Red dashed lines represent individual diagnostic activities; green dotted lines represent collaborative diagnostic activities. This figure is adapted from Radkowitsch, Sailer, et al. (accepted).

I suggest that the results of these processes (i.e., evidence and hypothesis) are stored in two distinct cognitive storages, namely the evidence space and the hypothesis space. The idea of an evidence space and a hypothesis space is borrowed from the SDDS model (Klahr & Dunbar, 1988). Although the idea of an evidence and hypothesis space in the CDR model is inspired by the SDDS model from Klahr and Dunbar (1988), I have quite a different notion of those spaces. Klahr and Dunbar (1988) understand the experiment and hypothesis space as a hypothetical space that includes a universe of possible hypotheses and means to test them.

They understand scientific reasoning as the search through those spaces. In contrast, the evidence space and hypothesis space in the CDR model are cognitive storages that contain outcomes of reasoning processes (evidence and hypothesis). Thus, they align more with the notion of the learners' hypothesis space introduced by Van Joolingen and De Jong (1997). Yet, as Klahr and Dunbar (1988) I assume that specifying a hypothesis, deriving predictions from this hypothesis and testing them, and evaluating the evidence with respect to the hypothesis are important components of the reasoning process and that, hence, the coordination between evidence (data) and hypothesis (theory) is essential for successful diagnosing. Based on the interaction between hypothesis and evidence space, hypotheses are weighted and/or a conclusion is drawn about the most likely diagnosis.

So far, the presented diagnostic activities describe individual diagnostic reasoning. In a collaborative diagnostic situation, the diagnosticians additionally engage in collaborative activities. How diagnosticians engage in collaborative activities is influenced by their internal collaboration scripts (Fischer et al., 2013). By engaging in collaborative activities (i.e., sharing, elicitation, negotiation, coordination), a shared evidence and a shared hypothesis space develops consisting of any evidence and hypotheses shared among the diagnosticians. More specifically, diagnosticians share, elicit, negotiate, or coordinate evidence and hypotheses. Only evidence and hypotheses stored in at least one individual space can become part of the shared spaces. In contrast to the individual spaces a shared space is not assumed to be a cognitive storage but an external system. However, I assume that diagnosticians have mental representations of the information contained in the shared spaces (Roschelle & Teasley, 1995). When using evidence and hypotheses from the shared space for individual diagnostic reasoning, evidence and hypotheses are evaluated optimally with respect to their validity. These evaluation processes can be influenced by knowledge or assumptions on the collaboration partner. For instance, the perceived credibility of a collaboration partner influences which information mentioned by the collaboration partner is remembered (Andrews & Rapp, 2014).

The individual and collaborative diagnostic activities described in the CDR model present a process perspective on collaborative diagnostic reasoning. However, in both aforementioned areas (diagnostic reasoning and collaboration), research has emphasized the importance of knowledge for individual and collaborative diagnostic reasoning. Based on the literature presented above (Section 1.2), I assume that medical knowledge and its organization affect when and how diagnosticians engage in individual diagnostic activities (i.e., evidence and hypotheses generation and evaluation). Further, meta-knowledge particularly affects when and how diagnosticians engage in collaborative activities (i.e., negotiation, sharing, elicitation, coordination). Beyond that, general social (e.g., openness; Stadler, Herborn, Mustafić, & Greiff, 2019) and cognitive factors (e.g., intelligence; Stadler, Becker, Gödker, Leutner, & Greiff, 2015) are assumed to affect collaborative diagnostic reasoning (Hetmanek, Engelmann, Opitz, & Fischer, 2018; Wimmers, Splinter, Hancock, & Schmidt, 2007).

Overall, the CDR model describes collaborative diagnostic reasoning processes based on theories and empirical findings from diverse domains. The model is based on general reasoning models (Fischer et al., 2014; Klahr & Dunbar, 1988), but goes beyond them by combining individual and collaborative diagnostic processes (e.g., Liu et al., 2015) and suggesting how these processes are influenced by prior knowledge. By that, the CDR model provides a first theoretical underpinning for the facilitation of collaborative diagnostic reasoning.

1.3 Instructional Support for Facilitating Collaborative Diagnostic Reasoning

Collaborative diagnostic reasoning is a complex competence, and the few available studies show that even for experienced physicians, collaborative diagnosing is difficult. Particularly the sharing and elicitation skills require facilitation (e.g., Larson et al., 1998; Tschan et al., 2009). Therefore, identifying conditions under which diagnosticians effectively learn collaborative diagnostic reasoning seems crucial. The prior sections have shown that for the learning of diagnostic reasoning, an early confrontation with patient cases is important for the development of illness scripts (Lubarsky et al., 2015). Yet, confronting inexperienced medical students with real patients could set the patients' health at risk and is very cost- and time-consuming, due to the necessity of close supervision (Ziv et al., 2003). Thus, simulating patients or diagnostic situations was established for giving medical students the opportunity to develop illness scripts in a risk-free and economic learning environment (Cook, Erwin, & Triola, 2010). A simulation is a simplified model of a real-world scenario that allows to engage in specific activities with the goal to develop skills and competences necessary to master the respective situation (Bradley, 2006). Since unguided problem solving heavily demands the learners' working memory capacity and, thereby, hinders learning (Kirschner et al., 2006; Mayer, 2004), further instructional support, for instance with *scaffolds*, is necessary. The term scaffolding is historically rooted in the interaction between a child and a human tutor (Wood, Bruner, & Ross, 1976). Nowadays, a scaffold is understood as a (computer-supported) instructional means that supports learning processes by focusing the learners' attention on relevant activities, modeling ideal problem solving processes, monitoring, increasing interest, and balancing affective aspects of learning (Pea, 2004; Tabak & Kyza, 2018). By that, scaffolds enable learners to master a task or activity the learners would not be able to solve without assistance, the ultimate goal being that learners internalize the scaffolded processes and develop self-regulated problem solving capacities (Wood et al., 1976). Instructional research has investigated the effectiveness of various kinds of scaffolds, for instance, self-explanation prompts (Heitzmann, Fischer, Kühne-Eversmann, & Fischer, 2015), worked examples (Bichler et al., 2019), or reflection prompts (Mamede & Schmidt, 2017). For supporting collaboration, socio-cognitive scaffolds such as computer-supported collaboration scripts have been found to be effective (Vogel, Wecker, Kollar, & Fischer, 2017). This makes them particularly suitable for facilitating collaborative diagnostic reasoning. Thus, in this thesis, I combine two approaches that are established instructional means in different contexts, namely simulations and computer-supported collaboration scripts. Whereas simulations are prominent in medical education (Cook et al., 2011), collaboration scripts have been used in the context of collaborative learning and mostly outside the medical context (Vogel et al., 2017). Both approaches are now described in more detail.

1.3.1 Agent-based simulations

Since simulations provide a risk free, ethically unobjectionable, and standardized learning environment, they have become particularly prominent in domains in which failure would have severe consequences (Ziv et al., 2003). Examples include pilot training (Hays, Jacobs, Prince, & Salas, 1992), military training (Smith, 2010), or medical training (Siebeck et al., 2011). Simulations have the advantage of providing opportunities of repeated practice for situations that occur rarely or come with a high risk (De Coninck, Valcke, Ophalvens, & Vanderlinde, 2019; Kaufman & Ireland, 2016). Further, it is easy to adapt simulations to the learners' needs by, for instance, increasing or reducing complexity of the simulation (Siebeck et al., 2011) or by providing additional instructional support such as opportunities for reflection, feedback, or prompts (Heitzmann et al., 2019). One assumption about why simulations are effective is rooted in the approaches to *situated cognition* which assume that knowledge is situated in a cultural context (Brown, Collins, & Duguid, 1989). These approaches emphasize the importance of providing authentic learning scenarios to enable learners transfer the learnt skills and knowledge to the professional context of application. For the development of diagnostic reasoning skills, providing opportunities for the application of knowledge seems vital for the encapsulation of pathophysiological knowledge and development of illness scripts (Schmidt & Rikers, 2007). The application of knowledge enables diagnosticians to process symptoms and findings in a clustered way making diagnostic reasoning much more efficient and less error prone (see Section 1.2.1, Lubarsky et al., 2015; Schmidt & Rikers, 2007). A meta-analysis on the effectiveness of virtual patients (i.e., simulated patients) for advancing the learning of diagnostic reasoning supports this line of argumentation by showing that learning with virtual patients has large positive effects on diagnostic reasoning compared to no instruction (Cook et al., 2010). Although there is a large body of data providing evidence for the effectiveness of simulations, there is also tremendous inconsistency between simulations which still warrants explanation (Cook, 2014). One explanatory factor could be the provision of additional instructional support since, for instance, prompts or feedback seemed to increase the effectiveness of virtual patients (Cook et al., 2010). This assumption aligns well with prior findings showing that solving ill-structured problems may overwhelm learners without further guidance (Belland, Walker, Kim, & Lefler, 2017; Kirschner et al., 2006). In general, while learning with simulations seems to be a promising approach for the facilitation of collaborative diagnostic reasoning, a more thorough understanding of the conditions under which learning with simulations is effective is greatly needed.

Simulations have not only been used to advance individual but also collaborative competences using, for instance, role-plays (Kaufman & Ireland, 2016). For example, Gardner and Ahmed (2014) used a virtual patient and role plays to increase performance of a trauma team consisting of 24 health professionals. Indeed, the intervention had positive effects on perceived credibility, perceived coordination, and team performance. Yet, such full-scale team simulations are time consuming and expensive which is why medical students rarely get the opportunity to participate actively, but rather spend much time observing peers participating in simulations (Zottmann et al., 2018). A solution that has been proposed in the context of collaborative problem solving is the use of agent-based simulations (OECD, 2017). Agents (i.e., computers simulating a person) have been used as pedagogical agent to support learning (e.g., AutoTutor; Nye, Graesser, & Hu, 2014) but also to substitute a collaboration partner (OECD, 2017; Rosen, 2015). Although human-to-agent collaboration is often less authentic due to limited conversations (Graesser, Kuo, & Liao, 2017), agent-based simulations can provide the necessary standardization for the training of specific subskills. The significance of standardizing research on collaboration can be illustrated by means of the CDR model. The CDR model (see Section 1.2.3) assumes that individual diagnostic reasoning processes are dynamically shaped by the collaborative processes and the diagnostic reasoning processes of the collaboration partner. For instance, if the collaboration partner shares evidence, a diagnostician ideally considers the new evidence for the own diagnostic reasoning. By standardizing the collaboration partner, instructors, thus, have more control over the dynamic interaction between collaborators. Therefore, in contrast to human-to-human collaboration, in human-to-agent collaboration it is easy to evoke situations in that participants need to show the targeted skill and that would otherwise occur rarely (Rosen, 2015). Moreover, agent-based collaboration allows for a focus on the repeated training of particularly difficult sub-tasks (Ericsson, 2004) and over longer periods of time. In human-to-human collaboration, such a focus would probably undermine the motivation of the respective collaboration partner.

Altogether, simulations seem promising for the learning of collaborative diagnostic reasoning. Further, the use of agents in simulations could provide a well-standardized, flexible, and economic environment for advancing collaborative diagnostic reasoning. Yet, complementing simulations with additional instructional support is necessary to fully exploit the potential of learning with simulations (Chernikova et al., 2020).

1.3.2 Computer-supported collaboration scripts

Collaboration scripts are instructional scaffolds that support collaboration processes among collaborators (Fischer et al., 2013). In the context of computer-supported learning, which includes learning with computer-based simulations (e.g., Roschelle & Teasley, 1995), collaboration scripts have also been called computer-supported collaboration scripts or CSCL scripts (Vogel et al., 2017). As I apply collaboration scripts on learning with computer-based simulations in this thesis, henceforth I use the term collaboration scripts or CSCL scripts to refer to computer-supported collaboration scripts. The notion of collaboration scripts as used in this thesis is based on the script theory of guidance that was introduced in Section 1.2.2. In short, based on the notion of cognitive scripts (Schank, 1999), Fischer and colleagues (2013) assume that people have cognitive scripts consisting of hierarchically structured knowledge guiding when and how to engage in collaborative activities. These cognitive scripts, which are also called *internal* collaboration scripts, are dynamically configured based on the collaborator's goals and the situational characteristics, which is why they allow to flexibly adjust one's behavior in collaborative situations (for a recent introduction see Kollar, Wecker, & Fischer, 2018). As demonstrated above, collaboration requires complex skills and is not always effective (e.g., Tschan et al., 2009). In other words, people often lack functional internal collaboration scripts hindering effective collaboration. Collaboration scripts support dysfunctional internal collaboration scripts by externally providing instruction necessary to engage in successful collaborative activities (Fischer et al., 2013; Vogel et al., 2017). Thus, in principle, collaboration scripts seem a suitable instructional scaffold for facilitating the learning of collaborative diagnostic reasoning.

At this point I would like to interject that collaboration scripts have mostly been used in the context of computer-supported collaborative learning (CSCL, e.g., Dillenbourg & Hong, 2008) and not within collaborative problem solving. From the perspective of cognitive and instructional psychology, instructors use CSCL because collaborating provokes higher-order cognitive activities such as negotiation and explaining, which in turn are beneficial for domain learning (Chi & Wylie, 2014; King, 2007). Thus, whereas in CSCL people collaborate with the goal of learning, in collaborative problem solving people collaborate with the goal to find a solution for a respective problem. Albeit CSCL and collaborative problem solving differ with respect to their goals, I consider the differences between both processes as rather negligible. Remarkably, collaborative learning often implies problem solving, and successful problem solving always implies learning (Dillenbourg, Baker, Blaye, & O'Malley, 1996), and collaboration scripts in both cases aim to advance collaboration. Yet, the focus on domain learning is reflected in the distribution of studies investigating the effectiveness of collaboration scripts for domain learning and the learning of collaboration skills (see Vogel et al., 2017).

The question of whether collaboration scripts are an effective instructional means to advance collaboration has been addressed empirically by several studies. For instance, Noroozi, Teasley, Biemans, Weinberger, and Mulder (2013) developed a collaboration script to support collaborators in developing a transactive memory. In their study, multidisciplinary dyads were asked to solve a problem together. Half of the dyads received sharing and elicitation support by prompting learners to make their expertise and knowledge explicit. The results suggest that supporting participants to share their knowledge helped them to converge their knowledge and to develop better problem solutions. Yet, a study from Rummel, Spada, and Hauser (2009) yielded different results. They supported dyads consisting of one medical student and one psychology student by sequencing their collaboration process and prompting specific collaborative activities (such as sharing of ideas and negotiation) during diagnosing ambiguous patient cases. They found that although scripting affected the learners' collaborative activities (for instance, they descriptively showed better information sharing and time management skills), scripted students did not outperform unscripted students in a free recall test about good collaboration and the quality of the final diagnosis. Vogel et al. (2017) addressed the ambiguity of these results by synthesizing research on the effectiveness of collaboration scripts in a meta-analysis based on a literature search conducted in 2013. They identified 12 studies examining the effect of collaboration scripts on collaboration skills that yielded a combined moderate positive effect. This meta-analysis provides first systematically aggregated evidence for the effectiveness of collaboration scripts. Nevertheless, the study leaves some important questions open, particularly with respect to the suitability of scripting for facilitating collaborative diagnostic reasoning. For instance, since the literature search on which the analyses were based was conducted in 2013, the sample was very small and several more recent studies were not included (e.g., Mende, Proske, Körndle, & Narciss, 2017; L. J. Schmitt & Weinberger, 2019). It is thus unclear whether the general results hold when more recent evidence is taken into consideration. Moreover, the effects vary tremendously between studies indicating that the collaboration scripts were not effective under all conditions. Vogel et al. (2017) addressed this issue by conducting moderator analyses and found some descriptive tendencies. For instance, more detailed collaboration scripts (i.e., scriptlet level collaboration scripts) were descriptively more effective for learning of collaboration skills compared to play or scene level collaboration scripts. Thus, these findings indicate that using a collaboration script on scriptlet-level could be more effective for facilitating the learning of collaborative diagnostic reasoning. Still, this factor leaves a considerable amount of variance unexplained, which is why examining further variables that could explain such heterogeneity seems warranted. One such factor that is difficult to address in a meta-analysis could be goodness of fit of the collaboration scripts for the learners needs (Kollar et al., 2018; Stegmann, Mu, Gehlen-Baum, & Fischer, 2011). The script theory of guidance argues that collaboration scripts should complement less functional or lacking internal collaboration scripts (Fischer et al., 2013). Studies investigating the effectiveness of collaboration scripts have, however, mostly used rather static approaches without assessing the internal collaboration script (Kollar et al., 2018). The few approaches that designed adaptive collaboration scripts typically faded the instruction after a predefined sequence and, thereby, still could have failed to adjust collaboration scripts to the learners needs (Stegmann et al., 2011; Wecker & Fischer, 2011). Thus, the existing evidence for the effectiveness of adaptive support is mixed and has not been investigated in the medical context. Albeit designing adaptive collaboration scripts that automatically assess the learners' internal collaboration scripts seem a promising approach to facilitate the learning of collaborative diagnostic reasoning, further systematical research in standardized contexts is currently absent.

A final issue that is important when intending to use collaboration scripts relates to the widely spread criticism of *over-scripting* (Dillenbourg, 2002; Wise & Schwarz, 2017). The critics argue that collaboration scripts could undermine the agency of scripted collaborators when providing support that does not fit the collaborators' needs and, by that, undermine their motivation to collaborate (Wise & Schwarz, 2017). This critique relates to the assumption that perceived agency is – besides perceived competence and social relatedness – a fundamental driver of human behavior (self-determination theory, Deci & Ryan, 1985). Large negative effects of collaboration scripts on perceived autonomy would be a strong counterargument against the use of collaboration scripts for facilitating collaborative diagnostic reasoning. Yet, there is little empirical evidence on this question. For instance, Rummel et al. (2009) investigated the effect of collaboration scripts on intrinsic motivation but did not find significant differences between scripted and unscripted groups. While descriptively scripted groups felt less pressure and more joy, no difference was found for the subscale perceived choice. Albeit these findings could provide first counterevidence against the over-scripting effect, they are certainly not sufficient, particularly since participants in this study self-reported their motivation after an unscripted collaboration phase that served as posttest which could have influenced their ratings (Rummel et al., 2009). Other studies such as those by Demetriadis, Egerter, Hanisch, and Fischer (2011) or Peterson and Roseth (2016) addressed motivational issues of collaboration scripts but used a more general conceptualization of motivation (e.g., motivation on account of the task or academic efficacy and task value). Hence, they did not directly address the criticism. Nevertheless, Demetriadis et al. (2011) found no significant effects of scripted collaboration on motivation. Peterson and Roseth (2016) found no significant effect on motivation for groups that collaborated synchronously and a small positive effect for groups collaborating asynchronously. Thus, further evidence concerning a plausible, yet not fully evidence-based criticism is necessary.

In summary, empirical studies suggest that collaboration scripts could be a suitable instructional scaffold to facilitate the learning of collaborative diagnostic reasoning (Vogel et al., 2017). Yet, several questions remain to be addressed: firstly, it is not yet clear whether positive effects of collaboration scripts on the learning of collaboration skills are robust when analyzing studies that were conducted recently. Secondly, the conditions under which collaboration scripts are more effective are still not investigated thoroughly (Vogel et al., 2017). Thirdly, the criticism that positive effects of collaboration scripts on learning come at the cost of negative effects on motivation and particular basic psychological needs empirical examination (Wise & Schwarz, 2017).

1.4 Outline of the Studies and the Agent-Based Simulation

The overarching goal of this thesis is twofold: first, it aims to advance the theoretical understanding of collaborative diagnostic reasoning using medicine as application context. Secondly, it seeks to identify conditions under which collaborative diagnostic reasoning can be facilitated effectively by using agent-based simulations and collaboration scripts. To do so, it comprises three studies using three different methodological approaches: a validation study examining the validity of an agent-based simulation (Study 1), a meta-analysis examining the general effectiveness of collaboration scripts (Study 2), and an intervention study examining the effectiveness of collaboration scripts in agent-based simulations (Study 3). In Study 1 and 3 of this thesis, I used an agent-based simulation, which is used as training and assessment instrument. These studies addressed the first goal. In Study 2 and 3, I examined the facilitation of collaborative diagnostic reasoning using collaboration scripts (see Figure 2). All three studies are introduced in more detail below. Next, I provide a concise description and theoretical rationalization of the simulation as used in Study 1 and 3.

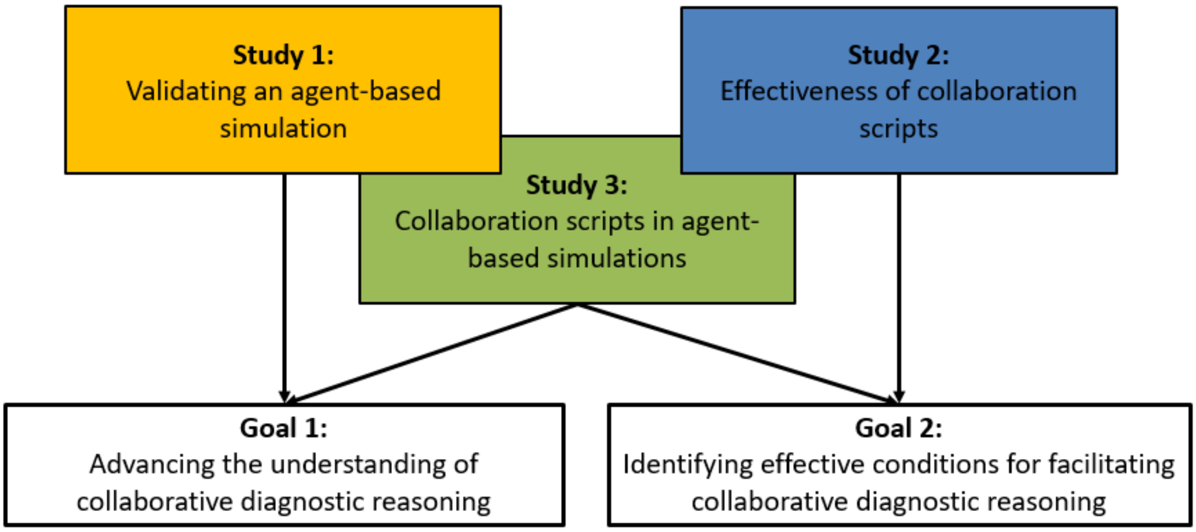

Figure 2. Overview of presented studies and their relation to the thesis' goals

The simulation was developed in collaboration with medical educators, physicians, psychologists, and software engineers. It models a situation in which two physicians with different professional backgrounds (an internist and a radiologist) collaboratively generate evidence for a patient case. This situation was chosen based on interviews with seven radiologists and internists, who suggested that this collaborative situation is highly relevant in the clinical routine, yet oftentimes is carried out ineffectively due to insufficient information sharing. These subjective reports find support in empirical findings suggesting that the sharing of clinical details in radiological requests is often ineffective (Davies et al., 2018).

In the simulation, a medical student acts in the role of an internist working in an emergency department in a hospital. The medical student consecutively receives health records of different patients that contain information about the patient's history, physical examination, and laboratory results. Afterwards, the medical student is asked to collaborate with a radiologist to generate evidence in form of a radiologic test. Thus, the learner requests a radiologic test from a simulated radiologist by choosing the kind of test (evidence elicitation) and sharing patient information (evidence sharing). Test requests that are not adequately justified are rejected from the radiologist. To share patient information effectively with the radiologist, the medical student needs to understand the radiologist's role, task, and responsibilities. Finally, the medical students are asked to document and justify the final diagnosis (for more information about the simulation see Study 1 and Radkowitsch, Sailer, et al., accepted).

In order to identify conditions under which facilitating collaborative diagnostic reasoning is effective, an assessment of the latter is necessary. Traditional approaches for assessing collaboration include quantifying collaborative activities by coding extensive amounts of collaborative dialogue (Andrews-Todd & Forsyth, 2018), rating the quality of specific collaborative aspects such as maintaining a shared understanding or time management based on video data (Meier et al., 2007),using knowledge tests about specific collaborative activities such as argumentation (Stegmann, Weinberger, & Fischer, 2007), or rating the quality of the solution generated during collaboration (Hao, Liu, von Davier, & Kyllonen, 2015). Since the present agent-based simulation is used for training purposes, an obvious alternative for assessing collaborative diagnostic reasoning is to use the agent-based simulation for the assessment as well. This approach also was used successfully by PISA 2015 to assess collaborative problem solving competences in a large-scale study. Their simulation required students to solve problems in collaboration with one or more agents by means of choosing among predefined messages (OECD, 2017). In an attempt to validate this approach, Herborn et al. (2018) compared the OECD's agent-based simulation to a reduced human-to-human collaboration. They found only minor differences between human-to-human and human-to-agent collaboration and in both collaboration modes performance was explained by the same latent factor. Thus, this study provided first evidence that agent-based collaboration could be comparable to human-to-human collaboration and, thus, be not only a suitable tool for *advancing,* but also for *assessing* collaborative problem solving. Besides this empirical finding, several theoretical considerations speak in favor of the use of agent-based simulations that are similar to those brought forward when discussing agent-based simulation as training tool: group-level measurements are highly dependent on group composition with respect to, for instance, knowledge distribution, personality, or motivation (Graesser et al., 2018). Such dependencies could bias the measurement. Thus, human-to-agent collaboration could provide the necessary standardization for assessing collaborative diagnostic reasoning. Furthermore, human-to-human collaboration usually produces a tremendous amount of data (e.g., Andrews-

Todd & Forsyth, 2018) the analysis of which is tedious. In contrast, human-to-agent collaboration requires careful considerations on how to design the interaction between collaborators which later simplifies data analysis considerably (Rosen, 2015), since log files can easily be matched to the previously defined quality standard.

1.4.1 Study 1

As described above, an agent-based simulation is used to facilitate and assess the learning of collaborative diagnostic reasoning. However, using a simulation to assess complex competences warrants thorough collection of validity evidence. This requires one to analyze whether conclusions drawn, and decisions made based on the outcome of the respective instrument are valid (Cook & Hatala, 2016). The first study addresses this issue by providing evidence of validity for the simulation used to assess collaborative diagnostic reasoning. As suggested by Cook and Hatala (2016), I do so by drawing upon a validity framework introduced by Kane (2006). This framework has the advantage that the respective measurement instrument is validated in the context of its intended use. In Kane's (2006) framework, he assumes that when making decisions based on the results of an assessment instrument, four different inferences (namely *scoring, generalization, extrapolation, implications*) are drawn based on underlying implicit assumptions. The validation process consists of the explication of these assumptions as well as the collections of respective warrants. Study 1 comprises a presentation of the processes and outcomes of collecting validity evidence for the agent-based simulation. The validation process is based on the development of the agent-based simulation and on a quasi-experimental study comparing the quality of process and outcome measures of medical students between the 5th and 7th semester, from the 9th semester and higher as well as internists with at least three years of working experience. All participants consecutively solved five patient cases that required the collaboration with a simulated radiologist. After the 2nd and 5th patient case, I further measured the learners' intrinsic cognitive load and their perceived authenticity of the simulation and the collaborative process. The following research questions were pertained in this study:

1) To what extent are the measures of collaborative diagnostic reasoning objective and consistent?
2) To what extent do medical practitioners perceive the simulation as authentic?
3) To what extent do groups with different levels of prior knowledge differ with respect to a) their collaborative diagnostic reasoning b) their diagnostic outcome and c) their reported intrinsic cognitive load?

1.4.2 Study 2

The second study addresses the general effectiveness of collaboration scripts for learning by means of a meta-analysis. As mentioned above, a prior meta-analysis was conducted based on a literature search from 2013 (Vogel et al., 2017). Since then, a number of new studies have been published. This raises the question of whether prior findings hold when using an updated sample. Additionally, CSCL scripts have recently been criticized for undermining learners' agency and thereby reducing learners' motivation. Little is known, however, whether these arguments are consistent with empirical data. Moreover, there is little evidence on the mechanisms that make collaboration scripts effective. The study addresses these questions, by investigating whether collaboration scripts that prompt different collaborative activities or a combination of them differ in their effectiveness. It is based on two literature searches, one in April 2017 and one in January 2020 to identify further studies examining the effectiveness of collaboration scripts. Based on 56 studies involving more than 5,600 participants, using a random effects meta-analysis, the following research questions were addressed in this study:

1) To what extent do collaboration scripts affect domain learning[2], collaboration skills, as well as motivation?
2) To what extent do collaboration scripts differ in their effectiveness depending on which collaborative activities (i.e., information sharing, negotiation, coordination or a combination) they prompt?

1.4.3 Study 3

The third study builds upon the two prior studies and is meant to answer the question under which conditions collaborative diagnostic reasoning can be effectively facilitated by learning with agent-based simulations. More specifically, the study uses an experiment to examine whether adjusting collaboration scripts to the internal collaboration scripts of the collaborator increases their effectiveness (Kollar et al., 2018). Moreover, it was also investigated how these collaboration scripts affect basic psychological need satisfaction. This directly addresses the criticism that collaboration scripts were prone to over-scripting (Dillenbourg, 2002). The study comprises of a one factorial pre-post experimental study in which 160 medical students solved several patient cases within the agent-based simulation. First, participants solved one patient case to assess the participants' prior competence of

[2] Although domain learning is not a variable of major interest in this thesis, domain learning was included as dependent variable in the meta-analysis as it is the main variable of interested in the CSCL community. However, in this thesis, I will not touch upon this variable further.

collaborative diagnostic reasoning. Afterwards, learners solved four further patient cases during which two thirds of the participants were supported with either a static or adaptive collaboration script. Both collaboration scripts were developed based on the literature presented in Section 1.2.2. That means that collaboration scripts included meta-knowledge about the radiologists' task, responsibility, and roles to enable learners to share patient information effectively. Further, based on findings of Vogel et al. (2017), the collaboration scripts were designed on scriptlet level. The third group of participants received no additional support. Afterwards all participants rated their basic psychological need satisfaction and solved another patient case to assess their learning gain. The concrete research questions addressed by the third study were the following:

1) To what extent do static and adaptive collaboration scripts affect collaborative diagnostic reasoning in a medical agent-based simulation?
2) To what extent do static and adaptive collaboration scripts affect basic psychological need satisfaction when learning to diagnose collaboratively with a medical agent-based simulation?

2 Study 1: Learning to Diagnose Collaboratively: Validating a Simulation for Medical Students

Anika Radkowitsch * Martin R. Fischer * Ralf Schmidmaier * Frank Fischer

Reference: Radkowitsch, A., Fischer, M. R., Schmidmaier, R., & Fischer, F. (2020). Learning to Diagnose collaboratively: Validating a simulation for medical students. *GMS Journal for Medical Education, 37*(5), Doc51. doi: 10.3205/zma001344.

Making Diagnoses OPEN ACCESS article

Learning to diagnose collaboratively: validating a simulation for medical students

Abstract

Objectives: Physicians with different professional backgrounds often diagnose a patients' problem collaboratively. In this article, we first introduce a process model for collaborative diagnosing (CDR model), describe the development of a simulation used to empirically examine the facilitation of collaborative diagnostic reasoning. Based on a contemporary validity framework [1], we further suggest indicators for validity and collect initial evidence with respect to the scoring, generalization, extrapolation, and implication inferences to assess the validity of the simulation when used to assess effects of learning interventions.
Method: In a quasi-experimental study, we assessed objectivity and reliability of the simulation and compared medical students with low and advanced prior knowledge to practitioners with high prior knowledge with respect to their diagnostic accuracy, diagnostic efficiency, information sharing skills, and their intrinsic cognitive load. Additionally, we obtained authenticity ratings from practitioners with high prior knowledge.
Results: The results yielded satisfying initial evidence for the validity of the scoring and the extrapolation inferences as ratings are objective, and the simulation and the collaborative process is perceived as rather authentic. Additionally, participants on different levels of prior knowledge differ with respect to their diagnostic accuracy, diagnostic efficiency, information sharing skills, and their reported intrinsic cognitive load. With one exception (information sharing skills), the generalization inference seems to be valid as well.
Conclusions: We conclude that collecting validity evidence for the simulation was an important step towards a better interpretation of the simulation. We found that the simulation is an authentic and valid representation of the chosen collaborative situation and that the collected validity evidence offers sufficient evidence for an initial validation of the simulation. Nevertheless, the validation process highlighted some important gaps that need further consideration. We further conclude that applying a validation model to the context of empirical research is promising and encourage other researchers to follow the example.

Keywords: collaboration, simulation, collaborative diagnostic reasoning, validation

Anika Radkowitsch[1,2]
Martin R. Fischer[1,3]
Ralf Schmidmaier[1,4]
Frank Fischer[1,2]

1 Ludwig-Maximilians-Universität München, Munich Center of the Learning Sciences, München, Germany

2 Ludwig-Maximilians-Universität München, Lehrstuhl für Empirische Pädagogik, Department Psychologie, München, Germany

3 Ludwig-Maximilians-Universität München, LMU Klinikum, Institut für Didaktik und Ausbildungsforschung in der Medizin, München, Germany

4 Ludwig-Maximilians-Universität München, LMU Klinikum, Medizinische Klinik und Poliklinik IV, München, Germany

1. Introduction

In their daily practice, physicians with different professional backgrounds often diagnose patients' problems collaboratively. For example, an internist diagnosing a patient suffering from fever and shortness of breath might consult a radiologist to conduct a CT scan the results of which will be discussed afterwards. In those situations, physicians need to be able to diagnose individually, that means being able to gather and integrate case-specific information with the goal to reduce uncertainty to make a medical decision [2]. But they also need collaborative competences such as sharing of relevant information, negotiation, and coordination skills [3]. A recent review shows that collaborative diagnostic reasoning has been scarcely investigated empirically yet [4]. The available empirical literature demonstrates that physicians often have difficulties to diagnose collaboratively. For example, the quality of the distribution and exchange of information among team members [5] and the experience of team members [6] seem to be key predictors for the quality of collaborative diagnostic reasoning. Such difficulties in information sharing also could affect the quality of subsequent negotiation processes. For instance, if an internist fails to share differential diagnoses and the respective symptoms, the radiologist will have a much harder time to interpret and to discuss the radiologic findings. Offering instructional support to foster collaborative diagnostic reasoning and in particular information sharing, therefore, seems vital. Simulation-based learning is an established method to foster complex competences and its effectiveness has been meta-analytically examined for health professions [7] as well as across domains [8] although it seems that scaffolding beyond mere problem-solving is beneficial for learning [9]. We developed, therefore, a simulation with the goal to identify instructional conditions under which simulations effectively advance collaborative diagnostic reasoning. Importantly, training and assessment of competences presupposes evidence of its validity. We follow Kane's [1] validity framework for the validation of instruments as suggested by Cook and Hatala [10].

In this paper, we want to collect initial evidence for validity of the simulation by constructing a validity argument for a simulation used to conduct experiments on the facilitation of collaborative diagnostic reasoning. For that, we first elaborate on a model of collaborative diagnostic reasoning and describe how simulations can be used to assess and facilitate complex competences. We further explain our validation approach based on Kane's [1] framework as well as validity indicators that are based on theory. Afterwards, we shortly describe the development of our simulation which included several evaluation and revision cycles (cf. [11]). Finally, we present a validation study that was conducted to analyze the validity indicators and discuss the extent to which the results add to our validity argument.

2. Collaborative diagnostic reasoning

Collaborative diagnostic reasoning means to accurately and efficiently diagnose a patient's problem by generating and evaluating evidences and hypotheses that can be shared with, elicited from, or negotiated among collaborators [12]. In the medical and psychological literature, however, diagnosing has been largely conceptualized as individual competence and by using varying terms such as clinical or diagnostic reasoning, clinical decision-making, or clinical problem-solving (e.g., [13], [14]). When diagnosing individually, physicians generate and evaluate evidence based on patient information, weigh the evidence with respect to differential hypotheses and draw conclusions (i.e., make a medical decision) based on the diagnostic process [14], [15]. The quality of individual diagnostic activities is influenced by professional medical strategic and conceptual knowledge [16]. However, more than one diagnostician is often involved in diagnosing a patient or making treatment decisions. For example, in medical consultations a responsible physician calls in the expertise of another health-care professional. Another example are discussion rounds such as tumor boards in which physicians with different professional backgrounds exchange and discuss patient information. In both examples, diagnosticians have the joint goal to make the best clinical decision. When diagnosing collaboratively, the professional medical knowledge, the knowledge about the patient, and outcomes of diagnostic reasoning processes might differ between the diagnosticians. Therefore, collaborative activities are necessary in addition to the individual diagnostic activities to coordinate the individuals' diagnostic processes. Based on the collaborative problem-solving framework by Liu and colleagues [3] and the scientific discovery as dual search (SDDS) model [17], Radkowitsch and colleagues [12] proposed a model for collaborative diagnostic reasoning (CRD model, see figure 1) describing collaborative diagnostic processes with individual and collaborative diagnostic activities. These collaborative activities are sharing, elicitation, negotiation, and coordination. According to the CDR model, evidences and hypotheses generated and evaluated during diagnostic processes are kept in individual diagnostic spaces (dashed lines and boxes). All evidences and hypotheses that are available to all collaborators are represented in shared diagnostic spaces (dotted boxes). For evidences and hypotheses to become part of a shared diagnostic space, the diagnosticians need to conduct the proposed collaborative activities (dotted lines). For example, an internist diagnosing a patient suffering from fever and shortness of breath might generate the hypothesis of pneumonia. In order to reduce the uncertainty of this hypotheses, the internist consults a radiologist to perform a radiologic test. The quality and relevance of the information that the internist shares with the radiologist may influence the hypotheses generated and the conclusions drawn by the radiologist and further affect, which information is shared, negotiated or elicited by the radiologist. In turn, the evidences and hypotheses shared, elicited, or negotiated by the radiologist may influence the internist's individual diagnostic process. Hence, the proposed collaborative activities are considered important for the quality of medical decisions. Based on models and findings on team cognition, we assume that the quality of collaborative activities is influenced by the team members' meta-knowledge [18], [19]. By meta-knowledge we mean the knowledge a team member holds about the other team members' roles, their knowledge, and their task. Meta-knowledge has been shown to particularly influence collaborative activities of collaborators (e.g., [20]). Among collaborative activities, information sharing has received particular attention. Sharing or rather the lack of sharing can affect the accuracy of the diagnoses, but at the same time diagnosticians often fail to share relevant information with others [5], [21].

3. Conducting research on advancing collaborative diagnostic reasoning with simulations

Simulations are an established method to foster competences in medical education as well as in other educational contexts such as teacher trainings [22], pilot trainings [23], or military trainings [24]. In all these contexts, the application of knowledge is a crucial part of professional

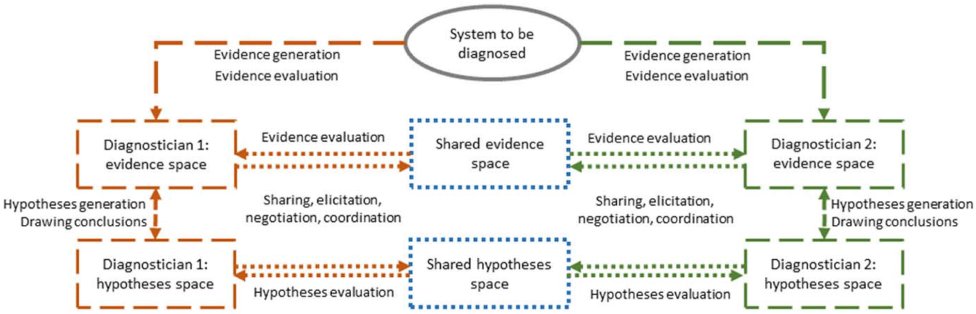

Figure 1: Model for collaborative diagnostic reasoning (CDR) adapted from Radkowitsch et al. [12]

practice [e.g., [25]. Simulations allow to practice the application of knowledge in a risk-free environment [26]. More importantly, however, simulations allow for the deliberate practice [27] of particularly difficult or complex subtasks. That means that within simulations, learners can repeatedly solve (sub-)tasks that they are yet not able to complete. Research on the deliberate practice has shown that this type of practice is particularly crucial during the development of professional expertise [27]. Besides, the application of knowledge in complex domains such as medicine can be overwhelming for learners. To facilitate learning, the complexity of these situation can be reduced in simulations and thereby offer a tradeoff between an approximation-of-practice and authentic representations of real-world situations [28]. Research on the effectiveness of simulations shows positive effects on cognitive, behavioral, and affective learning outcomes in medicine as well as in other domains [7], [8], [29]. However, a recent review shows that to advance diagnostic competences, the provision of additional instructional support beyond the opportunity to solve problems is beneficial [9]. We propose a research agenda to investigate conditions under which diagnostic competences are effectively advanced when learning with simulations [2]. For empirical laboratory research on complex competences it is necessary to focus on empirically measurable aspects. Hence, we focus on information sharing as subskill of collaborative diagnostic reasoning.

When conducting research on the effectiveness of different instructional means, educational research typically uses controlled experiments. That means that two or more groups of learners receive different types of support in an intervention phase. By using unsupported pre- and post-tests, the learning gain of the different groups of learners is assessed [e.g., [30]. The average performance of groups is then compared to identify the effects of the intervention. To realize the proposed research agenda [2], we developed a simulation that will be used in experiments to facilitate but also to assess collaborative diagnostic reasoning, in particular the sharing of information during diagnosing. During the intervention, learners will receive different versions of the simulation. During the pre- and posttest, the simulation will be used to assess the competence levels of groups of learners. Hence, it is an important prerequisite that the simulation differentiates between different competence levels, as well as that the simulation is suitable for the competence level of the targeted group. Using simulations for the assessment of competences is a common approach in medical education [31]. For example, simulations are used to assess procedural skills such as conducting rectal examinations [32], medical communication skills [33], or diagnostic reasoning [34]. When using simulations to assess competences, it is highly relevant that the simulation consists of authentic representations of real-world situations in which the respective competences is typically used [31], [35]. For assessing diagnostic reasoning skills, simulations usually present patient cases for which learners need to come up with the most likely diagnosis [31]. A systematic review on simulations shows that the evaluation of simulations with respect to their validity as assessment tool lacks thoroughness [36]. Therefore, in the present paper we seek to examine whether the simulation developed to realize our research agenda is a valid instrument for the assessment of between group differences of competence levels.

4. Validating a simulation of collaborative diagnostic reasoning: constructing a validity argument

In his validity framework, Kane [1] describes validation as the process to collect and to evaluate validity evidence to judge the appropriateness of interpretations of the results of the assessment. Four typical inferences are drawn when concluding from a test score to a real score which need critical examination with respect to their validity: scoring, generalization, extrapolation, and implications. Each of these inferences are typically based on implicit assumptions that need to be considered during a validation process [10], [37]. In this paper, we explicate these assumptions for the simulation-based assessment of collaborative diagnostic reasoning that has the goal to identify conditions under which collaborative diagnostic reasoning can be effectively facilitated. All considered assumptions as well as their warrants are listed in table

Table 1: Inferences, assumptions, and warrants for the development of the argument of validity

Inferences and assumptions		Warrants
Inference 1	*Scoring: from an observation to a score*	
Assumption 1.1	The observed performances are reliably and objectively scored	The assessment conditions are standardized for all participants (computer-based assessment; identical laboratory conditions). All answers were recorded electronically based on logfiles. All scorings are based on sample solutions or automatic coding developed by experts. The raters show high inter-rater reliability.
Inference 2	*Generalization: from a single score to an overall score*	
Assumption 2.1	Scores on a single performance align with an overall score	All case material was carefully developed by experts based on a blueprint and is structurally identical. All participants complete several structurally identical tasks. All but one measures show a sufficient internal consistency (Cronbach's alpha).
Inference 3	*Extrapolation: from an overall score to the expected target competence/skill*	
Assumption 3.1	The simulation is an authentic representation of a real scenario.	The scenario was carefully chosen based on interviews with practitioners. The case material was developed and evaluated by experts. Practitioners rated the simulation as well as the collaboration as authentic representation of the real-world practice.
Assumption 3.2	The simulation is sensitive to competence differences.	The simulation as well as the case material was carefully developed and evaluated by experts. Participants with different prior knowledge levels differ in their rated performance accordingly.
Assumption 3.3	The simulation affects cognitive processes.	The higher participants prior knowledge, the lower is their indicated intrinsic cognitive load.
Assumption 3.4	Differences between group of learners are due to the intervention.	The simulation will be used in controlled experiments.
Inference 4	*Implications: Interpretations of test results*	
Assumption 4.1	The acquired data can be used to assess the effectiveness of simulations.	Most of the assumptions were positively tested.

1. The first inference, scoring, refers to matching an observation to a single score [38]. For example, in our simulation a medical student proposes a diagnosis for a patient case which is then scored by the experimenter. A valid scoring procedure requires the observations to be correctly transformed into a consistent score and that raters of the accuracy of the final diagnoses show reliable ratings as indicated by high inter-rater agreements (assumption 1.1). The second inference, generalization, refers to generalizing the single score to a test score [38]. In our simulation, we generalize from the information sharing skills shown in one simulated patient case to the information sharing skills shown in several other simulated patient cases. A valid generalization inference is shown, if scores on a single performance (e.g., a final diagnosis of one patient) aligns with an overall score (e.g., all final diagnoses given during the test setting). Hence, high internal consistency of the measures are indicators for plausible extrapolation inferences (assumption 2.1). Extrapolation refers to generalizing from the test score to the real performance [38]. In our simulation, we would hope that medical students who are better in collaborative diagnostic reasoning in our simulation would also be better in collaborative diagnostic reasoning when working with real patients and colleagues. Hence, validity evidence should ideally show that the collaborative diagnostic reasoning of groups of learners shown within our simulations is representative for their collaborative diagnostic reasoning outside the simulation. To ensure that, we propose several validity indicators: First of all, it would be strong evidence for a valid extrapolation inference if experienced practitioners from the field rated the simulation as authentic (assumption 3.1) [35]. We consider experienced practitioners able to judge whether the simulated setting represents real life practices. Secondly, a valid assessment requires that medical practitioners and medical students with high prior knowledge show better test performance (i.e., more accurate and more efficient diagnostic performance) compared to medical students with low prior knowledge (assumption 3.2). The assumption is that on average those showing higher performance in real life settings on average also show higher performance within the simulation. A third validity indicator for the extrapolation inference are differences between persons with different levels of prior knowledge with respect to cognitive load. The cognitive load theory assumes that

learning imposes different kinds of cognitive load on learners. Particularly, the intrinsic cognitive load which is caused by the complexity of the learning material should be lower for people with high prior knowledge compared to less knowledgeable medical students [39]. With higher prior knowledge, the learning material becomes less complex as the material is better cognitively organized and, therefore, imposes less intrinsic load (assumption 3.3). Importantly, to assess the effectiveness of different kinds of simulations, we compare groups of learners rather than individuals. That means that all decisions will be based on group means rather than individual test results. Therefore, a further assumption is that differences between groups of learners result from the intervention and not from random or systematic prior differences between groups (assumption 3.4). Therefore, it is important to use an experimental approach. The final inference, implications, refers to the conclusions drawn, and decisions made based on the test results [1], [10], [38]. Hence, the final assumption is that the resulting data can be used to draw inferences on the effectiveness of different kinds of simulations (assumption 4.1). If the prior assumptions were met, then the implications drawn from the results would be valid.

Considering the intended use of the instrument to be validated is important for the construction of a validity argument as this helps to prioritize the evidence [10]. The intended use of the simulation described in this paper is to assess collaborative diagnostic reasoning of groups of learners in experimental studies. Although every described validity evidence is considered important for the construction of the validity argument, some of the evidences are considered crucial. For our intended use, we argue that particularly the identification of different levels of competence among participants with different levels of prior knowledge would offer the most important validity evidence as this evidence is closest to the final use of the simulation. Although due to content specificity of diagnostic skills, it seems hard to achieve reliable measures in medicine [13], [31], it is particularly important to have coherent measures that allow generalizing from one item to another as this would offer evidence that the same skill is assessed in different items.

5. Research questions of the validation study

Based on the validity framework and the validity indicators described above, we conducted a validation study to answer the following research questions:

1. Scoring: To what extent are the measures of collaborative diagnostic reasoning objective?
2. Generalization: To what extent are the measures of collaborative diagnostic reasoning consistent?
3. Extrapolation:
   1. To what extent do medical practitioners perceive the simulation as authentic?
   2. To what extent do groups with different levels of prior knowledge differ with respect to a) their collaborative diagnostic reasoning (information sharing skills, diagnostic efficiency, and diagnostic accuracy) within the simulation and b) to the reported intrinsic cognitive load?

6. Method

6.1. Development of the simulation to assess collaborative diagnostic reasoning

Our goal is to develop a tool for the assessment of the specific subskills of collaborative diagnostic reasoning as defined above. We chose a simulation-based approach to assess collaborative diagnostic reasoning [7], [8]. As described above, the construct of collaborative diagnostic reasoning is rather broad and can be assessed in a broad range of contexts. For example, different physicians such as internists, surgeons, or gynecologists could collaborate with nurses or other health-related professionals. We assume that the context of collaboration (such as the meta-knowledge about the collaborators' profession) influences collaborative diagnostic processes. We, therefore, decided to narrow down the simulated context to a situation that is relevant in real-world practices and particularly difficult for learners. Hence, we defined the simulated context as a collaborative situation between internists and radiologists based on practitioners' experiences. Interviews with seven practitioners from both disciplines were conducted to identify a specific situation that is considered as being problematic frequently. The interviews yielded that the main problem is unspecific test requests, that is unprecise justifications for the test (e.g., missing relevant patient information) and a lack of clustering of patient information. As a consequence, we decided to focus on information sharing during the request of a radiologic examination as an important and specific aspect of collaborative diagnostic reasoning. Next, we decided to use a computer-based simulation and chose the case-based learning platform CASUS (https://www.instruct.eu/). Computer-based simulations have several advantages compared to other types of simulations such as standardized patients (e.g., [33]). First, the use of the simulation is extremely economical once the material is developed as several participants can interact with the simulation at the same time and, for example, no actors are needed. Secondly, web-based simulations are easily accessible for participants and, hence, time and place restrictions are low. Thirdly, all case material as well as instructions are standardized and, therefore, do not confound the assessment. To develop the simulation, paper prototypes of the scenario and patient cases were constructed and evaluated by an expert committee from medicine, software development, and psychology. Whereas internists, radiologists, and a general practitioner developed the case material for ten patient cases, a software developer programed the simulation. The case

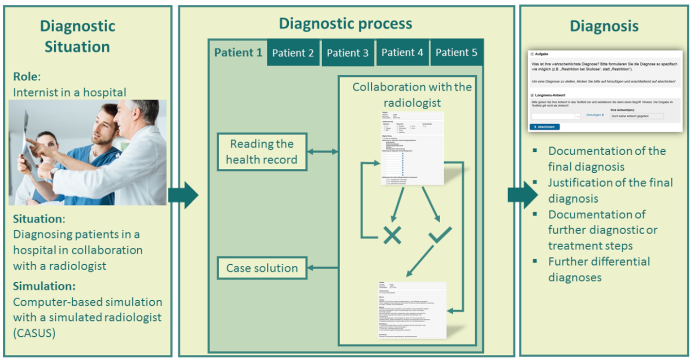

Figure 2: Schematic representation of the simulation

material was then evaluated and revised in a one-day expert-workshop, with focus on the case structure, the most plausible solution, as well as the sample solution. Finally, the simulation was implemented on the CASUS learning platform (see figure 2).

In a pilot study, the simulation with one patient case was presented to eight medical students (M_{age}=24.5, SD_{Age}=3.9; $M_{Semester}$=7.6, $SD_{Semester}$=1.2) to evaluate the user experience of the simulation (UEQ; [40]). Results indicated high values on the subscales attractiveness, perspicuity, stimulation, and novelty, but rather low values on the subscale dependability. To increase the perceived control for participants, a fiction contract containing information about the simulated scenario and the role learners are expected to take up as well as a technical familiarization giving detailed instructions on how to handle the simulation were developed. After having read the fiction contract and the familiarization, participants start the first simulated patient case. Participants first receive a patient file that they scan for symptoms and findings in the role of an internist. The patient file consists of a short patient presentation, medical history, a description of the physical examination, as well as the most important laboratory values. Afterwards, learners request a radiologic test from a simulated radiologist. For that, they are asked to fill in a request form by choosing among 42 different combinations of methods and body parts and by sharing patient information or differential diagnoses that are considered relevant for the radiologist. Only learners who appropriately justified their request (i.e., show high information sharing skills) receive a description of the radiologic findings, and, if provided by the learner, an evaluation of a specific differential diagnose from the simulated radiologist. We decided beforehand with radiologists which information is needed to justify a specific radiologic test. After having read the radiologic result, medical students can ask questions about the radiologic findings, share further information, or request further examinations. To solve the patient case, participants suggest a diagnosis and back it up with justifying findings and suggest further differential diagnoses and treatment or diagnostic measures. For a more detailed description of the simulation and the process of development, see [12]. In sum, in our simulation medical students are supposed to gather and integrate information from a patient file, and to collaboratively generate radiologic evidence by sharing relevant patient information with the radiologist. By that the medical student elicits relevant information from the radiologist, which they then integrate into prior information to arrive at a final diagnosis. Bearing in mind our definition of collaborative diagnostic reasoning, the simulation allows us to separately assess and facilitate both, collaborative diagnostic reasoning (i.e., information sharing) as well as individual diagnostic reasoning (i.e., the final diagnosis).

6.2. Sample and design

A quasi-experimental study with a one-factorial design consisting of three levels (low vs advanced vs high prior knowledge level) was conducted. We defined medical students between the 5^{th} and 8^{th} semester (N=45, N_{female}=31) of a total of 12 semesters as low prior knowledge (PK) (M_{PK}=6.4 semesters, SD_{PK}=0.7) as they had only few courses on internal medicine and radiology according to their study plan. Medical students from the 9^{t}h semester and above (N=28, N_{female}=19) were categorized as advanced prior knowledge (M_{PK}=11.5 semesters, SD_{PK}=1.9) as they already participated in courses for internal medicine and radiology according to their study plan. Internists and residents for Internal Medicine after completion of the 3 years of common trunk (N=25, N_{female}=11) were categorized as high prior knowledge (M_{PK}=13.6 years, SD_{PK}=10.5) as they are expected to have practical experience.

6.3. Procedure

The study was conducted as a laboratory study with a maximum of eight participants at a time. All participants consecutively worked individually on five computer-based patient cases as described above for as long as they wanted. The participants were asked to work efficiently. After the second and the fifth case, participants com-

Table 2: Internal consistencies for all instruments

Instrument	Cronbach's alpha
Diagnostic accuracy	.66
Diagnostic efficiency	.53
Information sharing skills	.33
Perceived authenticity simulation	.85 and .90
Perceived authenticity collaboration	.94 and .95
Intrinsic cognitive load	-

pleted a test measuring perceived authenticity as well as intrinsic cognitive load. Afterwards, participants were debriefed and thanked for their participation with 25€.

6.4. Measures

Within the simulation, we obtained three measures to assess the collaborative diagnostic reasoning: diagnostic accuracy, diagnostic efficiency, and information sharing skills. We used Likert-scaled items to assess the perceived authenticity of the simulation as well as the perceived intrinsic cognitive load (see table 2).

Diagnostic accuracy

The solution of the patient case (i.e., the suggested final diagnosis), differential diagnoses, and further necessary diagnostic or treatment steps were used to score the diagnostic accuracy. Depending on how specific the given diagnosis was, participants received 0, 0.5 or 1 point for each diagnosis and up to one additional point each for the quality of the differential diagnoses and the quality of the indicated further steps. Points were given based on the sample solution that was developed in the expert workshop. The mean diagnostic accuracy across the five patient cases (ranging from 0 to 3) was calculated for each participant.

Diagnostic efficiency

The diagnostic accuracy weighted by the time needed to solve a single patient case indicated the diagnostic efficiency. The mean diagnostic efficiency across the five patient cases was calculated for each participant.

Information sharing skills

The information sharing skills were operationalized as the inverted proportion of requests rejected by the simulated radiologist due to insufficient justification per case. Whether a justification is perceived as sufficient or insufficient by the simulated radiologist was defined beforehand in collaboration with expert radiologists based on how relevant information is for a radiologist to conduct a radiologic test. For this measure, values were obtained directly via the logfiles. The mean score of all five patient cases (ranging from 0 to 1) was calculated for each participant. A mean score of 1 means that all requests in all patient cases were accepted by the radiologist.

Perceived authenticity

The perceived authenticity was assessed with three items each with respect to the overall simulation and with respect to the collaborative process [41] on a 5-point Likert scale ranging from 1 (does not apply) to 5 (does apply). The perceived authenticity of the simulation as well as the authenticity of the collaborative process was assessed twice. An example item for authenticity is "I perceive the [simulation] / [the collaboration with the radiologist] as authentic".

Intrinsic cognitive load

Intrinsic cognitive load was assessed with one item on a 5-point Likert scale ranging from 1 (very easy) to 5 (very difficult) [42]. The item text was "How easy or difficult do you find the collaboration with a radiologist at the moment?".

6.5. Statistical analyses

To answer research question 1, we obtained the intraclass correlation (ICC) based on a two-way random effects model with absolute agreement for the main diagnoses, the differential diagnoses, and the indicated further steps. For that, two raters independently coded 20% of the cases.

To address research question 2, we calculated the internal consistency measure Cronbach's alpha with respect to the diagnostic efficiency, to the information sharing skills, and to the diagnostic accuracy.

To answer research question 3.1., we calculated the mean of both measurement times and contrasted it to a threshold of 3.0 using a one-sample t-test. The means above the threshold indicate that participants with high levels of prior knowledge on average rate the overall simulation and the collaborative process as rather authentic or authentic.

To address research question 3.2., we conducted ANOVAs and Bonferroni post-hoc tests with the independent variable prior knowledge and the dependent variables diagnostic accuracy, diagnostic efficiency, information sharing skill, as well as intrinsic cognitive load. If preconditions for calculating an ANOVA were not met, we con-

Table 3: Means and standard deviations per variable and group.

	Low PK *M* (*SD*)	Advanced PK *M* (*SD*)	High PK *M* (*SD*)
Accuracy	1.03 (0.28)	1.32 (0.31)	1.37 (0.42)
Efficiency	3.70 (1.06)	5.40 (1.64)	6.69 (2.55)
Information sharing	0.65 (0.18)	0.79 (0.11)	0.78 (0.19)
Intrinsic cognitive load	3.54 (0.71)	2.62 (0.70)	2.3 (0.76)

Notes. PK = Prior knowledge

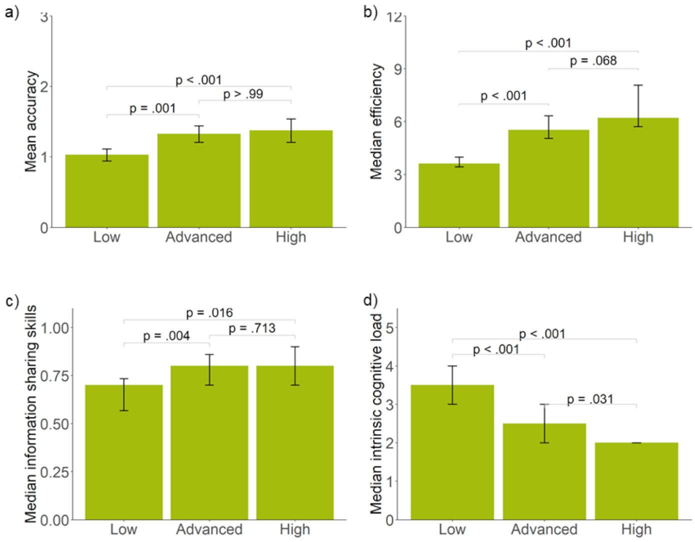

Figure 3: Differences of prior knowledge groups with respect to a) diagnostic accuracy, b) diagnostic efficiency, c) information sharing skill, and d) intrinsic cognitive load. Error bars indicate 95% Confidence Intervals.

ducted the non-parametric Kruskal-Wallis-Test and Wilcoxon post-hoc tests instead. Confidence intervals are calculated with bootstrapping.

7. Results of the validation study

Scoring

With respect to the first research question, we obtained high values for all three variables: The interrater agreement for the quality of the final diagnoses and for the further indicated steps was ICC=1. For the differential diagnoses, the interrater agreement was ICC=0.94. This indicates that raters objectively scored the observations during the simulation.

Generalization

With respect to research question 2, analyses yielded a Cronbach's alpha of .66 for the diagnostic accuracy, a Cronbach's alpha of .53 for the diagnostic efficiency, and a Cronbach's alpha of .33 for the information sharing skills. This indicates that the evidence for the generalization inference being valid is acceptable for the diagnostic accuracy and the diagnostic efficiency but limited for the information sharing skills.

Extrapolation

With respect to research question 3.1., participants with high prior knowledge rated the perceived authenticity of the overall simulation as M=3.89 (SD=0.91) and the authenticity of the simulated collaborative process as M=3.57 (SD=0.91). Both authenticity ratings are significantly above the threshold of 3 ($t(24)$=4.9, p<.01 and $t(24)$=3.14, p<.01). This indicates that, on average, practitioners with high levels of prior knowledge perceive the simulation as rather authentic or authentic. Concerning research question 3.2., see table 3 for the descriptive statistics and figure 3, a-d for between-group comparisons. The results show that the prior knowledge groups differ significantly with respect to the diagnostic accuracy ($F(2,95)$=11.62, p<.001, η^2=0.20). The high and advanced prior knowledge group show significantly higher accuracy than the low prior knowledge group but are not significantly different from each other. However, we found solution rates of up to 0.94 (i.e., the correctness of the final diagnosis) for three of the five patient cases indicating ceiling effects for the final diagnoses. The prior knowledge groups also differ significantly with respect to the diagnostic efficiency ($\chi^2(2)=34.29$, $p<.001$, $\eta^2=0.34$) and with respect to the information sharing skills ($\chi^2(2)=12.48$, $p<.002$, $\eta^2=0.11$). For both outcomes, the high and advanced prior knowledge groups again outperform the low prior knowledge group but do not differ significantly from each other. The prior knowledge groups further differ with respect to the reported intrinsic cognitive load ($\chi^2(2)=38.25$, $p<.001$, $\eta^2=0.38$). The high prior knowledge group reported the lowest intrinsic cognitive load, followed by the advanced, and the low prior knowledge groups. All comparisons are statistically significant.

8. Discussion

The objective of this study was to collect initial validity evidence for the simulation we developed to conduct further experimental research on facilitating collaborative diagnostic reasoning in medical education. The validation of the simulation was based on a theoretical model describing collaborative diagnostic processes (CDR model; [12]). The simulation focusses on one of the proposed collaborative activities, namely information sharing. The CDR model suggests that which information is shared by one diagnostician influences the diagnostic processes of another diagnostician. In case of the simulation, which information is shared by a learner in the role of an internist influences whether a radiologist conducts a radiologic test and how it is interpreted. An argument for initial validity was constructed by applying Kane's [1] validity framework to the context of experimental research based on a simulation. The underlying assumptions were made explicit and supported by warrants (see table 1). However, the strength of these warrants varies between inferences. We were able to show quite clearly that the single observations within the simulation can be assessed objectively as all materials were developed and evaluated by expert committees from different disciplines, and some of the variable scores are generated automatically (scoring). This reduces human errors during the transformation of the observation to a single score. For the variables where coding was necessary, inter-rater reliability was high. We conclude that no further evidence for the validity of the scoring procedure is necessary. Further, we found satisfying validity evidence for the question whether the results of the simulation can be transferred to real-world scenarios by comparing participants with different prior knowledge with respect to their performance and their indicated cognitive load in the simulation (extrapolation). We find that medical students and practitioners with high levels of prior knowledge indeed show higher information sharing skills than medical students with low levels of prior knowledge. This indicates that the simulation enables differentiating between levels of competence of different groups which is the intended use of the simulation. However, there is one exception. We found rather high solution rates for the patient cases, even with students on low levels of prior knowledge, indicating ceiling effects for the case solution included in the measures diagnostic accuracy and diagnostic efficiency. Higher case difficulty would allow to better distinguish between different levels of the competences under consideration which is why case difficulty was increased by adding further distracting information. Nevertheless, it is a recurrent finding in medical education that intermediates and experts do not differ in the accuracy of the diagnoses, but rather in the efficiency with which they come up with the correct solution [31]. An explanation for this effect is that the knowledge of experts is better organized (i.e., encapsulation of knowledge) compared to the knowledge of intermediates. This superior organization of knowledge enables experts to more efficiently come to a correct diagnosis [43]. This pattern of effects is illustrated in our data as the difference between intermediates and experts is descriptively larger for diagnostic efficiency than for diagnostic accuracy. Furthermore, the simulation was rated as rather authentic by practitioners from the field. Ultimately, when conducting experiments with the simulation to compare learning gains of groups of learners, it is of prime importance to additionally rule out prior differences between groups as confounding factors. This could be achieved by randomly distributing learners to experimental groups and by controlling for prior knowledge. Assuming that the simulation is used in randomized experiments, the validation study yielded satisfying evidence for the extrapolation inference. The weakest evidence was found for the assumption that scores from a single observation can be reliably summarized to an overall score (generalization). For two of the three variables of interest (diagnostic efficiency and diagnostic accuracy), the validity evidence is acceptable. For the information sharing skills, we obtained only low internal consistency indicating that across patient cases, learners show varying levels of information sharing quality. One explanation for the generally rather low value might be the small number of observations as the likelihood of higher reliability values increases with the number of observations. Generally, low consistency across different patients is a well-known problem in medical education and is also known as content specificity [13]. That means that the diagnostic accuracy between patient cases correlates poorly (0.1-0.3) [13]. That the consistency across patient cases is particularly low for collaborative diagnostic activities such as information sharing might be explained by the CDR model: Whereas individual diagnostic processes are influenced by medical knowledge, collaborative diagnostic reasoning is further influenced by the professional collaboration knowledge (e.g. meta-knowledge). For example, a student might know which information to share for a patient suffering pneumonia, but not for a patient suffering lung cancer. Hence, the measure for information sharing skill might be affected by both, professional medical content knowledge and professional meta-knowledge about the collaboration partners' discipline. Hence, the presented evidence for the generalization inference, particularly for information sharing skills, of our simulation gives rather limited support for the validity which is why further evidence is necessary.

8.1. Limitations

Of course, the present study is not without limitations that must be considered when interpreting its findings. First of all, the simulation is meant to represent collaborative diagnostic reasoning, however, we focus on a very specific subskill which is the sharing of information in diagnostic situations. This is a narrow focus and the results will not easily generalize to other subskills such as negotiation of differential diagnoses. However, we consider the subskill sharing as a particularly important part of collaborative diagnostic reasoning as prior literature has shown how important and how error-prone the sharing of relevant information is for the field of medicine (e.g., [5], [21]). Similar findings have also been reported in other fields (e.g., [20], [44]). The simulation will be used to scaffold the learning of sharing processes and we are convinced that our findings will be of use in other diagnostic situations in which sharing among diagnosticians is necessary as well.

Additionally, our validity argument is based to a large extent on a comparison between experts and novices. Such comparisons have been criticized as novices and experts differ in several variables which are oftentimes unrelated to the construct under investigation such as the probability of having grey hair ("grey hair index", [45], p. 830). However, we do not intend to argue that the expert-novice comparison shows that we're actually measuring the construct of interest. Instead, we argue that the expert-novice comparison shows that we are able to measure competence differences between groups using the simulation. Also, the intended use of the simulation is not to make judgements about individual competences of learners but rather to compare learning gains of groups to make judgements about the simulation's effectiveness under different instructional conditions. Therefore, we consider the results of comparisons between different levels of prior knowledge as a meaningful contribution to our validation argument.

9. Conclusion

In this article, we presented the collection of initial validity evidences for the simulation which we developed to investigate the facilitation of collaborative diagnostic reasoning – and more particularly information sharing – with simulations. Our validation process allows concluding that the simulation that was developed based on theory is indeed authentic enough with respect to both diagnostic process and collaboration. Importantly, more advanced students and practitioners are more efficient than students in earlier phases of their studies and experience less intrinsic cognitive load. More knowledgeable learners are also better able to interact successfully with the simulated radiologist. Thus, we were able to find initial validity evidence that the simulation can be used to assess whether interventions differ in their impact on the learning of collaborative diagnostic reasoning. With respect to the assessment of the information sharing skills as subcomponent of the collaborative diagnostic reasoning there is, however, a need for improvement concerning the reliability. As the reliability of assessments is considered one of the most important evidence components, this is still an important gap in the validity argument. Refining the measurement and increasing the number of observations might help to close this gap.

Collecting validity evidence about simulations for diagnostic reasoning still seems uncommon [36]. Yet, the construction of a validity argument helped us to understand the strength and weaknesses of the simulation for its intended use. This is an important step and will help us to interpret the results of planned experiments. Besides some gaps in the validity argument that will be addressed further, the simulation is a solid instrument to empirically examine the advancement of collaborative diagnostic reasoning of medical students.

Funding

The research presented in this contribution was funded by a grant of the Deutsche Forschungsgemeinschaft (DFG) to Frank Fischer, Martin R. Fischer and Ralf Schmidmaier (FI 792/11-1).

Competing interests

The authors declare that they have no competing interests.

References

1. Kane MT. Validation. In: Brennan RL, editor. Educational Measurement. Westport: Praeger; 2006. p.17-64.
2. Heitzmann N, Seidel T, Opitz A, Hetmanek A, Wecker C, Fischer M, Ufer S, Schmidmaier R, Neuhaus B, Siebeck M, Stürmer K, Obersteiner A, Reiss K, Girwidz R, Fischer F. Facilitating diagnostic competences in simulations: A conceptual framework and a research agenda for medical and teacher education. Front Learn Res. 2019;7(4):1-24. DOI: 10.14786/flr.v7i4.384
3. Liu L, Hao J, von Davier AA, Kyllonen P, Zapata-Rivera D. A tough nut to crack: Measuring collaborative problem solving. In: Rosen Y, Ferrara S, Mosharraf M, editors. Handbook of Research on Computational Tools for Real-World Skill Development. Hershey, PA: IGI Global; 2015. p.344-359. DOI: 10.4018/978-1-4666-9441-5.ch013
4. Kiesewetter J, Fischer F, Fischer MR. Collaborative clinical reasoning - A systematic review of empirical studies. J Contin Educ Health Prof. 2017;37(2):123-128. DOI: 10.1097/CEH.0000000000000158
5. Tschan F, Semmer NK, Gurtner A, Bizzari L, Spychiger M, Breuer M, Marsch SU. Explicit reasoning, confirmation bias, and illusory transactive memory: A simulation study of group medical decision making. Small Group Res. 2009;40(3):271-300. DOI: 10.1177/1046496409332928

6. Farand L, Lafrance JP, Arocha JF. Collaborative problem-solving in telemedicine and evidence interpretation in a complex clinical case. Int J Med Inform. 1998;51(2-3):153-167. DOI: 10.1016/S1386-5056(98)00112-9
7. Cook DA, Hatala R, Brydges R, Zendejas B, Szostek JH, Wang AT, Erwin PJ, Hamstra SJ. Technology-enhanced simulation for health professions education: A systematic review and meta-analysis. JAMA. 2011;306(9):978-988. DOI: 10.1001/jama.2011.1234
8. Gegenfurtner A, Quesada-Pallarès C, Knogler M. Digital simulation-based training: A meta-analysis. Br J Educ Technol. 2014;45(6):1097-1114. DOI: 10.1111/bjet.12188
9. Chernikova O, Heitzmann N, Fink M, Venance T, Seidel T, Fischer F. Facilitating diagnostic competences in higher education - A meta-analysis in medical and teacher education. Educ Psychol Rev. 2020;32:157-196. DOI: 10.1007/s10648-019-09492-2
10. Cook DA, Hatala R. Validation of educational assessments: A primer for simulation and beyond. Adv Simul (Lond). 2016;1:31. DOI: 10.1186/s41077-016-0033-y
11. Allen M, Sites R. Leaving ADDIE for SAM: An agile model for developing the best learning experiences. East Peoria, IL: ASTD Press; 2012.
12. Radkowitsch A, Sailer M, Fischer MR, Schmidmaier R, Fischer F. Diagnosing collaboratively: A theoretical model and a simulation-based learning environment. In: Fischer F, Opitz A, editors. Learning to diagnose with simulations - Examples from teacher education and medical education. New York, NY: Springer Briefs in Education Series. accepted.
13. Norman G. Research in clinical reasoning: Past history and current trends. Med Educ. 2005;39(4):418-427. DOI: 10.1111/j.1365-2929.2005.02127.x
14. Simmons B. Clinical reasoning: Concept analysis. J Adv Nurs. 2010;66(5):1151-1158. DOI: 10.1111/j.1365-2648.2010.05262.x
15. Fischer F, Kollar I, Ufer S, Sodian B, Hussmann H, Pekrun R, Neuhaus B, Dorner B, Pankofer S, Fischer M, Strijbos JW, Heene M, Eberle J. Scientific reasoning and argumentation: Advancing an interdisciplinary research agenda in education. Front Learn Res. 2014(5):28-45.
16. Schmidmaier R, Eiber S, Ebersbach R, Schiller M, Hege I, Holzer M, Fischer MR. Learning the facts in medical school is not enough: Which factors predict successful application of procedural knowledge in a laboratory setting? BMC Med Educ. 2013;13:28. DOI: 10.1186/1472-6920-13-28
17. Klahr D, Dunbar K. Dual space search during scientific reasoning. Cogn Sci. 1988;12:1-48. DOI: 10.1207/s15516709cog1201_1
18. Wegner DM. Transactive memory: A contemporary analysis of the group mind. In: Mullen B, Goethals GR, editors. Theories of group behavior. New York, NY: Springer; 1987. p.185-208. DOI: 10.1007/978-1-4612-4634-3_9
19. Fiore SM, Rosen MA, Smith-Jentsch KA, Salas E, Letsky M, Warner N. Toward an understanding of macrocognition in teams: Predicting processes in complex collaborative contexts. Human Fact. 2010;52:203-224. DOI: 10.1177/0018720810369807
20. Engelmann T, Hesse FW. Fostering sharing of unshared knowledge by having access to the collaborators' meta-knowledge structures. Comput Hum Behav. 2011;27:2078-2087. DOI: 10.1016/j.chb.2011.06.002
21. Larson JR, Christensen C, Franz TM, Abbott AS. Diagnosing Groups: The Pooling, Management, and Impact of Shared and Unshared Case Information in Team-Based Medical Decision Making. J Person Soc Psychol. 1998;75(1):93-108. DOI: 10.1037/0022-3514.75.1.93
22. Seidel T, Sturmer K. Modeling and Measuring the Structure of Professional Vision in Preservice Teachers. Am Educ Res J. 2014;51(4):739-771. DOI: 10.3102/0002831214531321
23. Hays RT, Jacobs JW, Prince C, Salas E. Flight simulator training effectiveness: A meta-analysis. Mil Psychol. 1992;4(2):63-74. DOI: 10.1207/s15327876mp0402_1
24. Smith R. The long history of gaming in military training. Simul Gaming. 2010;41(1):6-19. DOI: 10.1177/1046878109334330
25. Kolodner JL. An introduction to case-based reasoning. Art Intellig Rev. 1992;6(1):3-34. DOI: 10.1007/BF00155578
26. Ziv A, Wolpe PR, Small SD, Glick S. Simulation-based medical education: An ethical imperative. Acad Med. 2003;78(8):783-788. DOI: 10.1097/00001888-200308000-00006
27. Ericsson KA. Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. Acad Med. 2004;79(10):70-81. DOI: 10.1097/00001888-200410001-00022
28. Grossman P, Compton C, Igra D, Ronfeldt M, Shahan E, Williamson P. Teaching practice: A cross-professional perspective. Teach Coll Record. 2009;111(9):2055-2100.
29. Cook DA, Hamstra SJ, Brydges R, Zendejas B, Szostek JH, Wang AT, Erwin PJ, Hatala R. Comparative effectiveness of instructional design features in simulation-based education: Systematic review and meta-analysis. Med Teach. 2013;35(1):e867-e898. DOI: 10.3109/0142159X.2012.714886
30. Mamede S, van Gog T, Sampaio AM, Delbone de Faria RM, Maria JP, Schmidt HG. How can students' diagnostic competence benefit most from practice with clinical cases? The effect of structured reflection on future diagnosis of the same and novel diseases. Acad Med. 2014;89(1):121-127. DOI: 10.1097/ACM.0000000000000076
31. Schuwirth LW, Van der Vleuten CP. The use of clinical simulations in assessment. Med Educ. 2003;37:65-71. DOI: 10.1046/j.1365-2923.37.s1.8.x
32. Siebeck M, Schwald B, Frey C, Röding S, Stegmann K, Fischer F. Teaching the rectal examination with simulations: Effects on knowledge acquisition and inhibition. Med Educ. 2011;45:1025-1031. DOI: 10.1111/j.1365-2923.2011.04005.x
33. Cömert M, Zill JM, Christalle E, Dirmaier J, Härter M, Scholl I. Assessing communication skills of medical students in objective structured clinical examinations (OSCE) - A systematic review of rating scales. PloS one. 2016;11(3):e0152717. DOI: 10.1371/journal.pone.0152717
34. Helfer RE, Slater CH, Goltz L. Measuring the process of solving clinical diagnostic problems. Med Educ. 1971;5(1):48-52. DOI: 10.1111/j.1365-2923.1971.tb02150.x
35. Shavelson RJ. Assessing business-planning competence using the Collegiate Learning Assessment as a prototype. Emp Res Vocat Educ Train. 2012;4(1):77-90.
36. Cook DA, Brydges R, Zendejas B, Hamstra SJ, Hatala R. Technology-enhanced simulation to assess health professionals: A systematic review of validity evidence, research methods, and reporting quality. Acad Med. 2013;88(6):872-883. DOI: 10.1097/ACM.0b013e31828ffdcf
37. Cook DA, Brydges R, Ginsburg S, Hatala R. A contemporary approach to validity arguments: A practical guide to Kane's framework. Med Educ. 2015;49(6):560-575. DOI: 10.1111/medu.12678
38. Kane MT. Validating the interpretations and uses of test scores. J Educ Measure. 2013;50(1):1-73. DOI: 10.1111/jedm.12000
39. Sweller J. Cognitive load theory, learning difficulty, and instructional design. Learn Instruct. 1994;4(4):295-312. DOI: 10.1016/0959-4752(94)90003-5

40. Laugwitz B, Held T, Schrepp M. Construction and Evaluation of a User Experience Questionnaire. In: Holzinger A, editor. HCI and usability for education and work. Lecture Notes in Computer Science. Berlin; Heidelberg: Springer; 2008. DOI: 10.1007/978-3-540-89350-9_6

41. Schubert T, Friedmann F, Regenbrecht H. The experience of presence: Factor analytic insights. Presence. 2001;10(3):266-281. DOI: 10.1162/105474601300343603

42. Opfermann M. There's more to it than instructional design: The role of individual learner characteristics for hypermedia learning. Berlin, GER: Logos-Verlag; 2008.

43. Charlin B, Boshuizen HPA, Custers EJ, Feltovich PJ. Scripts and clinical reasoning. Med Educ. 2007;41(12):1178-1184. DOI: 10.1111/j.1365-2923.2007.02924.x

44. Kozlov MD, Große CS. Online collaborative learning in dyads: Effects of knowledge distribution and awareness. Comput Hum Behav. 2016;59:389-401. DOI: 10.1016/j.chb.2016.01.043

45. Cook DA. Much ado about differences: Why expert-novice comparisons add little to the validity argument. Adv Health Sci Educ Theory Pract. 2015;20(3):829-834. DOI: 10.1007/s10459-014-9551-3

Corresponding author:
Anika Radkowitsch
Ludwig-Maximilians-Universität München, Lehrstuhl für Empirische Pädagogik, Department Psychologie, Department Psychologie und Munich Center of the Learning Sciences, Leopoldstr. 13, D-80802 München, Germany
anika.radkowitsch@psy.lmu.de

Please cite as
Radkowitsch A, Fischer MR, Schmidmaier R, Fischer F. Learning to diagnose collaboratively: validating a simulation for medical students. GMS J Med Educ. 2020;37(5):Doc51.
DOI: 10.3205/zma001344, URN: urn:nbn:de:0183-zma0013448

This article is freely available from
https://www.egms.de/en/journals/zma/2020-37/zma001344.shtml

Received: *2018-11-11*
Revised: *2020-04-24*
Accepted: *2020-06-29*
Published: *2020-09-15*

3 Study 2: Good for Learning, bad for Motivation? A Meta-Analysis on the Effects of Computer-Supported Collaboration Scripts

Anika Radkowitsch * Freydis Vogel * Frank Fischer

Reference: Radkowitsch, A., Vogel, F., & Fischer, F. (2020). Good for learning, bad for motivation? A meta-analysis on the effects of computer-supported collaboration scripts. *International Journal of Computer-Supported Collaborative Learning, 15*(1), 5-47. doi:10.1007/s11412-020-09316-4

International Journal of Computer-Supported Collaborative Learning (2020) 15:5–47
https://doi.org/10.1007/s11412-020-09316-4

Good for learning, bad for motivation? A meta-analysis on the effects of computer-supported collaboration scripts

Anika Radkowitsch[1] · Freydis Vogel[2] · Frank Fischer[1]

Received: 8 September 2019 / Accepted: 6 March 2020 / Published online: 16 April 2020

Abstract
Scripting computer-supported collaborative learning has been shown to greatly enhance learning, but is often criticized for hindering learners' agency and thus undermining learners' motivation. Beyond that, what makes some CSCL scripts particularly effective for learning is still a conundrum. This meta-analysis synthesizes the results of 53 primary studies that experimentally compared the effect of learning with a CSCL script to unguided collaborative learning on at least one of the variables motivation, domain learning, and collaboration skills. Overall, 5616 learners enrolled in K-12, higher education, or professional development participated in the included studies. The results of a random-effects meta-analysis show that learning with CSCL scripts leads to a non-significant positive effect on motivation (Hedges' $g = 0.13$), a small positive effect (Hedges' $g = 0.24$) on domain learning and a medium positive effect (Hedges' $g = 0.72$) on collaboration skills. Additionally, the meta-analysis shows how scaffolding single particular collaborative activities and scaffolding a combination of collaborative activities affects the effectiveness of CSCL scripts and that synergistic or differentiated scaffolding is hard to achieve. This meta-analysis offers the first counterevidence against the widespread criticism that CSCL scripts have negative motivational effects. Furthermore, the findings can be taken as evidence for the robustness of the positive effects on domain learning and collaboration skills.

Keywords CSCL scripts · Collaboration skill · Domain learning · Meta-analysis · Motivation

✉ Anika Radkowitsch
Anika.Radkowitsch@psy.lmu.de

1 Department of Psychology and Munich Center of the Learning Sciences, Ludwig-Maximilians-Universität München, Munich, Germany

2 Learning Sciences Research Institute, School of Education, University of Nottingham, Nottingham, UK

Introduction

During collaborative learning, learners engage in collaborative activities that are expected to facilitate deep elaboration of the learning material, such as explaining, argumentation, or identifying and negotiating socio-cognitive conflicts (King 2007). Engaging in such activities should enhance learners' acquisition of knowledge and skills (Chi 2009). However, without guidance, learners often have difficulties engaging in the expected collaborative activities, and thus fail to take advantage of collaborative learning opportunities. For instance, learners do not engage in higher-level argumentation (Bell 2004) or do not take unshared information into account (Stasser and Titus 1985). As an explanation of why collaborative learning falls short of expectations, King (2007) argues that learners might have a limited conception of how to interact with each other. In particular, interacting with others in learning environments based on educational technology, such as MOOCS or computer-mediated interactions among remote learners, creates difficulties for learners unfamiliar with these new learning settings (Fischer et al. 2013).

However, while educational technology requires learners to engage in unfamiliar collaborative learning settings, it also makes it possible to shape and scaffold collaborative learning (Dillenbourg et al. 2009). For example, scripts for computer-supported collaborative learning (CSCL scripts) have been designed to provide just-in-time scaffolds to structure and sequence collaborative learning activities. CSCL scripts repeatedly engage learners in beneficial collaborative practices, eventually supporting the development of collaboration skills and domain learning (Fischer et al. 2013). However, studies on the effectiveness of CSCL scripts show heterogeneous results for domain learning and collaboration skills. In the light of the relevance of guidance for collaborative learning in computer-supported learning environments such as MOOCs and blended learning environments, and the diversity and complexity of previous findings, summarizing efforts seem warranted.

In a qualitative approach to summarizing CSCL research, Wise and Schwarz (2017) collected widespread views of CSCL among experts in the field and developed them into thought-provoking positions. One result of this effort holds that CSCL scripts are prone to overscripting by including prescriptions for users' interactions that are too specific. This leads to a high risk of undermining learners' self-determination and agency and thus reduces intrinsic forms of motivation to fully engage in collaborative learning activities (Deci and Ryan 1985). They further argue that positive effects, if they exist at all, are restricted to the enhancement of collaboration skills. In a recently conducted quantitative approach, a meta-analysis of the effects of CSCL scripts found a significant small effect of CSCL scripts on domain learning and a significant large effect on collaboration skills (Vogel et al. 2017). These results partially support the argument that CSCL scripts mainly affect the learning of collaboration skills; however, they also show that despite possible overscripting effects, CSCL scripts can be effective for domain learning as well. Yet it is still an open question whether the effectiveness of CSCL scripts is decreased by the hypothesized negative influence on motivation (Dillenbourg 2002).

Apart from this, however, the recent meta-analysis on the effects of CSCL scripts did not sufficiently clarify how and under which circumstances CSCL scripts are effective for learning, particularly for domain learning. Although some moderators were tested, a significant amount of variance remained, leaving the question of what makes CSCL scripts particularly effective mostly unanswered. Moreover, several new studies on CSCL scripts have been published since this meta-analysis was conducted in 2013. Thus, a quantitative analysis of the effects of CSCL scripts beyond their influence on domain learning and collaboration skills seems necessary at this stage. Thus, the present meta-analysis contributes to answering the question of whether CSCL scripts negatively influence learners' motivation, whether the effects of CSCL scripts on domain learning and collaboration skills remain robust when integrating more recent CSCL script studies, and to what extent specific features of CSCL scripts explain their effectiveness.

Scaffolding collaboration using CSCL scripts

In their script theory of guidance, Fischer et al. (2013) argue that collaborative learners' failure to engage in high-level collaborative processes is an indicator of a lack of internal collaboration scripts. Internal collaboration scripts are flexible cognitive structures consisting of knowledge about specific collaborative practices. In line with dynamic memory theory (Schank 1999), internal collaboration scripts consist of knowledge components about situations that are flexibly stored in a person's memory and activated in a more or less likely sequence depending on a person's goals and situational characteristics. This general knowledge structure can be dynamically changed, allowing for spontaneous reactions to situational changes (Schank 1999). If learners do not have well-functioning internal collaboration scripts for a specific learning situation, they will engage in processes that are less beneficial for learning, and hence not take full advantage of the given learning environment. To support learners in overcoming dysfunctional internal collaboration scripts, the information needed to engage in beneficial collaborative learning processes can be provided externally. For this purpose, CSCL scripts scaffold learners towards collaboration in a specific context by affording specific activities (e.g., argumentation) in a likely sequence (e.g., first reading a text, then formulating arguments based on the text), and implicitly or explicitly distributing roles among learners (e.g., pro and con positions). That means that CSCL scripts can be understood as scaffolding for the social interactions necessary for collaborative learning (Kollar et al. 2006).

Like other types of scaffolds, CSCL scripts support learners in solving tasks they would not be able to solve without the scaffold (Wood et al. 1976). Therefore, CSCL scripts are expected to be most effective when they fit the learners' prior knowledge and skills (i.e., the internal script they have available, Fischer et al. 2013). What distinguishes CSCL scripts from other types of scaffolds is the particular knowledge and skills targeted by the different scaffolds. Other scaffolds such as self-explanation prompts provide help on a content-related level to support the development of individual problem-solving processes in a specific domain (e.g., Heitzmann et al. 2015). Such scaffolds address conceptual, metacognitive, strategic, or motivational aspects of individual problem-solving (Belland et al. 2017). They particularly target the development of individual problem-solving skills to enable learners to solve similar problems in the specified domain on their own. In contrast, CSCL scripts are scaffolds that support collaborative learning and collaborative problem-solving processes such as coordination, sharing of information and ideas, or negotiation (Liu et al. 2015). Instead of directly addressing the content specific to the learning domain, CSCL scripts mainly target the development of skills necessary to interact in a social learning setting or to solve problems collaboratively. By that, CSCL scripts indirectly facilitate domain learning by engaging learners in beneficial activities (King 2007). Hence, CSCL scripts facilitate both collaboration skills and domain specific knowledge and skills.

To explain the organization of CSCL scripts, Fischer et al. (2013) identify four hierarchical script components: *play*, *scene*, *scriptlet*, and *role*, with the play component representing the highest level. The play component encompasses knowledge about the general task (e.g., argumentation), including knowledge about each individual's role and the sequence of scenes. The scene component consists of knowledge about situations making up a given scene (e.g., formulating a counterargument to an argument that has been stated). The scriptlet component consists of knowledge about activities and their sequence within a specific scene (e.g., when formulating an argument, first develop a claim, then support the claim with data and warrants). Finally, the role component consists of knowledge about participants' specific roles during a scene (e.g., participants in an argumentation with different positions). External CSCL scripts can provide scaffolding at all four script component levels. With their optimal external scripting level principle, Fischer et al. (2013) address the overscripting effect by identifying a mechanism to counteract overscripting. The principle suggests that learners benefit most from external CSCL scripts that address the hierarchical level whose subordinate components are already available to the learners. For example, if a learner already knows how to formulate a sound argument including data and warrants, but doesn't know how to respond to a collaboration partner's argument with a counterargument, the CSCL script should prompt the learner how to produce counterarguments. Taking the optimal external scripting level into account, a CSCL script could lead to overscripting if it addresses internal script components on a level the learners are already able to activate. For instance, if the learners already know that raising an argument involves formulating a claim, reasons and warrants, then a detailed prompt to formulate these components of an argument would not be optimal. In this case, the CSCL script might reduce learners' autonomy, decreasing their motivation and distracting them from using the functional internal script components they already have available (Dillenbourg 2002; Fischer et al. 2013).

In sum, collaborative learning requires complex skills, which are cognitively represented in internal scripts learners partially share socially. Often, it seems that learners do not have a functional internal script for a specific situation available. This leads to the use of less functional collaborative activities in collaborative learning situations. Hence, learners fail to engage in high-quality collaboration and thus do not take full advantage of the collaborative learning situation. External CSCL scripts are an instructional means that may compensate for erroneously activated internal script components or a lack of internal collaboration scripts and thus help learners engage in social interactions that lead to the enhancement of both domain learning and collaboration skills. CSCL scripts are specified by the target of social processes and can be distinguished from other scaffolds that target content-related processes in the learning domain.

Effects of CSCL scripts on motivation

External CSCL scripts have been criticized for constraining learners during collaboration and thereby undermining learners' motivation (Dillenbourg 2002). Motivation comprises all factors that are important for the selection, initiation, and maintenance of actions (e.g., Heckhausen 1974) and is hence highly important for individual learning (Urhahne 2008). The critique mentioned above is related to phenomena best explained by self-determination theory (SDT). In SDT, Deci and Ryan (1985) describe the experience of *basic psychological needs* as a key determinant of human behavior and emphasize the importance of the social environment for individual motivation (Ryan and Deci 2008b). The basic psychological needs are *autonomy*, *competence*, and *relatedness*. Autonomy refers to whether a specific behavior exhibited by an individual is perceived as congruent and volitional. Competence concerns the perceived efficacy of one's behavior. Relatedness refers to one's personal feeling of connection with a specific community and the significance thereof (Ryan and Deci 2008a). These three basic psychological needs are seen "as a nutrient essential for psychological growth, integrity, and wellness" (Ryan and Deci 2008a, p. 657). These needs are considered universal, and humans naturally seek to fulfill them by exhibiting or avoiding certain behaviors. This theory has received particular attention in educational contexts (e.g., Krapp et al. 2014). It is assumed that learners are autonomously motivated and thus engage in learning activities in learning environments that promote autonomy, perceptions of competence, and social relatedness. In contrast, learners experience controlled motivation or even amotivation and thus only superficially engage in learning activities if their basic psychological needs are not appropriately addressed (Ryan and Deci 2008a).

With respect to external CSCL scripts, this leads to two conceivable scenarios: On the one hand, the highly coercive nature of external CSCL scripts might diminish learners' perceived autonomy during the collaboration process and thus reduce motivation to learn collaboratively (Dillenbourg 2002). On the other hand, external CSCL scripts might enhance learners' perceived competence by enabling them to experience success early on. This might increase learners' motivation to stay engaged in the collaborative learning situation, reduce negative effects of unequal participation (i.e., social loafing and sucker effects; Latané et al. 1979; Schnake 1991) and thus further enhance motivation to learn collaboratively (Weinberger et al. 2009). While the most prominent articles criticizing CSCL scripts for undermining learners' motivation are conceptual (Dillenbourg 2002; Wise and Schwarz 2017), only a few studies have examined the effect of CSCL scripts on motivation empirically. For example, Peterson and Roseth (2016) found a non-significant negative effect on motivation when comparing learners in a synchronous CSCL environment learning with a CSCL script to those learning without a CSCL script. However, in the same study, a positive effect of a CSCL script on motivation was found for students in an asynchronous learning environment. In a similar vein, Demetriadis et al. (2011) found that learners supported by an external peer-review script experienced higher motivation from the tasks and the collaboration itself and were less motivated by grading than learners who were not supported by an external CSCL script.

To summarize, SDT suggests that in order for learners to be autonomously motivated, the learning environment should support their perceptions of autonomy, competence, and relatedness. Although external CSCL scripts have been criticized for hampering learners' motivation (e.g., Wise and Schwarz 2017), research examining how external CSCL scripts influence basic psychological needs and, thus also, motivation is still inconclusive. The present meta-analysis addresses this ambiguity by quantitatively synthesizing existing findings on the effect of CSCL scripts on motivational outcomes for the first time.

Effects of CSCL scripts on domain learning and collaboration skills

Several studies examining the effects of CSCL scripts on learning have been published in the last two decades (e.g., Choi et al. 2005; Tsovaltzi et al. 2015). Past research syntheses on CSCL scripts have shown that they have significantly higher effects on learners' collaboration skills than on domain learning (Vogel et al. 2017). A reason for this difference in effect sizes might be related to the cognitive resources available for the internalization of collaborative activities and the elaboration of content-related knowledge and skills based on these activities. During beneficial social interactions, learners co-construct knowledge that is internalized and integrated by each individual learner (Teasley 1997; Weinberger et al. 2007). The most beneficial interactions typically include higher cognitive processes (Chi and Wylie 2014). Collaboration skills are internalized by the repeated practice and engagement in collaborative activities scaffolded by CSCL scripts, while domain learning is enhanced by the cognitive elaboration induced by the collaborative activities a CSCL script requires the learners to engage in. Thus, learners' limited cognitive resources must be divided among both the internalization of collaboration skills and the elaboration of domain knowledge based on these activities. Learners with less elaborate internal collaboration scripts who are scaffolded with external CSCL scripts might still need to put a higher amount of cognitive effort into the collaborative task, leaving fewer cognitive resources available for the elaboration of domain knowledge (Kirschner et al. 2018). This might lead to higher effects of CSCL scripts on collaboration skills than on domain learning. Overall, CSCL scripts seem to positively affect domain learning and the learning of collaboration skills. It seems plausible to assume that the effect of CSCL scripts on domain learning depends on how functional a learner's internal collaboration scripts are.

Effectiveness of different collaborative activities scaffolded by CSCL scripts

Although the meta-analysis from Vogel et al. (2017) yielded important insights to the extent to which CSCL scripts affect learning, it only allows for limited conclusions on the question of *how* CSCL scripts affect learning. On a general level, it seems plausible to assume that CSCL scripts foster domain learning and collaboration skills by engaging learners in functional collaborative activities. Learners internalize the knowledge and skills needed to perform such collaborative activities by developing or modifying their internal collaboration scripts (Fischer et al. 2013). These collaborative activities (e.g., negotiation) are typically associated with higher-order cognitive processes that are beneficial for domain learning (Chi and Wylie 2014; King 2007). Vogel et al.'s (2017) meta-analysis approached the question of how CSCL scripts affect learning by analyzing whether scripts asking learners to engage in transactive activities are more effective than scripts asking learners to engage only in activities that are not transactive. During transactive activities, two or more learners elaborate on each other's ideas. In the best case, transactive activities result in new outcomes that neither learner would have come up with alone (Teasley 1997). Based on the ICAP framework (Chi and Wylie 2014) the collaborative activities expected to be most beneficial for learning are those in which learners mutually build on each other's contributions while cognitively processing the given learning material(Teasley et al. 2008). Such transactive activities should lead to a deeper elaboration of the learning material and hence higher domain learning gains (Teasley 1997). Vogel et al.'s (2017) meta-analysis compared CSCL scripts that did and did not scaffold transactive activities during collaboration. Their descriptive results indicated significant positive effects only for CSCL scripts scaffolding transactive activities on domain learning, but detrimental effects on collaboration skills. The differences between the effects of CSCL scripts scaffolding and not scaffolding transactive activities were not significant and left a substantial amount of unexplained variance. Since transactivity encompasses a large number of more specific collaborative activities (e.g., answering questions, expressing critiques, synthesizing different arguments, completing others' ideas), it seems to be reasonable to analyze more thoroughly how scaffolding more specific collaborative activities might explain the effectiveness of CSCL scripts. In doing so, we apply the collaborative problem-solving framework of Liu et al. (2015), which describes three different social activities that are needed for successful collaborative problem solving. CSCL scripts can support learners in engaging in these social activities, thus enhancing both domain learning and the learning of collaboration skills.

For collaborative learning to be successful, learners must demonstrate, alongside cognitive competences such as task-specific and domain-specific problem-solving skills, the skills to engage in social activities (e.g., Tschan et al. 2009; Zhuang et al. 2008). In their collaborative problem solving framework, Liu et al. (2015) propose three different social activities that are beneficial for learning: sharing ideas or information, negotiating, and regulating or coordinating problem solving. Sharing ideas or information refers to the skill needed for exchanging unshared information or ideas in a way that is appropriate for the collaboration partners. Information sharing is assumed to be beneficial for constructing and maintaining a shared representation and understanding of the problem (Roschelle and Teasley 1995) as well as for better decision making and problem solving (Lu et al. 2012). Negotiating refers to the learner's ability to criticize, elaborate on, and question their learning partners' contributions (Liu et al. 2015). This type of collaborative activity has been shown to particularly enhance individual understanding and learning (Chi and Wylie 2014; Noroozi et al. 2013c; Teasley 1997). Coordinating problem solving includes meta-cognitive activities such as goal setting, evaluating strategies, or regulating motivation and emotions (Liu et al. 2015). According to Järvelä and Hadwin (2013), coordination of collaborative activities is essential for efficient, effective, and enjoyable learning. It is worth mentioning that this framework refers to collaborative problem solving in general without explicitly mentioning computer-supported collaboration. However, collaborative problem-solving approaches are increasingly used in CSCL research as they are more general and, thus suitable for systematic reviews. Furthermore, it is assumed that the same skills apply to learning in CSCL environments. However, learners must additionally be able to handle and coordinate the technological component of learning in CSCL. In some CSCL environments, learners communicate using communication technology such as audio or video conferences or an asynchronous chat program (e.g., a forum). If the discourse is limited due to missing eye contact or lack of opportunities to interpret facial expressions and gestures, more explicit communication strategies such as turn-taking must be used to maintain successful collaboration (Rummel et al. 2009).

In short, the collaborative activities described here seem essential for collaborative learning, and successful collaborative learners ought to have corresponding "functional configuration of internal script components" (Fischer et al. 2013, p. 61). Providing learners with external CSCL scripts allows them to overcome less functional configurations of internal collaboration scripts (Fischer et al. 2013). Consequently, the stimulation of specific collaborative activities using CSCL scripts can be considered a crucial mechanism for the effectiveness of CSCL scripts in facilitating learning. Therefore, we seek to systematically examine to what extent stimulating these collaborative activities explains effects of CSCL scripts on domain learning and the enhancement of collaboration skills.

Effectiveness of combining scaffolds for different types of collaborative activities

The application of all of the aforementioned collaborative activities are perceived to be beneficial for domain learning. Learners who engage in all of these collaborative activities during a collaborative problem-solving task exhibit a higher task performance compared to learners who engage in only some or none of them (Andrews-Todd and Forsyth 2018). Assuming that learners lack corresponding functional internal scripts, it seems to be reasonable to design CSCL scripts with specific scaffolding for each of these collaborative activities. Hence, the development of the corresponding collaboration skills as well as domain learning are needs that are addressed by CSCL scripts.

Tabak (2004) suggests a framework for the distribution of different scaffolds for one need or for different needs. She describes the combination of different scaffolds targeting different needs as differentiated scaffolding, while different scaffolds targeting the same need can also be combined deploying their maximal strengths in the form of synergistic scaffolding. Applying Tabak's (2004) arguments to scaffolding different collaborative activities with CSCL scripts, differentiated and synergistic scaffolding could be relevant. For differentiated scaffolding, the proximate needs that are scaffolded using Liu et al.'s (2015) framework are the specific corresponding collaboration skills. Here, a scaffold targeting one collaborative activity exclusively meets the need to enhance the corresponding collaboration skill and not others. Thus, the effects of a CSCL script combining scaffolds for different collaborative activities should be as high as when the scaffolds were individually applied. Also, these scaffolds should not negatively interfere with each other (Tabak 2004).

However, scaffolding the different collaborative activities also addresses one joint need, namely domain learning. For domain learning, combining different scaffolds would ideally lead to synergistic scaffolding, as the scaffolds mutually increase each other's effectiveness (Tabak 2004). In the case of synergistic scaffolding, the effect of combined scaffolds on domain learning would thus exceed the effects of CSCL scripts scaffolding each collaborative activity individually. However, existing experimental studies with CSCL scripts combining different scaffolds indicate that synergistic scaffolding is hard to achieve and the combination of scaffolds can even decrease their effectiveness (Kollar et al. 2014; Noroozi et al. 2013b). A potential reason for the loss in effectiveness may be that combinations of different scaffolds can be overwhelming, particularly for weak learners (Schwaighofer et al. 2017). However, these studies focused on combining different CSCL scripts or combining CSCL scripts with other types of scaffolds. It is unclear whether these results are transferable to different types of activity prompts within a single CSCL script. Thus, it remains an open question whether CSCL scripts should scaffold different collaborative activities in a combined way within the same learning situation or scaffold one activity at a time in order to achieve the greatest benefits for domain learning. We assume that collaboration skills are acquired by repeatedly engaging in collaborative activities for practice, as described above (Fischer et al. 2013). In light of this, both the individual as well as the combined scaffolding of collaborative activities could be beneficial for learning. On the one hand, learners might benefit from combined scripting for different collaborative activities, as several skills necessary for successful collaboration are addressed simultaneously (Tabak 2004). On the other hand, learners might need to deliberately practice a single collaborative activity, particularly when the intervention period is short. Therefore, the meta-analysis at hand seeks to clarify to what extent scaffolding collaborative activities separately or in combination lead to better effects of CSCL scripts on domain learning and collaboration skills.

Goals and research questions

The first goal of this meta-analysis is to meta-analytically address the hypothesis that CSCL scripts are prone to overscripting and hence to reducing learners' motivation. The second goal is to assess whether the positive effects of learning with CSCL scripts on domain learning and collaboration skills found by Vogel et al. (2017) hold when examining an updated and extended sample of primary studies on learning with CSCL scripts. In addition, we seek to distinguish between the effects of CSCL scripts on different outcomes related to domain learning and collaboration skills. The third goal is to test the proposed mechanisms for how CSCL scripts facilitate domain learning by stimulating collaborative activities to support the internalization of functional internal script components. Particularly, we are interested in how stimulating specific types of collaboration skills exclusively and in combination might differentially affect the effectiveness of CSCL scripts on domain learning. Based on the literature presented above, we propose the following research questions.

RQ 1: What is the overall effect of collaborative learning with CSCL scripts compared to unstructured collaborative learning on motivation?

RQ 2: What is the overall effect of collaborative learning with CSCL scripts compared to unstructured collaborative learning on domain learning and collaboration skills?

RQ 3: To what extent do the effects of CSCL scripts differ if they prompt one, two, or three different types of collaborative activities?

RQ 4: To what extent do the effects of CSCL scripts on domain learning and collaboration skills differ between CSCL scripts prompting a) negotiation, b) information sharing, and c) coordination individually, and in combination with each other?

Method

Criteria for inclusion

To be included in the present meta-analysis on the effects of CSCL scripts on motivation, domain learning, and collaboration skills, primary studies had to meet the following criteria:

Independent variable (1): Only (quasi-)experimental studies with scaffolding collaborative learning by means of a CSCL script as an independent variable (i.e., comparing groups of learners scaffolded with a CSCL script to groups of learners learning without a CSCL script) were included in the meta-analysis. More specifically, CSCL scripts as an instructional means were defined as follows. The scaffolding of a CSCL script must target the collaborative process of at least two learners. Additionally, at least one activity scaffolded by the CSCL script should ask learners to address their learning partner(s). Typically, a CSCL script supports the collaborative processes by structuring the collaborative task into sub-tasks, for example, by prompting specific activities in a specific order. In these sub-tasks, learners address their learning partners by, for instance, explaining, discussing, or asking questions. Some CSCL scripts also assign roles to the participants. However, scaffolds that only implicitly assigned roles to learners (for example, by distributing learning materials among students to induce knowledge interdependence) and did not include any further scaffolds targeting the collaborative process, were not defined as CSCL scripts (e.g., Molinari et al. 2009). Comparisons of two different CSCL scripts (e.g., Rummel et al. 2009), comparisons of CSCL scripts with other instructional means such as example-based learning (e.g., Rummel and Spada 2005), with individual learning (e.g., Peterson and Roseth 2016), or with waiting conditions were not considered in this meta-analysis. Hence, the only difference between the experimental group and the control group is the provision of the CSCL script in the experimental group.

Dependent variable (2): The present meta-analysis seeks to identify the effect of CSCL scripts on motivation and learning outcomes, specifically domain learning and collaboration skills. Therefore, only studies reporting at least one of these dependent variables were included in the analysis. We are primarily interested in the effect of CSCL scripts on motivation. Self-report questionnaires that assessed motivational factors (e.g., motivation to participate) based on self-determination theory or expectancy-value models were categorized as motivation. If several subscales for motivation were reported, all subscales were included but treated as statistically dependent. Negatively framed subscales (e.g., tension and pressure subscale in the Intrinsic Motivation Inventory) were inverted before being included in the analyses. We defined domain learning as knowledge and skills in the subject domain students learned about during the intervention. For example, if learners analyzed student cases based on attribution theory during the intervention, tests assessing knowledge and application of attribution theory were categorized as domain learning (e.g., Stegmann et al. 2007). We defined collaboration skills as knowledge and skills that are needed to engage in collaborative activities. For example, argumentation or knowledge about argumentation, as well as knowledge about the general procedure of collaboration were categorized as collaboration skills. All dependent variables were required to be assessed after and independently of the treatment and could be assessed either on the individual or on the small group level. Hence, performance measures assessed during the intervention were not included in the analyses.

Context of the study (3): Only studies using computer-supported learning settings with groups of at least two learners were included in the meta-analysis. Within this, technological means could be used to facilitate learners' communication, the learners' documentation of learning process and/or learning outcome, as well as the CSCL script could be provided by technological means.

Study design (4): Only studies that reported data from an experimental or a quasi-experimental design were included in the meta-analysis, i.e., studies needed to compare a treatment condition supported by CSCL scripts with a control condition not supported by CSCL scripts.

Availability of data (5): Only studies that reported enough data to calculate an effect size were included in this meta-analysis. To calculate Hedges' g, information on the mean, standard deviation, and number of participants in the experimental group and the control group are needed. Alternatively, F, t, r, or $\chi 2$ statistics can be used to estimate Hedges' g, if the size of both groups is known. However, before excluding studies reporting insufficient data, the authors were contacted in order to obtain missing data.

Article quality (6): Only articles that were published in English and in peer-reviewed journals were included in the meta-analysis. Publication in peer-reviewed articles is seen as an indication of quality. Additionally, published articles are accessible to a broad scientific community, which enables fellow researchers to replicate meta-analytic findings. Furthermore, articles published in conference proceedings were excluded as they often are a first version of articles later published in peer-reviewed journals. To avoid the inclusion of unidentifiable duplicates, these articles were not considered in the meta-analysis.

Literature search

Two comprehensive literature searches were conducted in April 2017 (first literature search) and in January 2020 (second literature search). For each literature search, we searched the databases ERIC and ISI Web of Science using the search terms "(scaffold* OR script*) AND (learn* OR know*) AND (collaborat* OR cooperat*) AND (computer* OR CSCL OR techno*)" and used a snowball system.

During the first literature search, the coding for inclusion was conducted in two steps for both the database search and the snowball system. The first step was based on the title and abstract of the article, while the second step was based on the full text. The database search resulted in 624 articles, excluding duplicates. As the search and inclusion criteria are based on the work of Vogel et al. (2017), all studies that were excluded during the literature search of Vogel et al. (2017), were also excluded from the present meta-analysis. This step left a total of 418 new studies. Then, based on a coding scheme, the titles and abstracts of all articles were coded for eligibility based on the aforementioned inclusion criteria 1, 3, and 4. The articles were coded as relevant or irrelevant by two independent coders, with double coding for a 10 % sample (Cohen's Kappa = .86). In the second step, the full texts of the 71 articles coded as relevant in the first step were obtained and assessed for eligibility based on inclusion criteria 1 to 6. Two of the authors double-coded the full texts independently and discussed any disagreements, ultimately resulting in 32 eligible articles. For the snowball system, the reference lists of eligible articles published in 2015 or later were scanned for the search terms to identify additional articles that were not found by the database search. This procedure resulted in 34 articles, which were then obtained as full texts and coded for inclusion by both coders independently. Nine additional articles were found to be eligible by both coders (Cohen's Kappa of the full text coding was 0.60). In total, the literature search in 2017 resulted in 41 articles.

For the second literature search in January 2020, the same key words and databases were used to identify articles that were published in 2017 and later. The second literature search yielded 202 articles (including 7 systematic reviews), for which the titles and abstracts were coded based on inclusion criteria 1, 3, and 4. Additionally, we screened the full texts and reference lists of the seven newly identified systematic reviews for additional relevant studies that had not been detected by the literature search. The database search and the snowball system resulted in a total of 76 articles (44 from the databases and 32 from the snowball system) that were included in the full text coding and coded for eligibility based on inclusion criteria 1 to 6. Of these, 12 articles were found to be eligible for the meta-analysis (9 from the databases and 3 from the snowball system). In the second literature search, all coding was independently conducted by two coders. All disagreements were discussed until agreement was reached.

Description of sample

In total, the literature search resulted in 52 articles reporting 56 relevant studies. Three studies were excluded after data extraction, as they reported duplicate data from studies that were also included in the sample (Bollen et al. 2015; Mäkitalo et al. 2005; Noroozi et al. 2013a). The total sample includes 49 articles reporting 53 relevant studies involving 5616 participants (M = 102.3, SD = 118.5). The articles were published between 2000 and 2020. The participants' age ranged from 6 to 64. Most studies (k = 36) analyzed participants enrolled in higher education, 16 studies analyzed participants enrolled in primary or secondary school, and one study analyzed participants enrolled in professional training. In most studies, participants learned content related to the natural sciences (k = 17) followed by the social sciences (k = 16). Other subjects were the humanities (k = 4), medical science (k = 3), computer science (k = 7), and other topics (k = 6).

Coding of studies

A coding scheme describing the outcome variables and their specific components, as well as the type of collaborative activity (i.e., scripting for negotiation, information sharing, and coordination) was developed. The coding scheme contained definitions of the relevant variables, detailed descriptions of the coding rules and the coding procedure, and examples for each variable and its categories. The CSCL script for each study was categorized for each type of collaborative activity according to whether or not it included prompts for that specific collaborative activity. Two of the authors double coded 30% of the articles resulting from the first literature search. The same two persons coded half of the remaining studies each. For all variables, the coding was cross-checked by the respective second coder and differences were discussed in order to reach consensus. Following the second literature search, all studies found to be eligible for inclusion were double coded, and disagreements were discussed until consensus was reached. During the coding of the articles resulting from the second literature search, we refined the descriptions of the collaborative activities in the coding scheme. Hence, all codes from the first literature search were screened again by two coders in order to incorporate the refinements. The individual effects and categorization of each study regarding the moderator levels and dependent variables are reported in Table A4. For a summary of the co-occurrence of different moderator levels and dependent variables within studies, see Table A5.

Dependent variable: For the outcome variable domain learning, we distinguished between knowledge assessed in *recall tests* and *application tests*. Recall tests typically ask learners to reproduce knowledge that was learned beforehand without making inferences themselves. This knowledge typically consists of facts, definitions, or theories. Examples of recall tests are multiple choice tests and recognitions tests. In contrast, *application tests* ask learners to apply domain knowledge to a given problem. Examples of application tests are problem solving tasks such as the application of a theory to a case scenario. For example, learners are asked to apply their knowledge about force and motion to calculate a car's braking distance. Instruments that consisted of both types of measures and did not report separate results were categorized as *mixed tests*. For the outcome variable collaboration skills, we distinguished between the assessment of *negotiation skills*, *information sharing skills*, *coordination skills,* and *mixed skills* based on the collaborative problem solving framework by Liu et al. (2015). We categorized instruments that assessed the quality of skills related to negotiation and argumentation, such as the quality of discussion or argumentation sequences, as negotiation skills. Measures that assessed the quality of information sharing, such as the formulation of individual arguments, were defined as information sharing skills. Measures that assessed the quality of meta-cognitive activities, such as planning and monitoring collaborative strategies, were classified as coordination skills. Finally, measures that assessed components of more than one skill were categorized as mixed skills.

Negotiation scripting: We coded whether the CSCL script prompted or facilitated negotiation among participants. We coded CSCL script activities as negotiation scripting if they were related to argumentation and negotiation, such as resolving conflict, reaching a consensus, compromising, formulating critiques, discussing content-related aspects, and engaging in argumentation sequences. However, prompts about how to construct an individual argument were not classified as negotiation but rather as information sharing prompts.

Information sharing scripting: We assessed whether the CSCL scripts prompted information sharing among collaborators. Information sharing prompts are defined as prompts and hints related to the interchange of information, knowledge, ideas or problem solutions. Examples of scripting information sharing are recognizing differences in the knowledge base of oneself and one's learning partner, initiating and requesting information sharing processes, and constructing individual arguments without engaging in a discourse.

Coordination scripting: We defined activities prompted by the CSCL script as coordination scripting if they related to the management of group processes or taking on responsibility. For example, learners might be explicitly prompted to coordinate among themselves by discussing learning goals or strategies, or the CSCL scripts replaces such activities by distributing roles or sequencing the learning process into different phases.

Combinations of collaborative activities: Based on the coding for each specific collaborative activity, we further coded a) the number of different collaborative activities prompted by the CSCL script (i.e., one, two, or three different collaborative activities), and b) the specific combinations of collaborative activities or single collaborative activity prompted by the CSCL script. The different types of combinations were negotiation and information sharing, negotiation and coordination, information sharing and coordination, a combination of all three collaborative activities, or only one of the three collaborative activities.

Statistical analyses

Hedges' *g* (Hedges 1981) was chosen as a statistical index for all effect sizes reported in this meta-analysis. Furthermore, a random-effects meta-analysis was conducted in which studies were weighted according to their precision defined by the inverse of their variance (Borenstein et al. 2009), as we assume that the true effect size varies between studies, as the treatment (i.e., the CSCL scripts), experimenters, task, sample composition, and other partially unknown covariates all vary between studies. Some effect sizes were statistically dependent because studies reported multiple outcomes for the same sample, or several experimental groups were compared to one control group. We addressed such dependencies by calculating a synthetic effect size for statistically dependent effects (Borenstein et al. 2009). To identify heterogeneity within the data, the ratio of the observed variance between studies to the within study error (Q) was calculated. The ratio of estimated true variance to total observed variance was reported using the I^2 statistic. Based on suggestions by Higgins et al. (2003), the amount of heterogeneity is considered low if 25–49% of the observed variance is due to estimated true variance, moderate if 50–74% is estimated true variance, and high if at least 75% of the observed variance is estimated true variance. The moderator analysis was conducted as proposed by Borenstein et al. (2009): A meta-regression with categorical moderator variables was conducted first. This tests the potential moderators by indicating whether Hedges' *g*s for each level of a moderator differ significantly from each other. In a second step, Hedges' *g*, its confidence interval, and the amount of heterogeneity within the level were calculated for each moderator level.

This meta-analysis was based on published articles. To detect and correct for a possible publication bias, trim'n'fill (Duval and Tweedie 2000) and the three-parameter selection model (3PSM, McShane et al. 2016) were used.

All analyses were conducted in R for Windows 3.6.2 (R Core Team 2019). The metafor (Viechtbauer 2010) package was used to aggregate the comparisons and to calculate the effect sizes. According to Cohen (1988), effect size values ranging between 0.20 and 0.49 are interpreted as small effects, effect size values between 0.50 and 0.79 are interpreted as medium effects, and effect size values of 0.80 or larger are interpreted as large effects.

Results

Summary effect of CSCL scripts on motivation

In total, the included studies reported 33 effects of CSCL scripts on motivation, of which 9 effects were independent. The analyses yielded a non-significant positive effect with a Hedges' *g* of 0.13 (see Table 1).

The 3PSM resulted in an adjusted Hedges' *g* of 0.14 ($SE = 0.10$, 95% *CI* [−0.06; 0.35], $p = .17$), which is not significantly different from the unadjusted Hedges' *g* ($X^2(1) = 0.03$, $p = .86$). The trim'n'fill method resulted in 2 missing studies on the right side and suggested an adjusted significant Hedges' *g* of 0.16 ($SE = 0.07$, 95% *CI* [0.01; 0.31], $p = .03$; see Fig. 1a). These results indicate that CSCL scripts do not negatively affect motivation and that the results might be underestimated due to publication bias.

Summary effect of CSCL scripts on domain learning

In total, we examined 124 effects of CSCL scripts on domain learning, of which 54 were independent. Overall, the meta-analysis resulted in a small significant positive summary effect, with Hedges' *g* of 0.24. When distinguishing the particular types of domain learning, CSCL scripts showed small significant positive effects on both recall tests (Hedges' $g = 0.27$) and application tests (Hedges' $g = 0.21$). Studies reporting mixed tests for domain learning yielded a positive but non-significant summary effect of Hedges' $g = 0.09$ (see Table 2).

Concerning the publication bias analyses for the summary effect of CSCL scripts on domain learning, the 3PSM resulted in an adjusted Hedges' *g* of 0.21 ($SE = 0.06$, 95% *CI* [0.08; 0.33], $p < .01$). The Likelihood Ratio Test yielded no significant difference between the adjusted and the unadjusted model ($X^2(1) = 0.37$, $p = .54$). Trim'n'fill suggested 2 missing studies on the left side of the distribution (see Fig. 1b), with a corrected Hedges' *g* of 0.22 ($SE = 0.04$, 95% *CI* [0.14; 0.31], $p < .01$). These results indicate that CSCL scripts positively affect the recall and application of domain learning, although the results could have been slightly influenced by publication bias.

Table 1 *Summary Effect Size and Mean Effect Sizes for Moderator Levels for the Effect of CSCL Scripts on Motivation*

				95% CI		Test for heterogeneity				
	k	Hedges' *g*	*SE*	*LBD*	*UBD*	Q_B	*df*	*p*	*Tau*²	*I*²
Summary Effect	9	0.13	0.07	−0.01	0.27	14.37	8	.073	0.018	42.77%

* $p < .05$; ** $p < .01$

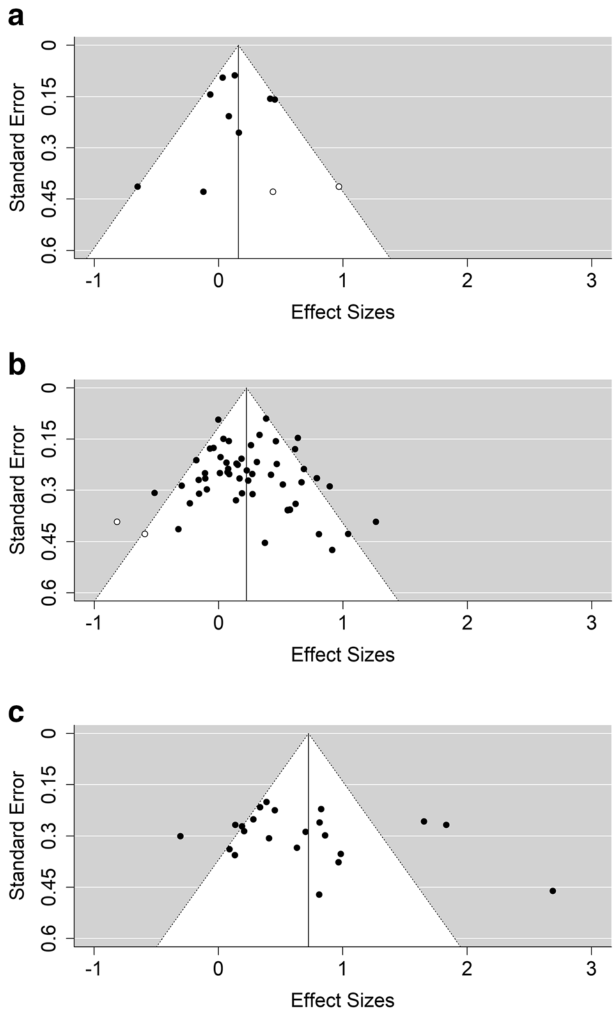

Fig. 1. *a-c.* Funnel plots for the effect of CSCL scripts on a) motivation, b) domain learning, and c) collaboration skills.

Summary effect of CSCL scripts on collaboration skills

We examined 41 effects of CSCL scripts on collaboration skills. Of these, 23 were independent. The meta-analysis resulted in a significant medium-sized positive summary effect of Hedges' $g = 0.72$. The effect sizes for two studies (Stegmann et al. 2012; Weinberger et al. 2010) were noticeable larger than the average effect size. Therefore, all analyses were conducted again excluding these studies. The resulting significant positive summary effect without these studies was also of medium size (Hedges' $g = 0.59$). When distinguishing between the different types of collaboration skills measures, the results again revealed significant positive effects of CSCL scripts on each of the three measures. For the effect of CSCL scripts on negotiation skills, the analyses resulted in a medium-sized Hedges' g of 0.59. For information sharing skills, the analyses resulted in

Table 2 *Summary Effect Size and Mean Effect Sizes for Moderator Levels for the Effect of CSCL Scripts on Domain Learning*

				95% CI		Test for heterogeneity				
	k	Hedges' *g*	*SE*	*LBD*	*UPB*	Q_B	*df*	*p*	*Tau*²	*I*²
Summary Effect	54	0.24**	0.04	0.16	0.33	101.77	53	<.001	0.040	46.23%
Recall test	32	0.27**	0.07	0.14	0.40	76.10	31	<.001	0.069	57.92%
Application test	30	0.21**	0.07	0.08	0.34	62.59	29	<.001	0.063	55.32%
Mixed test	3	0.09	0.12	−0.14	0.33	0.59	2	.746	0	0%
Number of collaboration prompts	58									
One: recall test	11	0.25*	0.11	0.02	0.46	32.11	10	<.001	0.073	64.04%
One: application test	10	0.44**	0.15	0.15	0.73	25.07	9	.003	0.129	67.75%
Two: recall test	5	0.25	0.25	−0.24	0.74	13.93	4	.008	0.222	73.33%
Two: application test	7	0.17*	0.08	0.004	0.33	6.96	6	.325	0.005	10.69%
Three: recall test	13	0.16*	0.08	0.01	0.31	14.48	12	.271	0.018	26.54%
Three: application test	12	0.25**	0.10	0.06	0.44	21.96	11	.025	0.053	48.97%
Combinations of collaborative activities	61									
Negotiation	8	0.22	0.13	−0.02	0.47	18.72	7	.009	0.072	67.26%
Information sharing	11	0.34*	0.14	0.07	0.61	24.11	10	.007	0.117	58.60%
Coordination	3	0.46**	0.12	0.22	0.70	3.27	2	.195	0.016	35.43%
Negotiation & information sharing	4	0.43**	0.16	0.11	0.75	2.10	3	.552	.007	6.26%
Negotiation & coordination	2[a]	–	–	–	–	–	–	–	–	–
Information sharing & coordination	8	0.15	0.08	−0.01	0.30	12.04	7	.099	0.010	19.40%
Three-way combination	25	0.20**	0.06	0.08	0.31	36.63	24	.048	0.026	34.26%

* $p < .05$; ** $p < .01$, [a] Not estimated due to a small number of studies

a medium sized Hedges' *g* of 0.60 (outliers excluded). For mixed measures of collaboration skills, the analyses resulted in a medium-sized Hedges' *g* of 0.52 (see Table 3).

Table 3 *Summary Effect Size and Mean Effect Sizes for Moderator Levels for the Effect of CSCL Scripts on Collaboration Skills*

				95% CI		Test for heterogeneity				
	k	Hedges' *g*	*SE*	*LBD*	*UBD*	Q_B	*df*	*p*	*Tau*²	*I*²
Summary Effect	23	0.72**	0.14	0.44	1.01	102.26	22	<.001	0.378	82.21%
Summary Effect[b]	21	0.59**	0.11	0.36	0.81	68.34	20	<.001	0.188	71.13%
Negotiation skills	10	0.59**	0.17	0.26	0.91	30.15	9	<.001	0.187	69.55%
Information sharing skills	12	0.92**	0.28	0.38	1.47	71.06	11	<.001	0.890	88.67%
Information sharing skills[b]	10	0.60**	0.20	0.21	1.00	38.53	9	<.001	0.296	76.92%
Coordination skills	0[a]	–	–	–	–	–	–	–	–	–
Mixed skills	5	0.52*	0.21	0.11	0.93	10.77	4	.029	0.369	63.36%
Number of collaboration prompts	27									
Number of collaboration prompts[b]	25									
One	10	1.09**	0.30	0.51	1.67	53.45	9	<.001	0.722	86.44%
One[b]	8	0.71**	0.20	0.32	1.10	23.59	7	.001	0.205	67.81%
Two	7	0.67**	0.22	0.23	1.11	17.58	6	.007	0.235	69.07%
Three	10	0.52**	0.18	0.18	0.87	37.50	9	<.001	0.239	76.81%

* $p < .05$; ** $p < .01$, [a] Not estimated due to a small number of studies, [b] Estimated excluding outlier

The 3PSM for the overall effect of CSCL scripts on collaboration skills (including outliers) resulted in a medium-sized adjusted Hedges' g of 0.76 ($SE = 0.22$, 95% CI [0.33; 1.18], $p < .01$) which is not significantly different from the unadjusted Hedges' g ($X^2(1) = 0.05$, $p = .83$). In line with this result, the trim'n'fill suggested no missing studies (see Fig. 1c). The 3PSM for the overall effect of CSCL scripts on collaboration skills (excluding outliers) resulted in a medium-sized adjusted Hedges' g of 0.62 ($SE = 0.18$, 95% CI [0.27; 0.97], $p < .05$), which is not significantly different from the unadjusted Hedges' g ($X^2(1) = 0.05$, $p = .82$). Trim'n'fill suggested no missing studies, indicating no publication bias. These results indicate that CSCL scripts positively affect collaboration skills, with no evidence of publication bias.

Moderator effects of number of different collaborative activities scripted

To answer the question of whether scaffolding a combination of different types of collaborative activities influences the effectiveness of CSCL scripts for recall and application tests, we compared CSCL scripts prompting one, two, or three different types of collaborative activities (RQ 3). For this analysis, only the number of different types of collaborative activities was relevant; we did not distinguish between the specific types of collaborative activities scaffolded by the CSCL script.

For domain learning, the effectiveness of CSCL scripts did not significantly depend on the number of different collaboration prompts ($Q(5) = 3.19$, $p = .67$). The amount of residual heterogeneity was significant ($Q(52) = 114.51$, $p < .01$). Descriptively, the highest effects were found for CSCL scripts prompting solely one type of collaborative activity (either information sharing, negotiation, or coordination), with higher effects for application tests compared to recall tests. The positive effect of CSCL scripts prompting two different types of collaborative activities was only significant for application tests, not for recall tests. CSCL scripts that prompted all three types of collaborative activities had a small significant positive effect on both recall tests and application tests (see Table 2).

The effect of CSCL scripts on collaboration skills did not depend significantly on the number of collaborative activities prompted ($Q(2) = 2.87$, $p = .24$). The amount of residual heterogeneity was significant ($Q(24) = 108.54$, $p = < .01$). When excluding the outliers, the effect of CSCL scripts on collaboration skills also did not depend significantly on the number of distinct collaborative activities prompted ($Q(2) = 0.55$, $p = .76$). The amount of residual heterogeneity was significant ($Q(22) = 78.67$, $p = <.001$). Descriptively, the results show that CSCL scripts prompting only one type of collaborative activity were most effective, followed by CSCL scripts prompting two and three types of collaborative activities (see Table 3).

Moderator effects of specific combinations of collaborative activities

To answer the question of whether scripting certain specific collaborative activities and combinations thereof affect the effectiveness of CSCL scripts, we compared CSCL scripts that scripted only negotiation, information sharing, or coordination to CSCL scripts that scripted the respective two-way or three-way combinations (RQ 4). For the effect of CSCL scripts on collaboration skills, the number of studies did not allow for conducting a moderator analysis with such a large number of moderator levels. Therefore, we conducted this analysis only for domain learning outcomes.

For domain learning, the effect of CSCL scripts did not depend significantly on the specific combination of collaborative activities ($Q(6) = 5.84$, $p = .44$). The amount of residual heterogeneity was significant ($Q(54) = 100.75$, $p = < .01$). Descriptively, the results show that for studies scaffolding only one type of activity, scaffolding coordination is most effective for enhancing domain learning, followed by scaffolding information sharing, while scaffolding only negotiation resulted in a non-significant effect size. When scaffolding a combination of two different collaborative activities, CSCL scripts scaffolding both negotiation and information sharing led to the second highest positive effect size, while the combination of information sharing and coordination is least effective, with a null effect on domain learning. CSCL scripts with a combination of all three types of collaboration skills resulted in the second smallest, yet still significant small effect on domain learning (see Table 2).

Discussion

The present meta-analysis of studies comparing collaborative learning supported by a CSCL script with unstructured collaborative learning did not yield evidence for a strong negative effect of learning with CSCL scripts on motivation. This finding contradicts the repeatedly formulated hypothesis that learning with CSCL scripts might be too coercive, reducing learners' autonomy and thus leading to a loss of motivation compared to externally less structured collaborative learning (e.g., Dillenbourg 2002; Wise and Schwarz 2017). Moreover, the meta-analysis shows that CSCL scripts have a small positive effect on domain learning and a medium to large positive effect on learning collaboration skills. Here, the inclusion of more recent primary studies confirms the results of a previous meta-analysis on CSCL script studies (Vogel et al. 2017). Within domain learning, CSCL scripts have a stronger effect on recall tests than on knowledge application tests. These findings are in line with theoretical assumptions and empirical findings on learning and transfer in CSCL contexts (Jeong et al. 2019). Within collaboration skills, the differences between negotiation and information sharing skills were rather small. Remarkably, no studies investigated the effect of CSCL scripts on coordination skills.

To analyze the mechanisms that are assumed to be responsible for the effectiveness of CSCL scripts, we compared CSCL scripts that prompted one, two, or three different collaborative activities as well as specific combinations of these activities based on a collaborative problem solving framework (Liu et al. 2015). Most studies prompted either one or all three collaborative activities. Generally, the results show that prompting only one specific activity tends to result in higher effect sizes compared to prompting a combination of two or three collaborative activities. However, this was more strictly the case for collaboration skills than for domain learning. For domain learning, combining two collaborative activities yielded even smaller effects than combining all three collaborative activities. When conducting more differentiated analyses comparing different combinations of particular collaborative activities, we found that particularly the combination of information sharing and coordination led to a small effect size. Among CSCL scripts that prompted only one collaborative activity, coordination scripting yielded the largest effect on domain learning. Among CSCL scripts prompting two collaborative activities, CSCL scripts combining negotiation and information sharing scripting were most effective. For collaboration skills, detailed analyses on the effect of scripting different combinations of collaborative activities were not possible due to the small number of studies. The results show, however, that scripting a greater number of different collaborative activities reduces the effectiveness of CSCL scripts for learning collaboration skills. However, the moderator analyses did not lead to a substantial reduction in heterogeneity between the CSCL script studies. Therefore, it is still necessary to identify other mechanisms that are relevant for the effectiveness of learning with CSCL scripts beyond the mechanisms proposed and analyzed in the present meta-analysis.

Overall effect of learning with CSCL scripts on motivation (RQ 1)

The first research question addressed the effect of collaborative learning with CSCL scripts compared to unstructured collaborative learning on motivation. In the light of the frequent critique that CSCL scripts might be too coercive and thus reduce learners' autonomy and motivation, it could be hypothesized that learning with CSCL scripts should have a negative effect on motivation (Dillenbourg 2002; Wise and Schwarz 2017). In contrast, the results showed a small positive but non-significant effect of learning with CSCL scripts on motivation with all studies either reporting non-significant or significant positive effects. Taking a closer look at the primary studies reporting effects of CSCL scripts on motivation enabled us to identify some patterns within the data. Interestingly, most studies considering effects on motivation used CSCL scripts that distributed roles among participants. There were, however, striking differences between the studies yielding positive and null effects. Most studies reporting non-significant effects used CSCL scripts that distributed roles among participants such as tutor and tutee or note-taker roles (G.-Y. Lin 2020; Peterson and Roseth 2016; Weinberger et al. 2005). In these cases, the participants were all undergraduate students. However, some CSCL scripts could even positively affect learners' motivation. For instance, a CSCL script that did not affect motivation in synchronous learning settings, had positive effects on motivation in in asynchronous learning settings (Peterson and Roseth 2016). Other motivating CSCL scripts applied a "natural" role distribution resulting from true knowledge interdependency (e.g. collaboration between psychologists and physicians, Rummel et al. 2009), or distributed roles with different responsibilities among school pupils (Taylor and Baek 2019). In the latter case, the pupils were more motivated when the roles rotated between learning sessions than when fixed roles were used. It seems plausible that undergraduate students have higher prior knowledge regarding collaboration than school students and therefore perceive such artificial roles as more constraining and disruptive of their natural collaboration than fourth and fifth grade students (Fischer et al. 2013). Nevertheless, the CSCL scripts included in this meta-analysis did not have negative effects on motivation, but rather were comparable to unstructured collaboration. Hence, these CSCL scripts were not detrimental, as suggested by critics, but seemed not to exploit their full potential for increasing motivation.

The duration of the intervention could be an alternative factor explaining differences in the effects of CSCL scripts on motivation. It seems plausible that during an intervention with CSCL script collaboration skills develop. Hence, learners could perceive a CSCL script as more coercive after having received support for some time. Therefore, due to the small number of effects we qualitatively compared studies with respect to their intervention duration. The intervention of two studies lasted for several weeks (Peterson and Roseth 2016; Taylor and Baek 2019). Two other studies used interventions that lasted for 60 to 80 min (Rummel et al. 2009; Weinberger et al. 2005). One study did not provide any information on intervention duration (G.-Y. Lin 2020). Although one would expect learners to become less motivated the longer they learn with a CSCL script, the included studies do not support such pattern. Studies using CSCL scripts in long term interventions yielded significant positive effects (Taylor and Baek 2019) on motivation as well as null effects (Peterson and Roseth 2016). Studies that used CSCL scripts in short term interventions resulted in null effects (Rummel et al. 2009; Weinberger et al. 2005). Therefore, the existing data does not allow for conclusions on how motivation is affected by CSCL scripts over time. Other factors in the design of CSCL scripts might affect learners' motivation such as fading or adapting the CSCL scripts to individual needs. However, to our knowledge no study on CSCL scripts addresses these aspects together with the intervention duration. Thus, it is necessary to address systematically the question of how different CSCL scripts affect motivation, how this effect changes over time, and how technology can help to exploit the full potential of CSCL scripts.

Another explanation for the non-significant effect on motivation might be that learning with CSCL scripts has ambivalent effects on different factors influencing learners' motivation. CSCL scripts might help learners easily achieve strong feelings of competence and relatedness because they provide a structure for collaborative learning processes and learners' involvement in a social context. Such feelings of relatedness and competence are connected to higher levels of motivation (Rienties et al. 2012; Ryan and Deci 2000). On the other hand, the coercive nature of CSCL scripts, which strictly define the activities learners are expected to engage in, could lead to a lower degree of autonomy, which in turn reduces motivation (Wise and Schwarz 2017). This combination of positive and negative effects of learning with CSCL scripts on motivation might balance out, leading to a non-significant effect size close to zero. Unfortunately, the small number of studies empirically examining the effect of CSCL scripts on motivation precludes more nuanced quantitative analyses. Thus, more primary studies addressing hypotheses that take a more differentiated view on motivational factors, for example with respect to the basic psychological needs, are needed. Critique of CSCL scripts rarely address the different aspects of motivation. If we regard motivation as a holistic construct, there is no meta-analytical evidence for an overscripting effect. Thus, this criticism remains a postulate without corresponding empirical evidence.

Overall effects of learning with CSCL scripts on domain learning and learning collaboration skills (RQ 2)

Based on the theoretical assumptions of the script theory of guidance, we further hypothesized that learning with CSCL scripts should have a positive effect on domain learning and learning collaboration skills (Fischer et al. 2013). The results support this hypothesis. Thus, we can conclude that CSCL scripts do indeed support the learning of beneficial collaboration processes that eventually lead to better elaboration of the learning content and ultimately to better collaboration skills and domain learning outcomes (King 2007).

As already detected in the previous meta-analysis by Vogel et al. (2017), the effect of learning with CSCL scripts on collaboration skills was substantially higher than the effect on domain learning. In the script theory of guidance, it is assumed that the guidance provided by an external script helps learners to participate in a specific CSCL practice, building and reconfiguring internal scripts they can then recall in other situations (Fischer et al. 2013). In doing so, the CSCL script helps learners engage in collaborative activities that are beneficial for domain learning. Therefore, substantially higher effects on collaboration skills might be due to a more direct link between CSCL scripts and the development of collaboration skills, while domain learning is more indirectly supported by accomplishing what the script suggests learners to do. The wide range in magnitude of the effects of CSCL scripts on collaboration skills, however, raises questions about how effective such scaffolding can be. Although the average effect size of CSCL scripts on collaboration skills is comparable to the effect sizes other scaffolds such as example-based learning have on learning (J. Chen et al. 2018; Jeong et al. 2019; Wittwer and Renkl 2010), some studies report effect sizes far beyond the usual effects of scaffolding (Noroozi et al. 2013c). These studies reveal a need for closer examination and challenge future CSCL script designs to increase the effectiveness for domain learning to a similar size. One reason for the extraordinary effectiveness of these CSCL scripts might be that they do not only provide support during collaborative phases but also ask learners to individually prepare for the joint learning phases. Prior studies have found the combination of individual and collaborative phases during collaborative learning to be more beneficial than individual learning or collaborative learning alone (Olsen et al. 2017). Individual phases allow learners to prepare for collaboration and give them time to think about and prepare their contributions before being engaging in communication with the learning partner, when answers are expected to be formulated immediately and little time for individual thinking is available. Another reason might be that these studies measure the internal collaboration script that was addressed by the CSCL script particularly well, whereas other studies with lower effects use broader measures.

Explaining the effectiveness of CSCL scripts prompting different combinations of collaborative activities (RQ 3 and 4)

For domain learning, we assumed that combining prompts for different types of collaborative activities would increase domain learning through synergistic scaffolding. Prior research has shown that learners who engage in all types of collaborative activities have the highest domain learning outcomes (Andrews-Todd and Forsyth 2018). Hence, successful synergistic scaffolding should lead to effects on learning outcomes when the CSCL script combines prompts for different collaborative activities above and beyond the effects achieved through separate prompts for each collaborative activity (Tabak 2004). To investigate this issue, we compared CSCL scripts that prompted one, two, or three different collaborative activities. In contrast to our expectations, CSCL scripts were descriptively most effective when prompting only one collaborative activity and least effective when prompting a combination of two collaborative activities. Notably, on all levels CSCL scripts were more effective in fostering domain learning as measured by application tests compared to recall tests. This is particularly surprising given that lower overall effect sizes were found for application measures compared to recall measures. This may indicate that exclusively prompting one collaborative activity is most effective for enhancing application-oriented knowledge and skills. One reason for the larger effect might be that performing these collaborative activities evokes higher-order cognitive processes that allow learners to connect the new information with prior knowledge and apply new information to a problem (Chi 2009). Why this is only valid for CSCL scripts that prompted only one or three collaborative activities remains a subject for further research.

The finding that CSCL scripts prompting one collaborative activity outperform CSCL scripts prompting all three types of collaborative activities indicates that CSCL scripts did not successfully induce synergistic scaffolding. One plausible reason for this lack of synergistic scaffolding might be connected to the fact that CSCL scripts that prompt different types of collaborative activities are increasingly demanding. It is possible that these pose an additional load on the learner; in particular, scripts for several collaborative activities pose an even higher cognitive load on learners (e.g., F. Kirschner et al. 2009). Possible solutions to take some load off learners when working with highly complex CSCL scripts might be to offer the scaffolds for the different types of collaborative activities independently of each other or in a specific sequence (Schwaighofer et al. 2017). However, this explanation would be in conflict with the finding that scripting three collaborative activities is more effective than scripting two collaborative activities. Primary research on how to combine prompts for different types of collaborative activities in one CSCL script can lead to synergistic scaffolding (Tabak 2004) is still at a nascent stage, and more research is needed to find the most beneficial design for such scaffolding.

To explore the effectiveness of specific combinations of collaborative activities, we compared the different combinations of collaborative activities prompted by the CSCL script in more detail. CSCL scripts addressing solely coordination or a combination of negotiation and information sharing were most effective, followed by scripting information sharing only and the three-way combination. Notably, the effect of CSCL scripts that only prompted negotiation was very variable and non-significant. Upon closer examination, the variability of this effect is reflected in the variability of the CSCL scripts used in this sample. The CSCL scripts range from very elaborate, highly structured discussion scripts (Noroozi et al. 2013c) to argumentation scripts sequencing the order of arguments and counter-arguments (Stegmann et al. 2007) and CSCL scripts that solely prompt to discuss a specific topic (Rau et al. 2017). Comparing these studies, it seems that CSCL scripts offering a higher degree of structure have larger effects on domain learning. Comparing the different combinations of collaborative activities also provided more detailed insights into the question of how synergistic scaffolding might be achieved (Tabak 2004). Specifically, combining the negotiation and information sharing prompts yielded a higher effect size than offering scaffolding for one of the two types of collaborative activities alone. Analogously to the interpretation of the positive effect of combining of individual and collaborative activities (Olsen et al. 2017), this could be seen as a successful combination of two activities that can lead to synergistic effects. CSCL scripts that only prompted negotiation yielded a non-significant effect, which was dramatically improved by combining negotiation with information sharing. Conversely, scripting information sharing alone already had a significant positive effect, yet the effect was even higher when combined with negotiation. Examining the CSCL scripts used in these studies in detail, it stands out that scripting negotiation only means that students are specifically prompted to engage in discussion; however, an information exchange phase is missing (e.g., Puhl et al. 2015; Wu et al. 2019). This phase might help students better engage in beneficial negotiation activities by encouraging prior listening and thinking about the information their learning partners share with them. This effect seems to be in line with the importance of individual phases in which students can first think about and establish their viewpoint before engaging in collaboration (Olsen et al. 2017). Conversely, the effectiveness of information sharing prompts when learning with CSCL scripts does not seem to be comparably dependent on negotiation prompts, since CSCL scripts scaffolding information sharing only already achieved a substantial effect. It seems that the information exchange is one of the most beneficial activities for learning with CSCL scripts. However, its effectiveness can be diminished by additionally scaffolding coordination but further boosted by including negotiation prompts.

When combining all three types of collaborative activities, any benefits resulting from two-way combinations of scaffolding in CSCL scripts seem to vanish. This could be a consequence of over-loading students with too many different activities to focus on (F. Kirschner et al. 2009). Nevertheless, since each type of collaborative activity investigated seemed to work successfully at least when offered separately or in combination with one other activity, it remains an important avenue for further research to determine how the scaffolding of different collaborative activities should best be combined to increase domain learning. Particularly, it would be interesting to gain more insight into how scaffolding coordination can remain beneficial when combined with other activities. Here, the results of the meta-analysis showed that although CSCL scripts scaffolding coordination alone were quite effective, combining coordination activities with other prompts led to far smaller or even non-significant effects. Overall, the results indicate that it is more important to focus on which specific combination of collaborative activities is prompted by the CSCL script than the number of different collaborative activities prompted by the script. Unfortunately, the low number of studies preclude more detailed analyses of whether the three collaborative activities differ in their potential to advance application-oriented and recall-oriented domain learning. In particular, there is still a lack of primary research on the combination of two different collaborative activities.

For the learning of collaboration skills, the three-way combination also led to the smallest effect size. This means that the CSCL scripts included in this meta-analysis failed to induce differentiated scaffolding of collaboration skills in a beneficial way (Tabak 2004). Ideally, prompting all of the different types of collaborative activities would lead to strong effects of the specific prompts on the corresponding collaboration skills. Thus, the differentiated scaffolding of different collaborative activities within a single CSCL script should not reduce their effectiveness. Consequently, the results of this meta-analysis could be interpreted either as suggesting that differentiated scaffolding is not possible in the way proposed by Tabak (2004), or that it is necessary to further study how to induce differentiated scaffolding when developing CSCL scripts combining scaffolds for different types of collaborative activities in order to support collaboration skills. In addition, the intervention period of most primary studies included in this meta-analysis was rather short. Therefore, it is possible that scripting several different collaborative activities over a short period of time overwhelms the learners by inducing a high cognitive load (e.g., F. Kirschner et al. 2009). It seems plausible that repeated practice of a single collaborative activity falls short if combined with other activity prompts. This might also be a question of measurement. Although CSCL scripts often address more than one collaborative activity, only a few studies measured a mixture of collaboration skills. Therefore, it seems plausible that those scripting only one or two collaborative activities were better able to measure the learning of collaboration skills as their measures better aligned with the specific skill scaffolded by prompting the respective collaborative activity. Also, the prompting of specific types of collaborative activities might not have addressed the optimal scripting level (Fischer et al. 2013). Designing

CSCL scripts with a combination of prompts for different types of collaborative activities might often result in prompting only higher scripting levels to avoid scripting that is too extensive and overwhelming scripting. For example, learners might be prompted to discuss the most plausible solution (Rummel et al. 2009) or exchange information (Ertl et al. 2006). Here, particularly inexperienced learners could require more detailed scripting (i.e., on a lower scripting level) for these activities.

Overall, the results of the meta-analysis show that CSCL scripts are beneficial for domain learning and for enhancing collaboration skills. However, the proposed mechanisms, that generally prompting a combination of negotiation, information sharing, or coordination might explain the effectiveness of learning with CSCL scripts, were not supported by the results of this meta-analysis. In particular, combining prompts for two or more types of collaborative activities led to lower effect sizes than only prompting one of these activities. Thus, neither differentiated nor synergistic scaffolding could be successfully achieved by generally combining prompts for different types of collaborative activities in one CSCL script. Thus, for the effectiveness of CSCL scripts, the studies included in this meta-analysis indicate that in some cases less is more when it comes to scaffolding different collaborative activities. This might be due to the additional cognitive load posed by CSCL scripts incorporating scripting for different types of collaborative activities. Thus, how to achieve differentiated and synergistic scaffolding (Tabak 2004) when combining different scaffolds in one CSCL script remains an open question. Sequencing and fading scaffolds throughout the application of a CSCL script might be a promising approach, and thus should be examined in future experimental studies on learning with CSCL scripts.

Implications for the critique that CSCL scripts decrease motivation by "overscripting"

Although motivation is vividly discussed as a factor responsible for the small or negative effects of learning with CSCL scripts (Dillenbourg 2002; Wise and Schwarz 2017), the number of CSCL script studies measuring motivation is rather small. Nevertheless, the non-significant overall effect of CSCL scripts on motivation reported in this meta-analysis does not support the repeatedly asserted critique that the strict structuring of collaborative learning through CSCL scripts leads to a reduction of autonomy, which in turn negatively influences motivation and ultimately impedes learning (Dillenbourg 2002). If such an effect exists at all, it might be reduced by an opposing effect of CSCL scripts enhancing learners' feelings of competence and social relatedness, which should lead to higher motivation (Järvelä et al. 2010). The empirical studies included in this meta-analysis suggest that CSCL scripts are not detrimental for motivation, but rather can have positive effects. This strongly indicates that the overscripting effect, which is originally based on a conceptual article (Dillenbourg 2002), has been overblown by the research community without being based on empirical evidence. Nevertheless, theories about supporting collaborative learning scenarios could be further developed by reflecting in more detail on the effect of structuring CSCL on motivation and integrating different factors that might have positive or negative effects on motivation, such as those proposed by self-determination theory (Deci and Ryan 1985). Future research on learning with CSCL scripts should measure aspects of motivation by default in order to achieve a more robust sample of effect sizes for motivation. Moreover, given the hypothesized detrimental effect of learning with CSCL scripts on motivation, future CSCL script designs should try to increase the positive effect of CSCL scripts on feelings of competence or social relatedness. Increasing the freedom afforded by CSCL scripts, for example by fading them out (e.g., Wecker and Fischer 2011) or adapting them to learners' needs (Rau et al. 2017), could presumably decrease a possible negative effect of CSCL scripts on learners' autonomy.

Implications for the script theory of guidance

The positive effects of CSCL scripts on both domain learning and learning collaboration skills are in line with the theoretical assumptions that learning with CSCL scripts induces beneficial collaborative processes, ultimately leading to better learning compared to unstructured collaborative learning (King 2007). These findings justify the principles formulated in the script theory of guidance that providing external scripts enables learners to engage in collaborative practice in a way that leads to learning of knowledge and skills (Fischer et al. 2013). Although CSCL scripts are mainly considered useful for learning to collaborate (Wise and Schwarz 2017), there are far more studies investigating their effect on domain learning. The small positive effect on domain learning is clearly stable, whereas the large positive effect on collaboration skills is deeply heterogeneous. Additionally, the relatively small number of studies analyzing effects on collaboration skills precludes more comprehensive moderator analyses that could help identify factors explaining the heterogeneity in effect sizes. Thus, in future research, fewer studies on the general effects of CSCL scripts on domain learning are needed, but more studies explicitly analyzing differential effects of CSCL scripts using different designs, CSCL scripts inducing different types of activities, or CSCL scripts implemented in various contexts. In contrast, more studies examining the general effect of CSCL scripts on collaboration skills are still needed. There is a particular lack of studies concerning the effect of CSCL scripts on coordination, despite the fact that studies facilitating coordination were most effective for domain learning. Moreover, given the various measures used to assess collaboration skills, the field would be strengthened by developing instruments for assessing particular collaboration skills such as information sharing or negotiating and widely applying them in studies on computer-supported collaborative learning.

Implications for the design of CSCL scripts aiming at differentiated and synergistic scaffolding

This meta-analysis has shown that combining prompts for different types of collaborative activities led to a reduction in the effectiveness of learning with CSCL scripts, particularly for learning collaboration skills. Consequently, the general functioning of differentiated scaffolding and synergistic scaffolding is doubtful (Kollar et al. 2014; Noroozi et al. 2013c; Tabak 2004). However, more detailed analyses of specific combinations of the three collaborative activities showed that combined scaffolding does not always lead to smaller effects. Thus, some types of combinations might be needed to take full advantage of learning with collaboration scripts. At the very least, a major future goal of scaffolding and scripting research should be to examine how and when CSCL scripts lead to successful differentiated and synergistic scaffolding. Two possible ways to increase the effectiveness of CSCL scripts with combined scaffolds would be to introduce sequencing and adapt the different scaffolds. In an experimental study concerning CSCL scripts for mathematical argumentation, Schwaighofer et al. (2017) showed that presenting different scaffolds in a specific sequence and fading out the scaffold presented first can increase the effectiveness of CSCL scripts, particularly for learners with weaker cognitive prerequisites. In another study, Wecker and Fischer (2011) showed that peer-monitored fading increases the effectiveness of CSCL scripts. It seems plausible to assume that these results are transferrable to scaffolding different collaborative activities within one CSCL script.

Limitations

Of course, this meta-analysis is not without limitations that must be considered when interpreting the results and drawing conclusions and implications. Although there is a large body of research and strong theoretical foundation concerning the effectiveness of CSCL scripts on learning, the design and context of CSCL scripts in empirical research varies greatly. The targeted collaboration skills range from the construction of individual arguments in short one-hour trainings to engaging in argumentative discourse, exchanging peer feedback and preparing mutual individual introductions, and even adhering to assigned roles for several weeks of collaborative learning activities. This leads to difficulties in finding comparable CSCL script studies and summarizing them in a reasonable way. In addition, there is wide variety in what constitutes a CSCL script. However, by applying a consistent definition of CSCL scripts and searching for studies on CSCL scripts without solely using the term script, but rather focusing on the mechanisms of CSCL scripts, we tried to identify an appropriate sample of primary studies representing what has been discovered empirically about learning with CSCL scripts. It was, however, necessary to constrain the literature search to specific search terms. Although we also used other terms than the term “script” itself, it is possible that we systematically missed studies from research areas using a different nomenclature. One such area might be the research on dynamic support such as conversational agents. We included two studies using conversational agents for the present meta-analysis (Adamson et al. 2014; Ulicsak 2004). Our decision to not use the rather broad search term “support” has led to missing out studies from this field of research (e.g., Wang et al. 2011). However, the respective analyses did not indicate any substantial publication bias. This leads us to conclude that we found a comprehensive sample of primary studies for the phenomenon under investigation. Nevertheless, it is possible that specific research areas are underrepresented.

Regarding the coding of the primary studies, a major problem is that most studies do not provide direct indicators for the factors that are theoretically and empirically assumed to positively affect the effectiveness of learning with CSCL scripts. The assumption is that the more learners engage in specific activities in the learning process (e.g., negotiation), the more they should benefit from collaborative learning (Chi and Wylie 2014; Liu et al. 2015). To analyze the effectiveness of learning with CSCL scripts based on this assumption, the most direct indicator primary studies could provide would be measures of the activities learners actually engage in throughout the collaborative learning process. However, most studies do not report such measures; thus, the most proximal information provided by the primary studies is the description of activities prompted by the CSCL script. What actually happened during the collaboration process in the different studies can only be estimated using this information, resulting in several sources of uncertainty. First, the accuracy of this information varies across studies. While some studies report the activities required in the CSCL script in detail (e.g., Rummel et al. 2009; Ulicsak 2004), others are less detailed, which might have led to inaccurate estimations of the learning activities that were actually used in the studies' learning processes (e.g., Hsu et al. 2015; Tsovaltzi et al. 2015). A second source of uncertainty is that, even when the activities required by the CSCL scripts are described in detail, we do not know to what extent the learners adhered to what they were asked to do during the learning process. Sometimes learners only complete the required activities on a very superficial level. Nevertheless, the description of activities required by the CSCL scripts is the most proximal estimation for the collaboration process available. Moreover, we assume that the presumed inaccuracy of this measure is well distributed across studies and thus should not substantially bias the results of the comparisons between different types of CSCL scripts.

Another limitation of this meta-analysis is that, although motivation features prominently in criticisms of learning with CSCL scripts, there were only a few studies measuring motivation (e.g., Peterson and Roseth 2016). Moreover, motivation is conceptualized in CSCL research in highly diverse ways, such as approaches building on expectancy value conceptualizations, distinguishing between extrinsic and intrinsic motivation or conceptualizations such as self-determination theory integrating the needs for competence, autonomy, and social relatedness (Deci and Ryan 1985). The broad variety of conceptualizations and the rather small number of independent effect sizes weaken our ability to interpret them coherently and do not allow for examinations of further moderating effects.

More generally, the number of CSCL script studies included in this meta-analysis is relatively low, which reduces the possibility of comparing between different studies on a more fine-grained level. The number of studies at each level of the examined factors was sometimes too low or too unevenly distributed to conduct a comparison and interpret the results in a reasonable manner. This was particularly true for studies analyzing the effects of CSCL scripts on collaboration skills and on motivation. Additionally, even within a moderator level the reported effect sizes are often very heterogeneous. This indicates that studies within moderator levels vary systematically due to unknown covariates. Therefore, such results should be interpreted carefully. Unfortunately, we cannot solve this issue until more primary studies are conducted that would allow more fine-grained analyses.

The final limitation that must be considered when interpreting the results of the meta-analysis at hand is that comparisons of different levels of the moderators are only comparisons between studies. Almost none of the studies compared the moderator levels as a within-study effect (e.g., Peterson and Roseth 2016). Thus, the differences between different moderator levels cannot be interpreted causally. They only suggest a direction for the empirical relationship, which might have been confounded by the specific study designs.

Conclusion

This meta-analysis investigated the effect of CSCL scripts on motivation, domain learning, and collaboration skills and proposed moderators to explain their effectiveness. Most importantly, this meta-analysis shows that the widely proposed hypothesis (Dillenbourg 2002; Wise and Schwarz 2017) that CSCL scripts reduce learners' motivation is not supported by the available data. Although a negative effect of CSCL scripts on motivation seems very unlikely given the present data, the existing empirical evidence might still not be enough to refute this criticism completely. Only a few studies empirically investigated the effect on motivation at all, and more differentiated analyses comparing different CSCL scripts' effects on motivation are lacking. It seems plausible that CSCL scripts foster some aspects of motivation, such as a feeling of competence, and at the same time diminish other aspects of motivation, such as fulfillment of the need for autonomy. Therefore, we call for a systematic empirical investigation of how different CSCL scripts affect the various aspects of motivation and how a possible influence interacts with the learning setting (e.g., duration of the intervention, synchronous versus asynchronous learning, learners' age, learners' prior knowledge). We suggest self-determination theory (Deci and Ryan 1985) as a potential theoretical foundation for obtaining a differentiated picture on motivational aspects that are important for students' learning.

Furthermore, the results are in line with prior studies' findings showing that CSCL scripts indeed have positive effects on domain learning and learning of collaboration skills. Hence, CSCL scripts engage learners in beneficial activities that enhance learning (Fischer et al. 2013). However, the results of this meta-analysis suggest that future research should focus on developing detailed measures of collaboration skills instead of simply replicating the well-established effect of CSCL scripts on domain learning. Although the effect of CSCL scripts on collaboration skills is larger compared to their effect on domain learning, it is also particularly heterogeneous, indicating that the effect differs widely between different types of CSCL scripts. Hence, future research may focus on identifying moderators that explain this heterogeneity. A further promising area of research would be to investigate how effects of CSCL scripts can be increased by designing dynamic CSCL scripts.

One aspect that seems to influence the effectiveness of CSCL scripts at least descriptively is the number of different collaborative activities prompted. For collaboration skills, this meta-analysis revealed that when it comes to designing CSCL scripts, less is more. Scaffolding fewer different types of collaborative activities generally led to higher effects of CSCL scripts on collaboration skills. For domain learning, these effects were less clear. The hypothesized effectiveness of combining different scaffolds and thus leading to differentiated and synergistic scaffolding (Tabak 2004) did not pan out for some combinations of collaborative activities. Hence, future research is needed that investigates how and which collaborative activities could be combined to enhance the effectiveness of CSCL scripts. Additionally, sequencing and fading different prompts within a single CSCL script might better exploit the potential of combining scaffolds that have been shown to be beneficial when being applied alone. However, the present meta-analyses failed to identify clear moderators explaining this variance.

Overall, the positive effect of CSCL scripts on domain learning and collaboration skills highlight the potential of CSCL scripts to facilitate learning. The finding that motivation is not significantly affected by learning with CSCL scripts stands in contrast to the wide-spread belief that CSCL scripts undermine learners' agency. Altogether, CSCL scripts seem to enable learners to make use of the different benefits of collaborative learning rather than hinder their learning processes.

Acknowledgements The research presented in this contribution was funded by a grant of the Deutsche Forschungsgemeinschaft (DFG; FI 792/11-1).

Funding Information Open Access funding provided by Projekt DEAL.

Appendix

Table 4 *Moderator Levels, Dependent Variables, and Individual Effect Sizes per Study*

Study	Type of Collaborative Activity			DV	Subjects			ES	
	Negotiation	Info sharing	Coordination		Unit	n exp.	n con.	g	V_g
Adamson et al. (2014), Study 4	Scripted	Scripted	Not scripted	DL: Mixed	Ind.	9	9	0.37	0.21
Belland et al. (2011)	Scripted	Scripted	Scripted	CS: Info. sharing	Ind.	41	45	0.33	0.05
Bouyias and Demetriadis (2012)	Not scripted	Scripted	Scripted	DL: Recall	Ind.	14	10	1.04	0.18
				CS: Info. sharing	Ind.	14	10	−0.36	0.16
				CS: Info. sharing	Ind.	14	10	−0.52	0.17
				CS: Negotiation	Ind.	14	10	1.15	0.19
Cáceres et al. (2018)	Scripted	Scripted	Scripted	DL: Appl.	Ind.	23	29	0.89	0.08
Chen and Chiu (2016)	Scripted	Scripted	Scripted	DL: Appl.	Ind.	35	37	0.31	0.06
High prior knowledge									
Low prior knowledge						35	37	−0.28	0.05
Chiu et al. (2013)	Not scripted	Scripted	Scripted	DL: Appl.	Group	11	11	0.81	0.18
				CS: Mixed	Group	33	33	0.54	0.06
				CS: Mixed	Group	33	33	1.11	0.07
Cho and Jonassen (2002)	Scripted	Scripted	Not scripted	CS: Info. sharing	Ind.	15	15	0.97	0.14
Illstructured problem									
Well-structured problem						15	15	0.13	0.13
Choi et al. (2005)	Scripted	Not scripted	Not scripted	CS: Info. sharing	Ind.	20	19	0.58	0.1
						20	19	−0.16	0.1
				CS: Negotiation	Ind.	20	19	0.17	0.1
				CS: Mixed	Ind.	20	19	0.52	0.1
Demetriadis et al. (2011)	Not scripted	Scripted	Scripted	CS: Mixed	Ind.	34	29	0.81	0.07
				DL: Recall	Ind.	34	29	0.01	0.06
				Motivation	Ind.	34	29	0.37	0.06
						34	29	0.51	0.06
						34	29	−0.63	0.07
Ertl et al. (2006), Study 2	Scripted	Scripted	Scripted	DL: Recall	Ind.	22	24	0.62	0.09
Without conceptual support									
						22	24	0.22	0.08

Table 4 (continued)

Study	Type of Collaborative Activity			DV	Subjects			ES	
	Negotiation	Info sharing	Coordination		Unit	n exp.	n con.	g	V_g
With conceptual support	Scripted	Scripted	Scripted	DL: Recall	Ind.	20	20	0.56	0.1
						20	20	−0.14	0.1
Combined support	Scripted	Scripted	Scripted	DL: Recall	Ind.	20	24	0.39	0.09
						20	24	0.09	0.09
Ertl et al. (2006), Study 3	Not scripted	Scripted	Scripted	DL: Appl.	Ind.	78	81	0.35	0.03
						78	81	0.3	0.03
Ertl et al. (2008)	Scripted	Scripted	Scripted	DL: Appl.	Ind.	39	39	0.03	0.05
Without content scheme									
						39	39	0.3	0.05
						39	39	−0.32	0.05
						39	39	−0.29	0.05
With content scheme	Scripted	Scripted	Scripted	DL: Appl.	Ind.	39	42	0.08	0.05
						39	42	0.93	0.05
						39	42	0.68	0.05
						39	42	0.77	0.05
Gelmini-Hornsby et al. (2011)	Not scripted	Scripted	Not scripted	CS: Info. sharing	Group	8	10	0.81	0.22
Gijlers and de Jong (2009)	Scripted	Scripted	Scripted	DL: Appl.	Ind.	22	22	0.08	0.09
Shared Scratchpad									
				DL: Recall	Ind.	22	22	−0.66	0.09
Shared Proposition Table	Scripted	Scripted	Scripted	DL: Appl.	Ind.	22	22	0.56	0.09
				DL: Recall	Ind.	22	22	0.33	0.09
Gijlers et al. (2013)	Scripted	Scripted	Not scripted	DL: Recall	Ind.	32	24	0.89	0.08
						32	24	0.48	0.07
Haake and Pfister (2010)	Scripted	Not scripted	Scripted	DL: Recall	Group	21	21	−0.52	0.09
Hron et al. 2000	Scripted	Scripted	Scripted	DL: Recall	Ind.	15	15	−0.47	0.13
Explicit structuring									
Implicit structuring	Scripted	Scripted	Scripted	DL: Recall	Ind.	15	15	0.15	0.13
Hsu et al. (2015)	Scripted	Scripted	Scripted	CS: Info. sharing	Ind.	26	24	0.7	0.08
Huang et al. (2012)	Not scripted	Scripted	Not scripted	DL: Recall	Ind.	30	30	0.79	0.07

Table 4 (continued)

Study	Type of Collaborative Activity			DV	Subjects			ES	
	Negotiation	Info sharing	Coordination		Unit	n exp.	n con.	g	V_g
Kollar et al. (2014) *With problem solving*	Scripted	Scripted	Scripted	CS: Negotiation	Ind.	22	26	0.21	0.08
				DL: Appl.	Ind.	22	26	−0.3	0.08
With heuristic worked examples	Scripted	Scripted	Scripted	CS: Negotiation	Ind.	26	27	0.19	0.07
				DL: Appl.	Ind.	26	27	0.24	0.07
Kollar et al. (2007) *Low internal collaboration scripts*	Scripted	Scripted	Scripted	CS: Negotiation	Ind.	22	20	0.41	0.09
				DL: Recall	Ind.	22	20	0.29	0.09
				DL: Appl.	Ind.	22	20	−0.51	0.09
High internal collaboration scripts	Scripted	Scripted	Scripted	CS: Negotiation	Ind.	26	22	0.86	0.09
				DL: Recall	Ind.	26	22	0.33	0.08
				DL: Appl.	Ind.	26	22	−0.55	0.08
Lee (2015) *Males*	Scripted	Scripted	Scripted	DL: Recall	Ind.	10	12	−0.26	0.17
				DL: Appl.	Ind.	10	12	0	0.17
						10	12	−0.44	0.17
Females	Scripted	Scripted	Scripted	DL: Recall	Ind.	29	25	0.58	0.08
				DL: Appl.	Ind.	29	25	0.19	0.07
						29	25	0.63	0.08
Lee (2018) *Immediate posttest*	Scripted	Scripted	Scripted	DL: Recall	Ind.	50	35	−0.08	0.05
				DL: Appl.	Ind.	50	35	0.14	0.05
						50	35	0.17	0.05
Delayed posttest	Scripted	Scripted	Scripted	DL: Recall	Ind.	50	35	0.18	0.05
				DL: Appl.	Ind.	50	35	0.58	0.05
						50	35	0.56	0.05
Lin (2020) *Low mastery goal orientation*	Scripted	Not scripted	Scripted	Motivation	Ind.	33	33	0.04	0.01
						33	33	0.02	0.01
						33	33	0.32	0.01

Table 4 (continued)

Study	Type of Collaborative Activity			DV	Subjects			ES	
	Negotiation	Info sharing	Coordination		Unit	n exp.	n con.	g	V_g
High mastery goal orientation	Scripted	Not scripted	Scripted	Motivation	Ind.	28	28	0.08	0.01
						28	28	0	0.01
						28	28	0.01	0.01
Lin et al. 2018 & Chen et al. (2018)	Scripted	Scripted	Scripted	DL: Appl.	Ind.	26	24	0.52	0.08
				CS: Info. sharing	Ind.	26	24	1.2	0.09
						26	24	1.24	0.09
						26	24	2.42	0.14
						26	24	2.48	0.14
Mende et al. (2017)	Not scripted	Scripted	Scripted	DL: Appl.	Ind.	43	45	−0.18	0.04
Noroozi et al. 2013b	Scripted	Scripted	Scripted	DL: Appl.	Ind.	30	30	0.8	0.07
Transactive Memory Script									
Transactive Discussion Script	Scripted	Scripted	Scripted	DL: Appl.	Ind.	30	30	0.92	0.07
Combined Scripts	Scripted	Scripted	Scripted	DL: Appl.	Ind.	30	30	−0.8	0.07
Noroozi et al. 2013c)	Scripted	Not scripted	Not scripted	DL: Appl.	Ind.	30	30	0.79	0.07
				CS: Negotiation	Ind.	30	30	1.89	0.09
						30	30	1.42	0.08
Peterson and Roseth (2016)	Not scripted	Not scripted	Scripted	DL: Recall	Ind.	82	70	0.58	0.03
Synchronous collaboration									
						82	61	0.69	0.03
				Motivation	Ind.	72	54	−0.12	0.03
						72	54	−0.08	0.03
						72	49	−0.06	0.03
						72	49	0	0.03
Asynchronous collaboration	Not scripted	Not scripted	Scripted	DL: Recall	Ind.	70	57	0.45	0.03
						70	55	0.47	0.03
				Motivation	Ind.	62	48	0.32	0.04
						62	48	0.31	0.04
						62	42	0.31	0.04
						62	42	0.71	0.04

Table 4 (continued)

Study	Type of Collaborative Activity			DV	Subjects			ES	
	Negotiation	Info sharing	Coordination		Unit	n exp.	n con.	g	V_g
Puhl et al. (2015)	Scripted	Not scripted	Not scripted	DL: Recall	Ind.	20	23	0.25	0.09
						20	19	1.09	0.11
				CS: Negotiation	Ind.	20	23	−0.06	0.09
						20	19	0.33	0.1
Rau et al. (2017)	Scripted	Not scripted	Not scripted	DL: Recall	Ind.	31	30	−0.05	0.06
				DL: Appl.	Ind.	31	30	0.46	0.07
						31	30	0.14	0.06
Rienties et al. (2012)	Not scripted	Scripted	Scripted	DL: Mixed	Ind.	61	82	0.67	0.03
						61	82	−0.59	0.03
Rummel and Spada (2005)	Scripted	Scripted	Scripted	DL: Appl.	Group	9	9	−0.68	0.21
						9	9	0.98	0.23
				CS: Mixed	Ind.	18	18	0.63	0.11
				DL: Recall	Ind.	18	18	1.42	0.13
Rummel et al. (2012)	Not scripted	Scripted	Scripted	DL: Appl.	Group	18	19	−0.23	0.1
						18	19	−0.46	0.11
						23	19	0.58	0.1
						23	19	0.44	0.09
					Ind.	38	39	0.08	0.05
						38	39	−0.04	0.05
Rummel et al. (2009) *Script plus*	Scripted	Scripted	Scripted	DL: Appl.	Group	8	8	−0.16	0.22
				CS: Mixed	Ind.	16	16	−0.36	0.12
				Motivation	Ind.	16	16	0.45	0.12
						16	16	0.36	0.12
						16	16	−0.21	0.12
						16	16	0.22	0.12
						16	16	0.11	0.12
						16	16	0.5	0.12
Script	Scripted	Scripted	Scripted	DL: Appl.	Group	8	8	−0.48	0.23
				CS: Mixed	Ind.	16	16	−0.25	0.12

Table 4 (continued)

Study	Type of Collaborative Activity			DV	Subjects			ES	
	Negotiation	Info sharing	Coordination		Unit	n exp.	n con.	g	V_g
				Motivation	Ind.	16	16	0.25	0.12
						16	16	0	0.12
						16	16	−0.39	0.12
						16	16	0.62	0.12
						16	16	0.01	0.12
						16	16	0.02	0.12
Schellens et al. (2007)	Scripted	Scripted	Scripted	DL: Recall	Ind.	286	223	0.38	0.01
Scheuer et al. (2014)	Scripted	Not scripted	Not scripted	DL: Recall	Ind.	24	20	−0.09	0.09
Schmitt and Weinberger (2019)	Not scripted	Scripted	Not scripted	DL: Recall	Group	20	20	−0.12	0.1
Verbalization script									
Strategy script	Scripted	Scripted	Not scripted	DL: Recall	Group	21	20	0.28	0.09
Combined scripts	Scripted	Scripted	Not scripted	DL: Recall	Group	20	20	0.09	0.1
Stegmann et al. (2012)	Not scripted	Scripted	Not scripted	DL: Appl.	Ind.	8	8	0.64	0.24
					Group	24	24	0.6	0.08
				CS: Info. sharing	Ind.	8	8	2.49	0.42
					Group	24	24	2.89	0.17
Stegmann et al. (2007)	Not scripted	Scripted	Not scripted	DL: Appl.	Ind.	10	10	0.66	0.19
Argument construction script									
				CS: Negotiation	Ind.	10	10	0.11	0.18
				CS: Info. sharing	Ind.	10	10	1.19	0.22
Argument sequence script	Scripted	Not scripted	Not scripted	DL: Appl.	Ind.	10	10	0.71	0.2
				CS: Negotiation	Ind.	10	10	1.25	0.22
				CS: Info. sharing	Ind.	10	10	0.66	0.19
Combined scripts	Scripted	Scripted	Not scripted	DL: Appl.	Ind.	10	10	0.3	0.19
				CS: Negotiation	Ind.	10	10	1.33	0.23
				CS: Info. sharing	Ind.	10	10	1.36	0.23
Taylor and Baek (2019)	Not scripted	Scripted	Scripted	DL: Appl.	Ind.	63	64	0.54	0.03
Fixed roles									
				Motivation	Ind.	63	60	0.33	0.03
Rotating roles	Not scripted	Scripted	Scripted	DL: Appl.	Ind.	60	64	−0.37	0.03
				Motivation	Ind.	60	60	0.57	0.03

Table 4 (continued)

Study	Type of Collaborative Activity			DV	Subjects			ES	
	Negotiation	Info sharing	Coordination		Unit	n exp.	n con.	g	V_g
Tsovaltzi et al. (2015), Study 1	Not scripted	Scripted	Not scripted	DL: Recall	Ind.	20	20	0.27	0.1
Tsovaltzi et al. (2015), Study 3	Not scripted	Scripted	Not scripted	DL: Recall	Ind.	32	32	0.07	0.06
Tsovaltzi et al. (2014)	Not scripted	Scripted	Not scripted	DL: Recall	Ind.	40	40	0.14	0.05
				CS: Info. sharing	Ind.	40	40	0.45	0.05
Ulicsak (2004)	Not scripted	Not scripted	Scripted	DL: Mixed	Ind.	26	25	0.21	0.08
						26	25	0.12	0.08
						26	25	0.13	0.08
van Aalst and Chan (2007), Study 2	Not scripted	Scripted	Not scripted	DL: Appl.	Ind.	14	9	1.31	0.21
				DL: Recall	Ind.	14	9	1.22	0.2
van Aalst and Chan (2007), Study 3	Not scripted	Scripted	Not scripted	DL: Appl.	Ind.	24	23	−0.14	0.08
				DL: Recall	Ind.	24	23	0.68	0.09
van Dijk et al. (2014)	Scripted	Scripted	Scripted	DL: Recall	Ind.	20	20	0.05	0.1
Without grouping preference									
						20	20	−0.37	0.1
Without grouping preference	Scripted	Scripted	Scripted	DL: Recall	Ind.	18	24	0.2	0.09
						18	24	0.14	0.09
Vogel et al. (2016)	Scripted	Scripted	Scripted	CS: Negotiation	Ind.	48	53	0.39	0.04
Weinberger et al. (2005), Study 1	Not scripted	Scripted	Not scripted	DL: Appl.	Ind.	48	48	−0.29	0.04
Epistemic script									
Social script	Scripted	Not scripted	Scripted	DL: Appl.	Ind.	48	48	0.21	0.04
Weinberger et al. (2005), Study 2	Scripted	Scripted	Scripted	DL: Recall	Ind.	11	12	0.04	0.16
Social script									
						11	12	0.61	0.17
						11	12	0.22	0.16
						11	12	−0.09	0.16
						11	12	0.35	0.16
						11	12	0	0.16
				Motivation	Ind.	11	12	−0.65	0.17
Social and epistemic script	Scripted	Scripted	Scripted	DL: Recall	Ind.	10	10	0.43	0.19
						10	10	−0.21	0.18

Table 4 (continued)

Study	Type of Collaborative Activity			DV	Subjects			ES	
	Negotiation	Info sharing	Coordination		Unit	n exp.	n con.	g	V_g
						10	10	0.54	0.19
						10	10	−0.13	0.18
						10	10	−0.13	0.18
						10	10	0.33	0.19
				Motivation	Ind.	10	10	−0.12	0.18
Weinberger et al. (2010)	Not scripted	Scripted	Not scripted	DL: Appl.	Ind.	9	9	0.91	0.22
				CS: Info. sharing	Ind.	9	9	3.17	0.48
Wu et al. (2019) *Model condition*	Scripted	Not scripted	Not scripted	DL: Recall	Ind.	75	383	−0.11	0.02
				DL: Appl.	Ind.	75	383	0.16	0.02
Draw condition	Scripted	Not scripted	Not scripted	DL: Recall	Ind.	107	383	−0.12	0.01
				DL: Appl.	Ind.	107	383	0.05	0.01

Note. The reported effect sizes are unaggregated effect sizes per study. DL = domain learning; CS = collaboration skills; Appl. = application oriented; Info. Sharing = information sharing; Ind. = Individual

Table 5 *Absolut Numbers of Independent Effect Sizes for the Combination of Dependent Variables and Moderator Levels Within One Statistical Comparison*

	k	Motivation	DL: Recall tests	DL: Application tests	DL: Mixed tests	CS: Negotiation skills	CS: Information sharing skills	CS: Coordination skills	CS: Mixed skills
k		9	32	30	3	10	12	0	5
Scripting Negotiation	12	0	4	3	0	3	1	0	1
Scripting Information sharing	19	0	7	6	0	1	5	0	0
Scripting Coordination	5	2	2	0	1	0	0	0	0
Scripting Negotiation and Information sharing	4	0	1	0	1	0	2	0	0
Scripting Negotiation and Coordination	3	2	1	0	0	0	0	0	0
Scripting Information sharing and coordination	14	2	2	5	1	1	1	0	2
Scripting three-way combination	44	3	15	16	0	5	3	0	2

DL = Domain learning, CS = Collaboration skills

References

References marked with an asterisk indicate studies included in the meta-analysis.

*Adamson, D., Dyke, G., Jang, H., & Rosé, C. P. (2014). Towards an agile approach to adapting dynamic collaboration support to student needs. *International Journal of Artificial Intelligence in Education, 24*(1), 92–124. https://doi.org/10.1007/s40593-013-0012-6.

Andrews-Todd, J., & Forsyth, C.-M. (2018). Exploring social and cognitive dimensions of collaborative problem solving in an open online simulation-based task. *Computers in Human Behavior.* https://doi.org/10.1016/j.chb.2018.10.025.

Bell, P. (2004). Promoting students' argument construction and collaborative debate in the science classroom. In M. C. Linn, E. A. Davis, & P. Bell (Eds.), *Internet environments for science education* (pp. 115–143). Mahwah, NJ: Erlbaum.

*Belland, B. R., Glazewski, K. D., & Richardson, J. C. (2011). Problem-based learning and argumentation: Testing a scaffolding framework to support middle school students' creation of evidence-based arguments. *Instructional Science, 39*(5), 667–694. https://doi.org/10.1007/s11251-010-9148-z.

Belland, B. R., Walker, A. E., Kim, N. J., & Lefler, M. (2017). Synthesizing results from empirical research on computer-based scaffolding in STEM education: A meta-analysis. *Review of Educational Research, 87*(2), 309–344. https://doi.org/10.3102/0034654316670999.

Bollen, L., Gijlers, H., & van Joolingen, W. (2015). Drawings in computer-supported collaborative learning - empirical and technical results. *Computing and Informatics, 34*, 559–587.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. West Sussex UK: John Wiley & Sons.

*Bouyias, Y., & Demetriadis, S. (2012). Peer-monitoring vs. micro-script fading for enhancing knowledge acquisition when learning in computer-supported argumentation environments. *Computers & Education, 59*(2), 236–249. https://doi.org/10.1016/j.compedu.2012.01.001.

*Cáceres, M., Nussbaum, M., Marroquín, M., Gleisner, S., & Marquínez, J. (2018). Building arguments: Key to collaborative scaffolding. *Interactive Learning Environments, 26*(3), 355–371. https://doi.org/10.1080/10494820.2017.1333010.

*Chen, C.-H., & Chiu, C.-H. (2016). Collaboration scripts for enhancing metacognitive self-regulation and mathematics literacy. *International Journal of Science and Mathematics Education, 14*, 263–280. https://doi.org/10.1007/s10763-015-9681-y.

Chen, J., Wang, M., Kirschner, P. A., & Tsai, C.-C. (2018). The role of collaboration, computer use, learning environments, and supporting strategies in CSCL: A meta-analysis. *Review of Educational Research, 88*, 799–843. https://doi.org/10.3102/0034654318791584.

Chi, M. T. H. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science, 1*, 73–105. https://doi.org/10.1111/j.1756-8765.2008.01005.x.

Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist, 49*, 219–243. https://doi.org/10.1080/00461520.2014.965823.

*Chiu, C.-H., Wu, C.-Y., Hsieh, S.-J., Cheng, H.-W., & Huang, C.-K. (2013). Employing a structured interface to advance primary students' communicative competence in a text-based computer mediated environment. *Computers & Education, 60*(1), 347–356. https://doi.org/10.1016/j.compedu.2012.09.002.

*Cho, K.-L., & Jonassen, D. H. (2002). The effects of argumentation scaffolds on argumentation and problem solving. *Educational Technology Research and Development, 50*(3), 5–22. https://doi.org/10.1007/BF02505022.

*Choi, I., Land, S. M., & Turgeon, A. J. (2005). Scaffolding peer-questioning strategies to facilitate metacognition during online small group discussion. *Instructional Science, 33*, 483–511. https://doi.org/10.1007/s11251-005-1277-4.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale NJ: Lawrence Erlbaum Associates.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York NY: Plenum.

*Demetriadis, S., Egerter, T., Hanisch, F., & Fischer, F. (2011). Peer review-based scripted collaboration to support domain-specific and domain-general knowledge acquisition in computer science. *Computer Science Education, 21*, 29–56. https://doi.org/10.1080/08993408.2010.539069.

Dillenbourg, P. (2002). Over-scripting CSCL: The risks of blending collaborative learning with instructional design. In P. A. Kirschner (Ed.), *Three worlds of CSCL. Can we support CSCL?* (pp. 61–91). Heerlen, Netherlands: Open Universiteit Nederland.

Dillenbourg, P., Järvelä, S., & Fischer, F. (2009). The evolution of research on computer-supported collaborative learning. In N. Balacheff, S. Ludvigsen, T. de Jong, A. Lazonder, & S. Barnes (Eds.), *Technology-Enhanced Learning: Principles and Products* (pp. 3–19). Berlin, Germany: Springer.

Duval, S., & Tweedie, R. (2000). A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association, 95*(449), 89–98. https://doi.org/10.1080/01621459.2000.10473905.

*Ertl, B., Fischer, F., & Mandl, H. (2006). Conceptual and socio-cognitive support for collaborative learning in videoconferencing environments. *Computers & Education, 47*, 298–315. https://doi.org/10.1016/j.compedu.2004.11.001.

*Ertl, B., Kopp, B., & Mandl, H. (2008). Supporting learning using external representations. *Computers & Education, 51*, 1599–1608. https://doi.org/10.1016/j.compedu.2008.03.001.

Fischer, F., Kollar, I., Stegmann, K., & Wecker, C. (2013). Toward a script theory of guidance in computer-supported collaborative learning. *Educational Psychologist, 48*, 56–66. https://doi.org/10.1080/00461520.2012.748005.

*Gelmini-Hornsby, G., Ainsworth, S., & O'Malley, C. (2011). Guided reciprocal questioning to support children's collaborative storytelling. *International Journal of Computer-Supported Collaborative Learning, 6*, 577–600. https://doi.org/10.1007/s11412-011-9129-5.

*Gijlers, H., & de Jong, T. (2009). Sharing and confronting propositions in collaborative inquiry learning. *Cognition and Instruction, 27*, 239–268. https://doi.org/10.1080/07370000903014352.

*Gijlers, H., Weinberger, A., van Dijk, A. M., Bollen, L., & van Joolingen, W. (2013). Collaborative drawing on a shared digital canvas in elementary science education: The effects of script and task awareness support. *International Journal of Computer-Supported Collaborative Learning, 8*, 427–453. https://doi.org/10.1007/s11412-013-9180-5.

*Haake, J. M., & Pfister, H.-R. (2010). Scripting a distance-learning university course: Do students benefit from net-based scripted collaboration? *International Journal of Computer-Supported Collaborative Learning, 5*, 191–210. https://doi.org/10.1007/s11412-010-9083-7.

Heckhausen, H. (1974). *Leistung und Chancengleichheit [Achievement and equal opportunities]*. Göttingen, Germany: Hogrefe.

Hedges, L. V. (1981). Distribution theory for Glasś estimator of effect size and related estimators. *Journal of Educational Statistics, 6*, 107–128. https://doi.org/10.3102/10769986006002107.

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. https://doi.org/10.1111/medu.12778.

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal, 327*, 557–560. https://doi.org/10.1136/bmj.327.7414.557.

*Hron, A., Hesse, F. W., Cress, U., & Giovis, C. (2000). Implicit and explicit dialogue structuring in virtual learning groups. *British Journal of Educational Psychology, 70*, 53–64. https://doi.org/10.1348/000709900157967.

*Hsu, C.-C., Chiu, C.-H., Lin, C.-H., & Wang, T.-I. (2015). Enhancing skill in constructing scientific explanations using a structured argumentation scaffold in scientific inquiry. *Computers & Education, 91*, 46–59. https://doi.org/10.1016/j.compedu.2015.09.009.

*Huang, H.-W., Wu, C.-W., & Chen, N.-S. (2012). The effectiveness of using procedural scaffoldings in a paper-plus-smartphone collaborative learning context. *Computers & Education, 59*, 250–259. https://doi.org/10.1016/j.compedu.2012.01.015.

Järvelä, S., & Hadwin, A. F. (2013). New frontiers: Regulating learning in CSCL. *Educational Psychologist, 48*, 25–39. https://doi.org/10.1080/00461520.2012.748006.

Järvelä, S., Volet, S., & Järvenoja, H. (2010). Research on motivation in collaborative learning: Moving beyond the cognitive–situative divide and combining individual and social processes. *Educational Psychologist, 45*(1), 15–27. https://doi.org/10.1080/00461520903433539.

Jeong, H., Hmelo-Silver, C. E., & Jo, K. (2019). Ten years of computer-supported collaborative learning: A metaanalysis of CSCL in STEM education during 2005–2014. *Educational Research Review, 28*, 100284. https://doi.org/10.1016/j.edurev.2019.100284.

King, A. (2007). Scripting collaborative learning processes: A cognitive perspective. In F. Fischer, I. Kollar, H. Mandl, & J. M. Haake (Eds.), *Scripting Computer-Supported Collaborative Learning: Cognitive, Computational and Educational Perspectives* (pp. 13–37). Bosten, MA: Springer US.

Kirschner, F., Paas, F., & Kirschner, P. A. (2009). A cognitive load approach to collaborative learning: United brains for complex tasks. *Educational Psychological Review, 21*, 31–42. https://doi.org/10.1007/s10648-008-9095-2.

Kirschner, P. A., Sweller, J., Kirschner, F., Zambrano, R., & J. (2018). From cognitive load theory to colaborative cognitive load theoy. *International Journal of Computer-Supported Collaborative Learning, 13*, 213–233. https://doi.org/10.1007/s11412-018-9277-y.

Kollar, I., Fischer, F., & Hesse, F. W. (2006). Collaboration scripts - a conceptual analysis. *Educational Psychology Review, 18*, 159–185. https://doi.org/10.1007/s10648-006-9007-2.

*Kollar, I., Fischer, F., & Slotta, J. D. (2007). Internal and external scripts in computer-supported collaborative inquiry learning. *Learning and Instruction, 17*, 708–721. https://doi.org/10.1016/j.learninstruc.2007.09.021.

*Kollar, I., Ufer, S., Reichersdorfer, E., Vogel, F., Fischer, F., & Reiss, K. (2014). Effects of collaboration scripts and heuristic worked examples on the acquisition of mathematical argumentation skills of teacher students with different levels of prior achievement. *Learning and Instruction, 32*, 22–36. https://doi.org/10.1016/j.learninstruc.2014.01.003.

Krapp, A., Geyer, C., & Lewalter, D. (2014). Motivation und emotion [motivation and emotion]. In T. Seidel & A. Krapp (Eds.), *Pädagogische Psychologie [Educational psychology]* (pp. 193–222). Weinheim, Germany: Beltz Verlag.

Latané, B., Williams, K., & Harkins, S. (1979). Many hands make light the work: The causes and consequences of social loafing. *Journal of Personality and Social Psychology, 37*, 822–832. https://doi.org/10.1037/0022-3514.37.6.822.

*Lee, Y.-H. (2015). Facilitating critical thinking using the C-QRAC collaboration script: Enhancing science reading literacy in a computer-supported collaborative learning environment. *Computers & Education, 88*, 182–191. https://doi.org/10.1016/j.compedu.2015.05.004.

*Lee, Y.-H. (2018). Scripting to enhance university students' critical thinking in flipped learning: Implications of the delayed effect on science reading literacy. *Interactive Learning Environments, 26*(5), 569–582. https://doi.org/10.1080/10494820.2017.1372483.

*Lin, G.-Y. (2020). Scripts and mastery goal orientation in face-to-face versus computer-mediated collaborative learning: Influence on performance, affective and motivational outcomes, and social ability. *Computers & Education, 143*, 103691. https://doi.org/10.1016/j.compedu.2019.103691.

*Lin, C. H., Chiu, C. H., Hsu, C. C., Wang, T. I., & Chen, C. H. (2018). The effects of computerized inquiry-stage-dependent argumentation assistance on elementary students' science process and argument construction skills. *Journal of Computer Assisted Learning, 34*(3), 279–292. https://doi.org/10.1111/jcal.12241.

Liu, L., Hao, J., von Davier, A. A., Kyllonen, P., & Zapata-Rivera, D. (2015). A tough nut to crack: Measuring collaborative problem solving. In Y. Rosen, S. Ferrara, & M. Mosharraf (Eds.), *Handbook of research on computational tools for real-world skill development* (pp. 344–359). Hershey, PA: IGI Global.

Lu, L., Yuan, Y. C., & McLeod, P. L. (2012). Twenty-five years of hidden profiles in group decision making: A meta-analysis. *Personality and Social Psychology Review, 16*, 54–75. https://doi.org/10.1177/1088868311417243.

Mäkitalo, K., Weinberger, A., Häkkinen, P., Järvelä, S., & Fischer, F. (2005). Epistemic cooperation scripts in online learning environments: Fostering learning by reducing uncertainty in discourse? *Computers in Human Behavior, 21*, 603–622. https://doi.org/10.1016/j.chb.2004.10.033.

McShane, B. B., Böckenholt, U., & Hansen, K. T. (2016). Adjusting for publication bias in meta-analysis: An evaluation of selection methods and some cautionary notes. *Perspectives on Psychological Science, 11*, 730–749. https://doi.org/10.1177/1745691616662243.

*Mende, S., Proske, A., Körndle, H., & Narciss, S. (2017). Who benefits from a low versus high guidance CSCL script and why? *Instructional Science, 45*(4), 439–468.

Molinari, G., Sangin, M., Dillenbourg, P., & Nüssli, M.-A. (2009). Knowledge interdependence with the partner, accuracy of mutual knowledge model and computer-supported collaborative learning. *European Journal of Psychology of Education, 24*, 129–144. https://doi.org/10.1007/BF03173006.

Noroozi, O., Biemans, H. J. A., Weinberger, A., Mulder, M., & Chizari, M. (2013a). Scripting for construction of a transactive memory system in multidisciplinary CSCL environments. *Learning and Instruction, 25*, 1–12. https://doi.org/10.1016/j.learninstruc.2012.10.002.

*Noroozi, O., Teasley, S. D., Biemans, H. J. A., Weinberger, A., & Mulder, M. (2013b). Facilitating learning in multidisciplinary groups with transactive CSCL scripts. *International Journal of Computer-Supported Collaborative Learning, 8*, 189–223. https://doi.org/10.1007/s11412-012-9162-z.
*Noroozi, O., Weinberger, A., Biemans, H. J. A., Mulder, M., & Chizari, M. (2013c). Facilitating argumentative knowledge construction through a transactive discussion script in CSCL. *Computers & Education, 61*, 59–76. https://doi.org/10.1016/j.compedu.2012.08.013.
Olsen, J. K., Rummel, N., & Aleven, V. (2017). Learning Alone or Together? A Combination Can Be Best! In B. K. Smith, M. Borge, E. Mercier, & K. Y. Lim (Eds.), *Making a Difference: Prioritizing Equity and Access in CSCL. 12th International Conference on Computer Supported Collaborative Learning (CSCL) 2017, Volume 1*. Philadelphia, PA: International Society of the Learning Sciences.
*Peterson, A. T., & Roseth, C. J. (2016). Effects of four CSCL strategies for enhancing online discussion forums: Social interdependence, summarizing, scripts, and synchronicity. *International Journal of Educational Research, 76*, 147–161. https://doi.org/10.1016/j.ijer.2015.04.009.
*Puhl, T., Tsovaltzi, D., & Weinberger, A. (2015). Blending Facebook discussions into seminars for practicing argumentation. *Computers in Human Behavior, 53*, 605–616. https://doi.org/10.1016/j.chb.2015.04.006.
R Core Team. (2019). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/.
*Rau, M. A., Bowman, H. E., & Moore, J. W. (2017). An adaptive collaboration script for learning with multiple visual representations in chemistry. *Computers & Education, 109*, 38–55. https://doi.org/10.1016/j.compedu.2017.02.006.
*Rienties, B., Giesbers, B., Tempelaar, D., Lygo-Baker, S., Segers, M., & Gijselaers, W. (2012). The role of scaffolding and motivation in CSCL. *Computers & Education, 59*, 893–906. https://doi.org/10.1016/j.compedu.2012.04.010.
Roschelle, J., & Teasley, S. D. (1995). The construction of shared knowledge in collaborative problem solving. In C. Ońalley (Ed.), *Computer-supported collaborative learning* (pp. 69–97). Berlin, Heidelberg, Germany: Springer.
*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*, 201–241. https://doi.org/10.1207/s15327809jls1402_2.
*Rummel, N., Spada, H., & Hauser, S. (2009). Learning to collaborate while being scripted or by observing a model. *International Journal of Computer-Supported Collaborative Learning, 4*, 69–92. https://doi.org/10.1007/s11412-008-9054-4.
*Rummel, N., Mullins, D., & Spada, H. (2012). Scripted collaborative learning with the cognitive tutor algebra. *International Journal of Computer-Supported Collaborative Learning, 7*, 307–339. https://doi.org/10.1007/s11412-012-9146-z.
Ryan, R. M., & Deci, E. L. (2000). The darker and brighter sides of human existence: Basic psychological needs as a unifying concept. *Psychological Inquiry, 11*, 319–338. https://doi.org/10.1207/S15327965PLI1104_03.
Ryan, R. M., & Deci, E. L. (2008a). Self-determination theory and the role of basic psychological needs in personality and the organization of behavior. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (pp. 654–678). New York, NY: Guilford press.
Ryan, R. M., & Deci, E. L. (2008b). Self-determination theory: A macrotheory of human motivation, development, and health. *Canadian Psychology, 49*(3), 182–185. https://doi.org/10.1037/a0012801.
Schank, R. C. (1999). *Dynamic Memory Revisited*. Cambridge, UK: Cambridge University Press.
Schellens, T., van Keer, H., Wever, B. d., & Valcke, M. (2007). Scripting by assigning roles: Does it improve knowledge construction in asynchronous discussion groups? *International Journal of Computer-Supported Collaborative Learning, 2*(2), 225–246. https://doi.org/10.1007/s11412-007-9016-2.
*Scheuer, O., McLaren, B. M., Weinberger, A., & Niebuhr, S. (2014). Promoting critical, elaborative discussions through a collaboration script and argument diagrams. *Instructional Science, 42*, 127–157. https://doi.org/10.1007/s11251-013-9274-5.
*Schmitt, L. J., & Weinberger, A. (2019). Fourth graders' dyadic learning on multi-touch interfaces—Versatile effects of verbalization prompts. *Educational Technology Research and Development, 67*(3), 519–539. https://doi.org/10.1007/s11423-018-9619-5.
Schnake, M. E. (1991). Equity in effort: The "sucker effect" in co-acting groups. *Journal of Management, 17*, 41–55. https://doi.org/10.1177/014920639101700104.
Schwaighofer, M., Vogel, F., Kollar, I., Ufer, S., Strohmaier, A., Terwedow, I., Ottinger, S., Reiss, K., & Fischer, F. (2017). How to combine collaboration scripts and heuristic worked examples to foster mathematical argumentation – When working memory matters. *International Journal of Computer-Supported Collaborative Learning, 12*(3), 281–305. https://doi.org/10.1007/s11412-017-9260-z.

Stasser, G., & Titus, W. (1985). Pooling of unshared information in group decision making: Biased information sampling during discussion. *Journal of Personality and Social Psychology, 48*, 1467–1467. https://doi.org/10.1037/0022-3514.48.6.1467.

*Stegmann, K., Weinberger, A., & Fischer, F. (2007). Facilitating argumentative knowledge construction with computer-supported collaboration scripts. *International Journal of Computer-Supported Collaborative Learning, 2*, 421–447. https://doi.org/10.1007/s11412-007-9028-y.

*Stegmann, K., Wecker, C., Weinberger, A., & Fischer, F. (2012). Collaborative argumentation and cognitive elaboration in a computer-supported collaborative learning environment. *Instructional Science, 40*, 297–323. https://doi.org/10.1007/s11251-011-9174-5.

Tabak, I. (2004). Synergy: A complement to emerging patterns of distributed scaffolding. *The Journal of the Learning Sciences, 13*, 305–335. https://doi.org/10.1016/j.chb.2018.12.010.

*Taylor, K., & Baek, Y. (2019). Grouping matters in computational robotic activities. *Computers in Human Behavior, 93*, 99–105. https://doi.org/10.1016/j.chb.2018.12.010

Teasley, S. D. (1997). Talking about reasoning: How important is the peer in peer collaboration? In L. R. Resnick, R. Säljö, C. Pontecorvo, & B. Burge (Eds.), *Discourse, tools and reasoning: Essays on situated cognition* (pp. 361–384). Berlin, Heidelberg, Germany: Springer.

Teasley, S. D., Fischer, F., Weinberger, A., Stegmann, K., Dillenbourg, P., Kapur, M., & Chi, M. T. H. (2008). Cognitive convergence in collaborative learning. *Proceedings of the 8th international conference on International conference for the learning sciences - Volume* 3, pp. 360–367.

Tschan, F., Semmer, N. K., Gurtner, A., Bizzari, L., Spychiger, M., Breuer, M., & Marsch, S. U. (2009). Explicit reasoning, confirmation bias, and illusory transactive memory: A simulation study of group medical decision making. *Small Group Research, 40*, 271–300. https://doi.org/10.1177/1046496409332928.

*Tsovaltzi, D., Puhl, T., Judele, R., & Weinberger, A. (2014). Group awareness support and argumentation scripts for individual preparation of arguments in Facebook. *Computers & Education, 76*, 108–118. https://doi.org/10.1016/j.compedu.2014.03.012.

*Tsovaltzi, D., Judele, R., Puhl, T., & Weinberger, A. (2015). Scripts, individual preparation and group awareness support in the service of learning in Facebook: How does CSCL compare to social networking sites? *Computers in Human Behavior, 53*, 577–592. https://doi.org/10.1016/j.chb.2015.04.067.

*Ulicsak, M. H. (2004). 'How did it know we werent talking?': An investigation into the impact of self-assessments and feedback in a group activity. *Journal of Computer Assisted Learning, 20*, 205–211. https://doi.org/10.1111/j.1365-2729.2004.00083.x.

Urhahne, D. (2008). Sieben Arten der Lernmotivation: Ein Überblick über zentrale Forschungskonzepte [seven kinds of learning motivation. An overview of central concept of research]. *Psychologische Rundschau, 59*, 150–166. https://doi.org/10.1026/0033-3042.59.3.150.

*van Aalst, J., & Chan, C. K. K. (2007). Student-directed assessment of knowledge building using electronic portfolios. *The Journal of the Learning Sciences, 16*, 175–220. https://doi.org/10.1080/10508400701193697.

*van Dijk, A. M., Gijlers, H., & Weinberger, A. (2014). Scripted collaborative drawing in elementary science education. *Instructional Science, 42*, 353–372. https://doi.org/10.1007/s11251-013-9286-1.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*(3), 1–48.

*Vogel, F., Kollar, I., Ufer, S., Reichersdorfer, E., Reiss, K., & Fischer, F. (2016). Developing argumentation skills in mathematics through computer-supported collaborative learning: The role of transactivity. *Instructional Science, 44*, 477–500. https://doi.org/10.1007/s11251-016-9380-2.

Vogel, F., Wecker, C., Kollar, I., & Fischer, F. (2017). Socio-cognitive scaffolding with computer-supported collaboration scripts: A meta-analysis. *Educational Psychology Review, 29*, 477–511. https://doi.org/10.1007/s10648-016-9361-7.

Wang, H.-C., Rosé, C. P., & Chang, C.-Y. (2011). Agent-based dynamic support for learning from collaborative brainstorming in scientific inquiry. *International Journal of Computer-Supported Collaborative Learning, 6*(3), 371–395. https://doi.org/10.1007/s11412-011-9124-x.

Wecker, C., & Fischer, F. (2011). From guided to self-regulated performance of domain-general skills: The role of peer monitoring during the fading of instructional scripts. *Learning and Instruction, 21*(6), 746–756. https://doi.org/10.1016/j.learninstruc.2011.05.001.

*Weinberger, A., Ertl, B., Fischer, F., & Mandl, H. (2005). Epistemic and social scripts in computer-supported collaborative learning. *Instructional Science, 33*, 1–30. https://doi.org/10.1007/s11251-004-2322-4.

Weinberger, A., Stegmann, K., Fischer, F., & Mandl, H. (2007). Scripting argumentative knowledge construction in computer-supported learning environments. In F. Fischer, I. Kollar, H. Mandl, & J. M. Haake (Eds.), *Scripting Computer-Supported Collaborative Learning: Cognitive, Computational and Educational Perspectives* (pp. 191–211). Boston, MA: Springer.

Weinberger, A., Kollar, I., Dimitriadis, Y., Mäkitalo-Siegl, K., & Fischer, F. (2009). Computer-supported collaboration scripts: Perspectives from educational psychology and computer science. In N. Balacheff, S. Ludvigsen, T. De Jong, A. Lazonder, S.-A. Barnes, & L. Montandon (Eds.), *Technology-enhanced learning - principles and products* (pp. 155–173). Dordrecht, Netherlands: Springer Netherlands.

*Weinberger, A., Stegmann, K., & Fischer, F. (2010). Learning to argue online: Scripted groups surpass individuals (unscripted groups do not). *Computers in Human Behavior, 26*, 506–515. https://doi.org/10.1016/j.chb.2009.08.007.

Wise, A. F., & Schwarz, B. B. (2017). Visions of CSCL: Eight provocations for the future of the field. *International Journal of Computer-Supported Collaborative Learning, 12*, 423–467. https://doi.org/10.1007/s11412-017-9267-5.

Wittwer, J., & Renkl, A. (2010). How effective are instructional explanations in example-based learning? A meta-analytic review. *Educational Psychological Review, 22*, 393–409. https://doi.org/10.1007/s10648-010-9136-5.

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry, 17*, 89–100. https://doi.org/10.1111/j.1469-7610.1976.tb00381.x.

*Wu, S. P., Corr, J., & Rau, M. A. (2019). How instructors frame students' interactions with educational technologies can enhance or reduce learning with multiple representations. *Computers & Education, 128*, 199–213. https://doi.org/10.1016/j.compedu.2018.09.012.

Zhuang, X., MacCann, C., Wang, L., Liu, L., & Roberts, R. D. (2008). Development and validity evidence supporting a teamwork and collaboration assessment for high school students. *ETS Research Report Series, 2*. https://doi.org/10.1002/j.2333-8504.2008.tb02136.x.

4 Study 3: Learning to diagnose collaboratively – Effects of adaptive collaboration scripts in agent-based medical simulations

Anika Radkowitsch * Michael Sailer * Martin R. Fischer * Ralf Schmidmaier * Frank Fischer

Reference: Radkowitsch, A., Sailer, M., Fischer, M. R., Schmidmaier, R., & Fischer, F. (2021). Learning to diagnose collaboratively – Effects of adaptive collaboration scripts in agent-based medical simulations. *Learning and Instruction, 75*. Doi: 10.1016/j.learninstruc.2021.101487.

Learning and Instruction 75 (2021) 101487

Contents lists available at ScienceDirect

Learning and Instruction

journal homepage: www.elsevier.com/locate/learninstruc

Learning to diagnose collaboratively – Effects of adaptive collaboration scripts in agent-based medical simulations

Anika Radkowitsch [a,b,*], Michael Sailer [a,b], Ralf Schmidmaier [a,c], Martin R. Fischer [a,d], Frank Fischer [a,b]

[a] *Munich Center of the Learning Sciences (MCLS), Ludwig-Maximilians-Universität München, Munich, Germany*
[b] *Department Psychologie, Ludwig-Maximilians-Universität München, Munich, Germany*
[c] *Medizinische Klinik und Poliklinik IV, LMU Klinikum, Ludwig-Maximilians-Universität München, Munich, Germany*
[d] *Institut für Didaktik und Ausbildungsforschung in der Medizin, LMU Klinikum, Ludwig-Maximilians-Universität München, Munich, Germany*

A R T I C L E I N F O

Keywords:
Adaptive collaboration scripts
Basic psychological needs
Collaborative diagnostic
Reasoning
Medical education

A B S T R A C T

We investigated how medical students' collaborative diagnostic reasoning, particularly evidence elicitation and sharing, can be facilitated effectively using agent-based simulations. Providing adaptive collaboration scripts has been suggested to increase effectiveness, but existing evidence is diverse and could be affected by unsystematic group constellations. Collaboration scripts have been criticized for undermining learners' agency. We investigate the effect of adaptive and static scripts on collaborative diagnostic reasoning and basic psychological needs. We randomly allocated 160 medical students to one of three groups: adaptive, static, or no collaboration script. We found that learning with adaptive collaboration scripts enhanced evidence sharing performance and transfer performance. Scripting did not affect learners' perceived autonomy and social relatedness. Yet, compared to static scripts, adaptive scripts had positive effects on perceived competence. We conclude that for complex skills complementing agent-based simulations with adaptive scripts seems beneficial to help learners internalize collaboration scripts without negatively affecting basic psychological needs.

Diagnosing collaboratively is part of many physicians' daily routines. For instance, physicians discuss their patients' symptoms with other physicians from different medical subspecialities. Although collaborative diagnostic reasoning is a crucial competence in routine medical care, empirical research is largely lacking (Kiesewetter, Fischer, & Fischer, 2017). The few available studies suggest that collaborative diagnosing is difficult, and physicians often fail to pool their knowledge appropriately, which can lead to wrong diagnoses (Tschan et al., 2009). Hence, understanding and facilitating collaborative diagnostic reasoning, and particularly sharing and elicitation of information, seems necessary. Simulations have been found to be effective for enhancing the learning of complex skills (Gegenfurtner, Quesada-Pallarès, & Knogler, 2014). Simulations are models of real-world scenarios in which learners can act as if they were in that situation, thereby practicing complex skills (Gegenfurtner et al., 2014). However, particularly during early stages of the development of complex skills, it seems beneficial to provide learners with additional scaffolding beyond providing realistic problem solving opportunities (Chernikova et al., 2020). Collaboration scripts are scaffolds that structure collaboration (Fischer, Kollar, Stegmann, & Wecker, 2013) and were found to be effective for facilitating collaborative learning (Radkowitsch, Vogel, & Fischer, 2020). Yet, collaboration scripts have been criticized for being too coercive, by that reducing learners' self-determination and thus reducing motivation and impairing learning (Dillenbourg, 2002). A promising solution for the respective criticism seems to provide adaptive support that adjusts to the learners' needs. However, the evidence for the effectiveness of adaptive scaffolding to increase motivation and learning is ambiguous (Stegmann, Mu, Gehlen-Baum, & Fischer, 2011). Hence, the present study addresses the questions to what extent adaptive collaboration scripts (1) enhance the learning of complex skills such as collaborative diagnostic reasoning in medicine and (2) ensure self-determination when learning with simulations.

* Corresponding author. Department Psychologie, Ludwig-Maximilians-Universität München, Leopoldstr. 13, 80802 München, Germany.
E-mail addresses: Anika.Radkowitsch@psy.lmu.de (A. Radkowitsch), michael.sailer@psy.lmu.de (M. Sailer), ralf.schmidmaier@med.uni-muenchen.de (R. Schmidmaier), martin.fischer@med.uni-muenchen.de (M.R. Fischer), frank.fischer@psy.lmu.de (F. Fischer).

https://doi.org/10.1016/j.learninstruc.2021.101487
Received 20 May 2020; Received in revised form 22 February 2021; Accepted 21 March 2021
Available online 19 April 2021

1. Background

1.1. Collaborative diagnostic reasoning

As in other collaborative problem-solving contexts (e.g., OECD, 2017), physicians need to be able to diagnose individually and to engage in collaborative activities when diagnosing collaboratively. For a proper understanding of collaborative diagnostic reasoning, examining both, individual and collaborative cognitive processes as well as their interaction is thus necessary. *Individual* diagnostic reasoning is an epistemic process (Heitzmann et al., 2019) with the goal to identify an accurate diagnosis (Simmons, 2010) and to reduce uncertainty to the degree that enables taking appropriate action (Charlin et al., 2012). Generally, epistemic reasoning processes have been described as coordination between hypotheses and evidence by generating hypotheses and deriving predictions from and testing of hypotheses in the light of evidence (Klahr & Dunbar, 1988). In the context of medical diagnostic reasoning, physicians suggest differential diagnoses (i.e., hypotheses) based on findings and symptoms (i.e., evidence) which in turn allow deriving predictions about further findings and symptoms. To test these predictions, the generation of further evidence is often crucial. For example, physicians perform specific physical examinations or conduct laboratory tests to get more information about the patient's health status based on prior hypotheses (Charlin et al., 2012). These diagnostic reasoning processes are based on different types of knowledge such as conceptual biomedical or strategic knowledge. These types of knowledge and their efficient organization are essential for efficient diagnostic reasoning processes (Feltovich & Barrows, 1984; Klein, Otto, Fischer, & Stark, 2019; Stark, Kopp, & Fischer, 2011). According to the illness script theory, medical knowledge is stored in so-called *illness scripts* which organize medical knowledge based on disease entities, their underlying pathophysiological processes, the resulting signs and symptoms, and enabling conditions. An efficient organization develops with increasing medical experience which allows to relate a patient's signs, symptoms, and enabling conditions to the respective disease. Thus, illness-scripts allow physicians to diagnose accurately and fast based on pattern recognition (Charlin et al., 2007; Feltovich & Barrows, 1984).

Yet, complex cases often require that different medical experts combine their efforts to diagnose the patient *collaboratively*. For example, attending physicians do not generate additional evidence themselves but consult more specialized diagnosing physicians such as radiologists or pathologists to generate and evaluate evidence. In such situations, collaboration has the function to pool knowledge and skills distributed among collaborators in order to reach the solution of a problem (OECD, 2017). Collaborators do so by engaging in collaborative activities of which sharing, elicitation, negotiation, and regulation were particularly in the focus of recent models of collaboration (e.g., Hesse, Care, Buder, Sassenberg, & Griffin, 2015; Liu, Hao, von Davier, Kyllonen, & Zapata-Rivera, 2015; OECD, 2017; Sun et al., 2020). By sharing and eliciting information, collaborators contribute to both the processing of information on a group level (Hinsz, Tindale, & Vollrath, 1997), as well as the formation of a shared mental representation of the problem and its possible solutions (Meier, Spada, & Rummel, 2007; Roschelle & Teasley, 1995) and requires collaborators to take the audience's background into account. Negotiating is particularly important in case of conflicts and can prevent groups from premature closure or from ignoring dissenting evidence (Nickerson, 1988; Patel, Kaufman, & Arocha, 2002). Regulation refers to coordinating goals and strategies to reach these goals and requires collaborators to reflect on their own collaborative activities (Järvelä & Hadwin, 2013). In short, when diagnosing collaboratively, physicians share, negotiate, or elicit information and the results of the individual diagnostic activities (e.g., evidence or hypotheses) and coordinate their collaborative diagnostic reasoning. By that physicians pool their knowledge and effort in order to reach the common goal which is, for instance, to identify the patients' disease or a suitable treatment.

In which collaborative activities physicians engage in depends on the situational needs. However, the pooling of information (i.e., sharing and elicitation) has received much attention in psychological and medical research as it has been found to be difficult but, at the same time, highly relevant for the success of collaboration. A number of studies in the medical context showed that physicians often fail to share or elicit crucial patient information among each other which negatively affected individual diagnostic reasoning processes such as the generation of hypotheses (Larson, Christensen, Franz, & Abbot, 1998; Tschan et al., 2009), or the generation and evaluation of evidence (Davies et al., 2018). Brady, Laoide, McCarthy, & McDermott (2012) showed that in radiology diagnostic errors often occur due to missing medical information. For instance, radiologists are not informed about prior surgeries or secondary diagnoses which may lead to misinterpretations of radiologic evidence and, thus, mislead their diagnostic reasoning process. Therefore, although general collaborative problem-solving processes often also require negotiating and regulation, in this paper we focus on the sharing and elicitation of evidence when collaboratively generating evidence since these collaborative diagnostic processes seem crucial but often deficiently functioning.

Under which conditions are collaborators successful? Different strands of research (e.g., transactive memory theory, shared mental models) highlight the relevance of *meta-knowledge*, that is knowledge about the other team members' roles, their responsibilities and their tasks (Cannon-Bowers, Salas, & Converse, 1993; Engelmann & Hesse, 2011; Wegner, 1987). Meta-knowledge indicates how knowledge is distributed among the team and allows team members to anticipate the team members' activities and to adapt the own activities accordingly. Thus, meta-knowledge is likely to affect which information is shared with or elicited from collaboration partners (Fiore et al., 2010).

In sum, collaborative diagnostic reasoning requires combining individual diagnostic skills such as the generation of evidence and collaborative skills such as sharing and elicitation of diagnostically relevant information. To successfully diagnose collaboratively, physicians need to apply medical knowledge as well as knowledge about the team members' roles and responsibilities to a patient case.

1.2. Facilitating collaborative diagnostic reasoning with agent-based simulations

Offering learners opportunities to apply their knowledge to realistic cases is considered crucial for the development of complex skills (Kolodner, 1992) such as collaborative diagnostic reasoning. By applying their knowledge to cases, physicians reorganize and encapsulate their knowledge to build efficient illness scripts (Feltovich & Barrows, 1984). Simulations provide learners with the opportunity to apply knowledge to specific cases in standardized settings (e.g., rare diseases, situations with high stakes) while simultaneously allowing to reduce the complexity of a real practice situation (Siebeck et al., 2011). Typically, these simulations focus on a smaller range of sub-skills, and allow for repetition, error, time-outs, and systematic debriefing. Several systematic reviews have been conducted, concluding that learning with simulations is beneficial for the development of complex skills in a broad range of conditions (Cook, 2014; Gegenfurtner et al., 2014). Thus, simulations offer beneficial learning opportunities by engaging learners with important aspects of a task, enabling them to apply knowledge and practice certain skills. However, prior research has repeatedly shown that unsupported problem solving is likely to overwhelm learners, particularly during early phases of skill development (Belland, Walker, Kim, & Lefler, 2017; Kirschner, Sweller, & Clark, 2006). Designing simulation-based learning environments with additional instructional support such as scaffolding has been found to be a promising way to enhance further the effectiveness of simulations (Chernikova et al., 2020).

When simulating collaborative tasks, a recent approach is the collaboration with simulated computer agents (Graesser et al., 2018). In agent-based collaboration, one or more learners collaborate with one or more computer agents which are computer programs designed to act similar to humans to solve a problem or task (Rosen, 2015). Such agents can fulfill different roles, for example as pedagogical agent (e.g., AutoTutor, Nye, Graesser, & Hu, 2014), or as substitute for a collaboration partner (e.g., Stadler, Herborn, Mustafić, & Greiff, 2019). Simulating collaboration partners during collaborative tasks allows controlling the effect of possibly influencing variables such as motivation or expertise of the collaborators, thus offering highly standardized research settings (Graesser et al., 2018). The targeted aspects of collaboration are often selected based on empirical analyses showing a particular need for advancing these skills.

1.3. *Scaffolding collaborative diagnostic reasoning with collaboration scripts*

To learn complex skills, instructional support beyond providing a problem scenario seems beneficial (Wood, Bruner, & Ross, 1976). External collaboration scripts are a scaffold used for facilitating interaction during collaborative learning by prompting specific collaborative activities (Fischer et al., 2013). For example, learners receive prompts to share particular information (Noroozi, Biemans, Weinberger, Mulder, & Chizari, 2013). The idea of external collaboration scripts is based on the *script theory of guidance* (Fischer et al., 2013), which assumes that internal collaboration scripts guide any behavior and cognition concerning collaboration. Such internal collaboration scripts are assumed to structure knowledge about specific collaborative practices. During collaborative practices, this knowledge is flexibly activated depending on the situational characteristics and the collaborators' goals. External collaboration scripts complement less functional cognitive script components. By engaging learners in beneficial collaborative activities, external collaboration scripts can facilitate the internalization of functional script components (Fischer et al., 2013). Beyond these theoretical considerations, there is empirical evidence for the effectiveness of collaboration scripts for advancing the development of collaboration skills. For example, Rummel and Spada (2005) found that dyads of medical students and psychological students who were supported with a collaboration script showed less deviations from an exemplary collaborative process compared to dyads not supported by an external collaboration script. In a meta-analysis, Radkowitsch et al. (2020) found that collaboration scripts effectively facilitate the learning of collaboration skills in the context of computer-supported collaborative learning. Yet, most of the prior studies targeted advancing domain learning, and none of the studies employed simulations systematically to advance collaboration skills. Thus, it is not clear whether collaboration scripts are effective when implemented in simulations and used to advance profession-specific skills such as sharing or elicitation of evidence.

1.3.1. *Adaptive collaboration scripts*

As described above, external collaboration scripts aim to complement less functional internal collaboration scripts. That means that to be effective, collaboration scripts should be adapted to the learners' actual proficiency level (Fischer et al., 2013). When providing too detailed guidance – for which collaboration scripts have been criticized (Dillenbourg, 2002) – the learning of more advanced learners is hindered because it restricts the learners' natural collaboration processes (i.e., their internal collaboration scripts). This phenomenon became known as *over-scripting* (Dillenbourg, 2002). Yet, most studies that aim to avoid over-scripting and to take learners' zone of proximal development (Vygotsky, 1978) into account, use static techniques such as fading. For example, Stegmann et al. (2011) compared static and fading collaboration scripts supporting argumentation to unstructured collaborative learning. They found that learning with static and fading collaboration scripts enhanced the quality of argumentation with descriptively larger effects for learning with the static collaboration script. By fading after a predefined sequence, learners' internal collaboration script could have been misaddressed, posing unnecessary extraneous cognitive load on learners. Consequently, learners could have failed to internalize the collaboration scripts (Wecker & Fischer, 2011).

Thus, to exploit the full potential of adaptive support modeling learners' collaborative activities and compare them to an ideal collaborative process seems beneficial. In the case of previously defined deviations, learners receive just-in-time support in the form of so-called adaptive collaboration scripts (Karakostas & Demetriadis, 2011; Tchounikine, Rummel, & McLaren, 2010). There are different ways to implement such adaptive support with varying degrees of complexity. For example, in a small-scale study on conducting chemical experiments, Tsovaltzi et al. (2010) used a wizard of Oz approach in which a human adept intervened in the collaborative process in predefined situations. Other studies used the amount of contribution of each collaborator (Constantino-Gonzalez, Suthers, & de los Santos, 2003), or learners' automatically assessed problem-solving strategies (Diziol, Walker, Rummel, & Koedinger, 2010) as indices for quality of collaboration. Yet, studies on adaptive collaboration scripts are scarce and provide no clear picture of their effectiveness for advancing collaboration. A reason for this could be that research on collaboration often takes place in rather unstandardized human-to-human collaborative settings. That means that the composition of learners affects collaboration processes (Kirschner, Sweller, Kirschner, & Zambrano, 2018) and consequently could also influence the effects of collaboration scripts. Thus, the reported effects of collaboration scripts could be affected by their particular implementation and by the high noise due to several real collaboration partners. Therefore, an agent-based realization of collaboration provides a standardized setting for investigating the effect of adaptive collaboration scripts.

1.3.2. *Collaboration scripts and psychological need satisfaction*

Collaboration scripts were criticized for restricting learners and thereby negatively affecting learners' self-determination and, thus, intrinsic motivation (Dillenbourg, 2002). This criticism relates to the self-determination theory (Deci & Ryan, 1985) that describes that the feelings of autonomy, competence, and social relatedness are basic psychological needs and thus crucial determinants of intrinsic motivation and, consequently, human behavior. The feeling of autonomy refers to whether the learners perceive their actions as congruent and volitional. The feeling of competence relates to the feeling of efficacy of own actions. The perception of social relatedness refers to a feeling of being connected with a specific community (Deci & Ryan, 1985). Whereas critics argue that collaboration scripts reduce the perceived autonomy (Wise & Schwarz, 2017), collaboration scripts might also enhance the feeling of competence and social relatedness by enabling them to experience successful collaboration (Radkowitsch et al., 2020). Yet, the empirical evidence for a negative effect of collaboration scripts on self-determination or intrinsic motivation is scarce. In their meta-analysis, Radkowitsch et al. (2020) identified only six studies investigating the effect of collaboration scripts on motivation, which yielded a combined null effect on motivation.

If collaboration scripts negatively affected self-determination due to their limitation of choices, then collaboration scripts that are less structured should have a less negative impact on learners' self-determination. Stegmann et al. (2011) found initial evidence for that as learners supported either with a high or low structured collaboration script were descriptively less intrinsically motivated compared to learners who collaborated freely. However, this effect was not significant, and the study does not allow for differentiated conclusions on how learners' basic psychological needs were affected by learning with collaboration scripts. Besides, participants of the study could be novices and they could have required a higher degree of structure for successful collaboration. Then, the collaboration script should not automatically interfere with the learners' motivation. As the learning environment – and hence also the collaboration script – is considered crucial for fostering psychological need satisfaction (Deci & Vansteenkiste, 2004), we argue that collaboration scripts could have a negative impact on learners' perceived autonomy if the provided structure interferes with the learners' needs for support. Thus, adaptive collaboration scripts should affect learners' perceived autonomy less than collaboration scripts that are not adapted to the learners' specific needs for support. Concerning the feeling of competence, we assume that adaptive collaboration scripts are better tailored to the learners' needs, which could enable learners to adapt their collaborative activities better. Concerning the feeling of social relatedness, we assume that collaboration scripts could have positive effects. As learning with collaboration scripts could lead to equal participation and increase team functioning, we assume that both adaptive and static collaboration scripts increase the perception of social relatedness with higher effects of an adaptive collaboration script.

1.4. The present study

We seek to identify conditions under which facilitating collaborative diagnostic reasoning of medical students using agent-based simulation and information sharing scripts (ISS) is effective. In this study, ISS are collaboration scripts that focus on facilitating the sharing and elicitation of diagnostic information during collaboration. More particularly, we aim at facilitating the elicitation and sharing of evidence, which are considered essential subskills of collaborative diagnostic reasoning. The goals of the present study are twofold: We examine the effects of adaptive and static ISS on (1) collaborative diagnostic reasoning, and (2) the satisfaction of basic psychological needs. We pose the following research questions:

(1) What are the effects of an adaptive and a static ISS on collaborative diagnostic reasoning and more specifically on a) evidence elicitation and b) evidence sharing in an agent-based simulation? We hypothesize that both static and adaptive collaboration scripts have positive effects on a) evidence elicitation and b) evidence sharing compared to an unsupported control group, with adaptive ISS resulting in larger effects than static ISS.
(2) What are the effects of an adaptive and a static collaboration script on the basic psychological need satisfaction in the context of diagnosing collaboratively in an agent-based simulation? We hypothesize that both ISS have negative effects on the perceived autonomy but positive effects on the perceived competence and perceived social relatedness compared to an unsupported control group. Learning with adaptive ISS should result in higher perceived autonomy, higher perceived competence, and higher perceived social relatedness than static ISS.

2. Method

2.1. Sample and design

We conducted an experiment with a one-factorial design with the three *levels adaptive ISS, static ISS*, and no further instructional support (*control group*). Before recruiting participants, we conducted a power analysis based on medium effect sizes reported in prior studies (Radkowitsch et al., 2020). To detect a medium effect size of $f = 0.25$ in an ANCOVA design, a minimum sample of 159 participants (53 per group) was needed (presuming $\alpha = .05$, $1 - \beta = 0.80$). Medical students in their 3rd clinical year and higher were recruited to participate in the study voluntarily. These medical students have usually already completed medical clerkships in which collaboration with other physicians is necessary. However, systematic training of professional collaboration is not a formal part of medical education and their prior experience in medical collaboration can, thus, be described as low. Further, they had no experience with the agent-based simulations that we used in the experiment. All participants were randomly distributed to one of three groups. The final sample consisted of 160 participants ($N_{female} = 110$) of whom 54 were in the adaptive ISS condition and 53 in the static ISS and control condition each. Participants were, on average, 25 years old (SD = 3.1) and in their 5.24 years of medical school (SD = 1.07) of a 6-year program.

2.2. The agent-based medical simulation

To foster and measure collaborative diagnostic reasoning, we developed a simulation as well as cases of suitable fictitious patients suffering from fever of unknown origin. We chose the collaborative generation of evidence between internists and radiologists as simulated scenario since this is a situation that is common in emergency departments but often deficiently functioning (e.g., Brady et al. (2012)) and which offers a high potential for future standardization. In these situations, internists consult radiologists to generate evidence in order to reduce the uncertainty with respect to a specific diagnosis. Radiologists are better able to reliably contribute to the diagnostic process if internists precisely specify and justify the kind of evidence needed to reduce uncertainty (i.e., elicitation of evidence) and if internists report any relevant signs, symptoms, and prior conditions of the patient that could influence the radiologists' diagnostic process (i.e., sharing of evidence). As such, this situation does not represent a mere distribution of tasks since the quality of the activity of one person depends on the quality of collaborative activities of another person. For the internist to optimally collaborate with the radiologist, meta-knowledge of radiologists' tasks, role, and responsibilities is beneficial (e.g., Brady et al., 2012).

The simulation was developed in collaboration with software developers, medical educators, physicians, and psychologists and validated in expert workshops, a pilot study, as well as a comprehensive validation study comparing cognitive processes of medical students and experienced internists when working in the simulation (see Radkowitsch, Fischer, Schmidmaier, & Fischer, 2020). We implemented the simulation in the learning platform CASUS (http://www.casus.net) with which most participants already had prior experience in their curriculum.

Participants acted in the role of an internist working in an emergency department and were required to collaborate with a simulated radiologist to generate further evidence in each patient case. Overall, the interaction between the participant and the simulated radiologist in each case consists of filling in the request form and receiving e-mail-like text messages from the simulated radiologist containing a short introduction, the decline or acceptance of the request, and the report of radiological findings. The participants first received the electronic patient file containing information about the patient's admission, medical history, physical examination, and laboratory tests. In the next step, the participants requested a radiological test to generate further evidence about the cause of the presented symptoms in order to reduce uncertainty respective a potential diagnosis. For that, participants first contacted the radiologist by pressing the button "request radiological test". A simulated radiologist then sent an e-mail-like prompt shortly introducing herself as the radiologist on duty and asking to fill in a request form. Participants then requested the radiologic test by choosing a specific radiologic method (e.g., computer tomography) and a body part (e.g., abdomen) using the form. The request form further required participants to share patient information and suspected diagnoses that justified the test and helped the radiologist to interpret the radiologic findings. For that learners could tick off patient information from a thematically clustered list containing all information from the electronic patient file and type in diagnoses in a free text field. The test chosen by the participant specified the required information. For example, tests based on x-rays required, inter alia, information about a potential pregnancy. The minimum amount of information necessary to justify the request was determined in advance by medical experts. Only when participants shared sufficient patient information, the simulated radiologist actually conducted the respective test and interpreted the findings. Participants then received a detailed report on the generated evidence in form of a text message. Otherwise, the simulated radiologist rejected the request stating that the shared information was not sufficient to justify the specific radiologic test and asked the participant to revise and resubmit the request. Finally, the medical students concluded the patient case by suggesting and justifying a final diagnosis.

2.3. Treatment

We supported participants in the treatment groups either with an adaptive or a static ISS that were both provided in text messages by the simulated radiologist during the learning phase. Both ISS consisted of three types of prompts containing meta-knowledge (i.e., information about the radiologists' role, task, and responsibilities). Firstly, the ISS included general, case-independent details on the radiologists' task and information that is helpful for them to complete the task (first type of ISS prompt). For instance, the radiologist explains that the request should, inter alia, provide information on the patient's main symptoms and their course as this helps the radiologist to judge what and where to look for. This first type of ISS prompt addressed the learners' evidence sharing skill since it should help learners to precisely identify information relevant for the collaborating radiologist to complete their task. Secondly, the ISS contained case-specific meta-knowledge about how radiologists generate evidence for specific suspected diagnoses (second type of ISS prompt). This information was included for 45 differential diagnoses that were most relevant for the patient cases. For instance, the radiologist explains that if the patient is suspected to suffer from a pneumonia, the radiologist typically tries to differentiate air-filled parts in the lung from liquid-filled parenchyma, and that x-rays often are not sufficient for a differentiated evaluation due to overlays. This type of ISS prompt should help learners to specify and justify the type of evidence needed from the collaborator, the radiologist, and thus addresses the learners' evidence elicitation skill. Thirdly, the ISS provided meta-knowledge about specific radiologic examinations and about how such imaging procedures could potentially harm patients (third type of ISS prompt). Here, the radiologist explained which information helps radiologists to judge the risk of a specific radiologic test. For example, the radiologist explains that radiocontrast can have negative effects on the patients' kidneys which is why radiologists require the kidney status in order to weighing up the benefits and risks of using radiocontrast. This ISS prompt particularly addressed evidence sharing since learners are supported in their decision about whether a specific information is relevant for the collaborator (i.e., the radiologist) or not. As the simulation, the collaboration script prompts were developed in collaboration with experts from medical education, medicine (internists and radiologists), and psychology. The ISS prompts suggested to engage in specific collaborative processes and provided information about why these collaborative processes were meaningful. But the ISS prompts themselves did not force the participants to engage in a specific step at a specific point in time and, thus, can be described as low coercive.

In the *adaptive ISS* condition, participants received the first type of ISS prompt at the beginning of the interaction with the simulated radiologist during the introduction of the radiologist. The second and third prompts were provided by the simulated radiologist whenever the participants submitted a request that was not adequately justified according to the criteria described above. The simulated radiologist checked whether the presented diagnoses and symptoms were compatible with the requested test (second type of ISS prompt) and whether all necessary information for the respective radiologic tests was given (third type of ISS prompt) and answered with the respective prompts. For instance, when participants requested a test with radiocontrast but failed to share information on the kidney function, the request was rejected, and participants received a prompt providing meta-knowledge about the importance of kidney status for the radiologist and asking the learner to share respective information.

Participants in the *static ISS* condition received the ISS in form of a letter and a booklet from the radiologist at the beginning of the learning phase. In the letter, the radiologist first explained the general procedure of a radiologist (first type of ISS prompt) and which specific information was needed for potentially harmful radiologic tests (third type of ISS prompt). A booklet further provided meta-knowledge about the evaluation of specific diagnoses (second type of ISS prompt). The learners could access the letter and booklet any time and as often as they wanted to. When participants in the static ISS condition failed to adequately justify their request, their requests were also rejected by the radiologist. Participants were then required to find the relevant information in the letter and booklet for themselves. Thus, the main difference between the static and the adaptive ISS is that learners in the adaptive ISS condition receive only the ISS prompt the system identified as their current need for support whereas learners in the static ISS condition received all ISS prompts at once. Therefore, adaptive ISS should be less likely to interfere with the learners' need for autonomy compared to the static ISS. Moreover, since the adaptive ISS should help learners to implement the script prompts and consequently enhance the learners' perceived competence. Finally, in the adaptive ISS condition, the simulated radiologist reacts more directly to the learners' action. Therefore, learners supported with an adaptive ISS should perceive a higher social relatedness compared to learners in the static ISS condition. After the learning phase, learners were asked to return the letter and the booklet.

The requests of participants in the *control condition* were also rejected when they failed to share and elicit the necessary information. However, participants in the control condition did not receive any meta-knowledge prompts from the radiologist.

2.4. Dependent variables

2.4.1. Collaborative diagnostic reasoning performance

We assessed the participants' performance of *evidence sharing* and *evidence elicitation* as two subskills of collaborative diagnostic reasoning. For that, we used log files produced during the interaction with the simulation. The log files consist of all clicks and text entries of the users within the simulation, such as the patient information or diagnoses chosen to share with the radiologist. We used R Studio Version 4.0.2 (R Core Team, 2020) to automatically evaluate the quality of evidence elicitation and evidence sharing based on the expert solutions produced from medical experts. More specifically, the patient information and diagnoses shared, and radiologic tests requested from the learners were automatically matched to relevant patient information, diagnoses and radiologic tests as defined by the expert solutions. Thus, no manual coding was necessary to evaluate the measures evidence elicitation and evidence sharing. The expert solution and scoring procedure are described in more detail below for each measure.

(Transfer) performance of evidence elicitation. As indicator for evidence elicitation, we used the medical relevance of the radiologic tests elicited from the simulated radiologist as defined by the expert solution. This indicator assesses whether learners are able to identify how radiologists would generate the needed evidence, and thus justify their request accordingly in order to convince the radiologist to conduct the test. For each requested test, learners received 1 point if the test was appropriate with respect to the indicated diagnosis and 0 points if learners chose an inappropriate radiologic test. The mean points for all requested radiologic tests were calculated for each patient case scenario. Hence, for each patient case scenario, a maximum of 1 point was possible. To analyze the *performance of evidence elicitation*, we calculated the mean evidence elicitation across all learning cases solved during the intervention. To analyze the *transfer performance of evidence elicitation*, we used the evidence elicitation score of an unsupported posttest case. The internal consistency across all cases, as indicated by Cronbach's alpha was .60.

(Transfer) performance of evidence sharing. We used the relevance of the evidence shared with the radiologist as defined by the expert solution as indicator for evidence sharing. This indicator assesses whether learners are able to identify which information a radiologist would need to optimally conduct the radiologic test and interpret its results. We evaluated the evidence shared during the first request depending on which test was chosen and calculated the proportion of shared relevant evidence to all relevant evidence. Hence, values range between 0 and 1 point with 1 point indicating that all relevant information was shared with the radiologist and 0 points indicating that no relevant information was shared. To analyze the *performance of evidence sharing*, we calculated the mean quality of evidence sharing across all learning cases solved during the intervention. To analyze the *transfer performance of evidence sharing*, we used the evidence sharing score of the unsupported posttest case. The internal consistency across all cases, as indicated by Cronbach's alpha was .76.

2.4.2. Psychological need satisfaction

We assessed psychological need satisfaction directly after the intervention using a scale adapted from Sailer, Hense, Mayr, and Mandl (2017). The scale consisted of three subscales assessing perceived competence, perceived autonomy, and perceived social relatedness. Participants answered all items on a seven-point Likert scale ranging from 1 (I do not agree) to 7 (I totally agree). Perceived competence was measured with four items (Cronbach's alpha = .92). A sample item is "I felt competent during the activity.". The subscale perceived autonomy consisted of three items (Cronbach's alpha = .85). A sample item is "I was able to decide for myself what I would do during the activity.". Social relatedness was measured using three items (Cronbach's alpha = .85). A sample item is "I felt like a part of a team.".

2.4.4. Treatment check

We assessed how many requested radiologic tests were rejected from the simulated radiologist during the intervention to determine whether learners in the adaptive ISS condition did receive any support. For that, we calculated the average absolute number of rejections for the first two test requests and additionally calculated the absolute number of participants whose requests were rejected at least once during the intervention. We included only the first two turns since for solving most patient cases two radiologic examinations are meaningful. The treatment check is successful if learners were rejected at least once per case. The results of the treatment check are reported below.

2.5. Procedure

All participants first answered demographic questions (age, sex, semester, 3 min), and then solved an unsupported pretest case within the agent-based simulation (15 min). During the intervention phase, all participants solved four patient cases (20 min each) in variations of the simulation corresponding to their experimental condition (see above). All patient cases were presented in the same order. Directly after the intervention, we assessed the participants' psychological need satisfaction (5 min). Finally, all participants solved an unsupported posttest case (15 min). All patient cases covered diseases related to fever with unknown origin.

2.6. Statistical analyses

We conducted all statistical analyses with R Studio using the R Version 4.0.2 (R Core Team, 2020) and report inferential results based on a 5% alpha level. We first examined correlations between pretest variables and outcome variables. We found small to moderate correlations between prior evidence elicitation and evidence elicitation performance ($r = 0.26$, $p < .01$) and between prior evidence sharing and evidence sharing performance ($r = 0.36$, $p < .01$) as expected and thus included the respective pretest measures as a covariate.

To analyze whether adaptive and static ISS enhance the performance of evidence elicitation during the intervention and transfer performance of evidence elicitation in an unsupported posttest (Research Question 1a), we conducted two ANCOVAs with the prior evidence elicitation performance as a covariate. To address research question 1b, we conducted two ANCOVAs with evidence sharing performance and transfer performance of evidence sharing as the dependent variable, respectively, and prior evidence sharing performance as a covariate. We addressed the second research question by conducting three ANOVAs with the three basic psychological needs measures perception of competence, autonomy, and social relatedness as dependent variables. Further, examining Q-Q plots and histograms yielded that measures for evidence elicitation, transfer performance of evidence sharing, and autonomy suffered from a non-normal distribution. Therefore, we additionally conducted non-parametric Kruskal-Wallis-Tests for these variables. For all analyses, we tested hypotheses with planned contrasts and analyzed further differences between groups using Tukey corrected post-hoc tests.

3. Results

3.1. Treatment check

As a treatment check, we descriptively analyzed the sum of rejections during the intervention. The control group was rejected most often ($M = 5.528$, $SD = 4.304$), followed by the adaptive ISS condition ($M = 3.759$, $SD = 3.291$) and the static ISS condition ($M = 3.528$, $SD = 3.129$). In the adaptive ISS, there were 5 participants (9.3%) who received no rejections on their requests at all during the intervention. In the static condition, the requests of 6 participants (11.3%) were not rejected by the radiologist. In the control condition, the requests of 4 participants (7.5%) were not rejected. For more detailed number of rejections per group see Table 1. Since these numbers are comparable between the treatment groups, we decided not to exclude these participants to not reduce the power of analyses. These findings show that the majority of participants in the adaptive ISS condition did receive instructional support during the intervention and numbers of participants whose requests were not rejected by the radiologist were comparable between groups. Thus, the treatment check was successful.

3.2. Effects of adaptive and static information sharing scripts on evidence elicitation

Concerning the *evidence elicitation performance*, the descriptive results show that learners supported with the static ISS scored highest, followed by learners in the adaptive ISS condition. Learners in the control group showed the lowest performance of evidence elicitation (see Table 2). The ANCOVA indicates a large effect of the intervention with significant differences between conditions ($F(2,156) = 13.362$, $p <$

Table 1
Absolut number of participants receiving a number of rejections during the intervention.

Number of Rejections	Adaptive ISS	Static ISS	Control condition
0	5	6	4
1	12	11	3
2	6	8	5
3	5	4	5
4	8	10	6
5	6	4	5
6	5	2	8
7	0	0	7
8	3	4	4
9	1	2	1
10	1	0	1
More than 10	2	2	4

Note: ISS = information sharing script.

Table 2
Means and standard deviations for collaborative diagnostic reasoning performance per condition.

	Adaptive ISS	Static ISS	Control condition
	M (SD)	M (SD)	M (SD)
Pretest			
Prior EE perf.	0.728 (0.335)	0.723 (0.312)	0.741 (0.330)
Prior ES perf.	0.644 (0.207)	0.641 (0.239)	0.660 (0.223)
Intervention phase			
EE perf.	0.748 (0.170)	0.858 (0.153)	0.681 (0.232)
ES perf.	0.854 (0.117)	0.815 (0.111)	0.756 (0.143)
Posttest			
EE transfer perf.	0.898 (0.240)	0.896 (0.207)	0.896 (0.241)
ES transfer perf.	0.831 (0.162)	0.767 (0.181)	0.762 (0.188)

Note: ISS = information sharing script, EE = evidence elicitation, ES = evidence sharing. For all variables, the theoretical minimum is 0 and the theoretical maximum is 1.

.001, partial η^2 = 0.146). The robust Kruskal-Wallis-Test also yielded significant differences ($\chi^2(2)$ = 22.431, p < .001). The planned contrasts reveal that the scripted groups significantly outperform unscripted groups, whereby learning with the static ISS led to significantly higher evidence elicitation performance than learning with the adaptive ISS (see Table 3). One-sided Tukey-corrected post-hoc comparisons further yielded, that the static ISS significantly enhanced evidence elicitation performance compared to the control group ($M_{Difference}$ = 0.179, SE = 0.035, p < .001), but the adaptive ISS did not ($M_{Difference}$ = 0.069, SE = 0.035, p = .066).

Concerning the *transfer performance of evidence elicitation*, the descriptive results show similar scores in all three groups with the highest score for learners supported with an adaptive ISS during the intervention (see Table 2). The ANCOVA yielded a null effect and no significant differences between conditions (F(2,156) = 0.003, p = .997, partial η^2 = 0.000). The robust Kruskal-Wallis-Test also yielded no significant differences ($\chi^2(2)$ = 0.109, p = .947).

Overall, the results suggest that, as hypothesized, learning with static ISS enhanced evidence elicitation performance, but in contrast to our expectation, learning with an adaptive ISS did not. Further, the effects did not transfer to an unsupported posttest.

3.3. *Effects of adaptive and static information sharing scripts on evidence sharing*

Concerning the *evidence sharing performance*, learners supported with an adaptive ISS yielded the highest score followed by learners supported with a static ISS. Learners who did not receive any additional support yielded the lowest score for the performance of evidence sharing (see Table 2). The ANCOVA revealed significant and medium-sized effects of the intervention (F(2,156) = 10.633, p < .001, partial η^2 = 0.120). Planned contrasts showed that learning with ISS significantly enhanced evidence sharing performance compared to learning without ISS. Yet, there were no significant differences between both ISS conditions (see Table 3). One-sided Tukey-corrected post-hoc tests further show that the adaptive ISS ($M_{Difference}$ = 0.101, SE = 0.022, p < .001) and the static ISS ($M_{Difference}$ = 0.063, SE = 0.022, p = .007) conditions significantly differ from the control condition.

Concerning the *transfer performance of evidence sharing*, the descriptive results yielded the highest score for learners supported with an adaptive ISS, followed by learners supported with a static ISS. Learners who received no additional support during the intervention showed the lowest score in the skill of evidence sharing (see Table 2). The ANCOVA yielded significant small differences between conditions (F(2,156) = 3.145, p = .046, partial η^2 = 0.039). The robust Kruskall-Wallis-Test yielded no significant differences ($\chi^2(2)$ = 5.674, p = .059). The planned contrasts revealed no significant differences between both ISS conditions and the control condition, but significant differences between the adaptive and the static ISS (see Table 3). One-sided Tukey-corrected post-hoc comparisons further show that only learning with adaptive ISS enhances the transfer performance of evidence sharing in an unsupported posttest compared to unstructured learning ($M_{Difference}$ = 0.074, SE = 0.032, p = .030), but learning with a static ISS did not ($M_{Difference}$ = 0.010, SE = 0.032, p = .696).

Hence, we find support for the hypothesis that adaptive and static ISS enhance evidence sharing performance. Yet, adaptive ISS had no larger effects than static ISS. Further, we find support for the hypothesis that adaptive ISS affects the transfer performance of evidence sharing. Yet, static ISS did not affect the transfer performance of evidence sharing.

3.4. *Effects of adaptive and static information sharing scripts on basic psychological need satisfaction*

Descriptively, learners supported with an adaptive ISS reported the highest *perceived competence*. The lowest feeling of competence was reported by learners supported by the static ISS (see Table 4). The ANOVA yielded a significant small effect of scripting on perceived competence (F(2,157) = 3.568, p = .031, η^2 = 0.043). The planned contrasts reveal that both ISS conditions did not differ significantly from the control condition concerning the perceived competence. Yet, the adaptive ISS differed significantly from the static ISS (see Table 3). One-sided Tukey-corrected post-hoc comparisons further show that neither learners supported with an adaptive ISS perceived significantly higher competence compared to the control group ($M_{Difference}$ = −0.383, SE = 0.250, p = .157), nor did learners supported with the static ISS ($M_{Difference}$ = 0.283, SE = 0.251, p = .998).

Table 4
Means and standard deviations for basic psychological need satisfaction per condition.

	Adaptive ISS	Static ISS	Control condition
	M (SD)	M (SD)	M (SD)
Perceived competence	4.963 (1.241)	4.297 (1.303)	4.580 (1.338)
Perceived autonomy	5.735 (1.100)	5.478 (1.395)	5.597 (1.185)
Perceived social relatedness	3.704 (1.458)	3.233 (1.334)	3.704 (1.660)

Note: ISS = information sharing script. For all variables the theoretical minimum is 1 and the theoretical maximum is 7.

Table 3
Planned contrasts for significant group comparisons.

DV	Contrast	*t*	*p*	*r*
EE performance	Scripted groups vs. control group	4.099	<.001	0.312
	Adaptive vs. static collaboration script	−3.173	.002	0.246
ES performance	Scripted groups vs. control group	4.271	<.001	0.324
	Adaptive vs. static collaboration script	1.715	.088	0.136
ES Transfer performance	Scripted groups vs. control group	1.506	.134	0.120
	Adaptive vs. static collaboration script	1.998	.048	0.158
Perceived competence	Scripted groups vs. control group	0.229	.819	0.018
	Adaptive vs. static collaboration script	2.660	.009	0.208

Note: EE = evidence elicitation, ES = evidence sharing.

Learners reported a relatively high *perceived autonomy*, with learners in the adaptive ISS condition reporting the slightly stronger feelings of autonomy compared to the other conditions. Learners supported with static ISS reported the lowest feeling of autonomy (see Table 4). The ANOVA showed that scripting had a non-significant effect on the perceived autonomy ($F(2,157) = 0.581$, $p = .560$, $\eta^2 = 0.007$). The robust Kruskal-Wallis-Test also yielded no significant differences ($\chi^2(2) = 0.657$, $p = .720$).

Concerning the *perceived social relatedness*, learners supported with adaptive ISS and unsupported learners reported the same level of feeling of social relatedness. Learners supported with static ISS reported the lowest feeling of social relatedness (see Table 4). The ANOVA revealed a non-significant effect of scripting on the perceived social relatedness ($F(2,157) = 1.774$, $p = .173$, $\eta^2 = 0.022$).

Overall, the results are in support of the hypothesis that adaptive scripting enhances the feeling of competence. Yet, the findings do not support the hypotheses that learning with adaptive or static ISS affects the perceived autonomy and perceived social relatedness.

4. Discussion

4.1. Effects of adaptive and static ISS on collaborative diagnostic reasoning

The presented study suggests that collaboration scripts can support specific collaborative diagnostic reasoning processes when learning with simulations. More specifically, the study shows that providing learners with knowledge about the collaboration partners' responsibilities and tasks helped them to successfully collaborate (Fiore et al., 2010) in a professional medical situation between internists and radiologists. In contrast to our hypotheses, static and adaptive collaboration scripts differed in their effectiveness for specific subskills during the intervention. We hypothesized that the adaptive collaboration script would better address the learners' zone of proximal development (Vygotsky, 1978) and, therefore, better help learners to apply the script to the collaborative tasks. Yet, the results support those assumptions only for the subskill evidence sharing, but not for evidence elicitation. The performance of evidence elicitation was only facilitated by the static ISS. These findings challenge the assumption that scaffolds adapted to the learners' needs consistently outperform static support. An explanation why evidence elicitation was only facilitated by the static ISS could lie in the implementation of the ISS. The static ISS was implemented as a letter and a booklet that were constantly present during the intervention phase. In contrast to the adaptive ISS which was only presented to the learners when they showed deviations from an optimal collaboration, the learners could apply the static ISS from the beginning and avoid errors. Besides, learners in the static ISS condition could have focused their attention on the booklet which visually dominated the letter due to its length and contained evidence elicitation support. Thus, implementing static or adaptive scaffolds could impact the learners' focus on the learning material.

An interesting finding is that only the effects of the adaptive ISS on evidence sharing transferred to an unsupported posttest, whereas effects of the static ISS on evidence sharing did not. This means that learners supported with a static collaboration script relied on the availability of the information and thus failed to internalize the meta-knowledge necessary to share patient information. Pea (2004) calls such static scaffolds "distributed intelligence" (p. 431) since they support the learners' momentary activity, but the required skills are not internalized and transferred to similar learning situations. At this point it is important to note that we did not directly measure script internalization. However, internalization of the collaboration script is a plausible explanation for the learners' processes in a posttest in which learners were not supported by a collaboration script.

A further explanation might be that the static collaboration script posed a higher extraneous cognitive load on learners than the adaptive collaboration script since learners were required to search through learning materials to identify relevant support. Thus, the high demands of the complex learning environment combined with the static collaboration script could have exceeded learners' cognitive capacity and impeded learning (Sweller, van Merriënboer, & Paas, 2019). It seems that – when combining complex learning environments such as simulations with other types of instructional support – adaptation of scaffolds is necessary in order to not overwhelm learners. Prior findings support this line of argumentation (e.g., Radkowitsch et al., 2020; Schwaighofer et al., 2017). For example, Schwaighofer et al. (2017) found that when combining different scaffolds, sequencing the scaffolds seems particularly relevant for learners with low working memory capacity. Moreover, compared to prior findings showing moderate to large positive effects of collaboration scripts on learning to collaborate in other learning settings (Radkowitsch et al., 2020), the effects of the collaboration scripts on learning to collaborate within this study (i.e. when learning with simulations) were rather small.

Concerning evidence elicitation, learners in all groups scored high in the transfer test. That indicates that learning of evidence elicitation was easier compared to learning of evidence sharing and that using the agent-based simulations for trial-and-error strategies was successful for internalizing this subskill. Yet, the rather high scores on the pretest and posttest case make a differentiated analyses difficult and further analyses with more differentiated measures are necessary. Nevertheless, for the learning of more complex skills such as evidence sharing, complementing agent-based simulations with adaptive collaboration scripts seems beneficial for internalizing the scripts.

4.2. Effects of adaptive and static ISS on basic psychological need satisfaction

Self-determination theory stresses the importance of the environment for the intrinsic motivation of learners (Deci & Vansteenkiste, 2004). Offering a learning environment enriched with collaboration scripts – which guide learners through collaborative activities and therefore have been criticized for diminishing learners' agency (Wise & Schwarz, 2017) – could thus reduce learners' self-determination and intrinsic motivation. Yet, our findings do not support this line of argumentation since we found no significant effects of adaptive and static ISS on perceived autonomy. Though, it is important to stress that both ISS were designed in a way that they relatively little affected learners' choices when compared to collaboration scripts that, for instance, structure learners' communication with sentence openers. Nevertheless, our findings show that it is generally possible to design static and adaptive collaboration scripts that have little negative impact on perceived autonomy.

Our results suggest that an adaptive design of collaboration scripts can have positive effects on perceived competence since we found that adaptive collaboration scripts significantly increased the perceived competence compared to learning with static ISS. We assumed that just-in-time prompts challenged learners at the right level (Vygotsky, 1978), enabling them to easily adjust their collaborative diagnostic activities, which could have led to an increased feeling of competence (Deci & Ryan, 1985). Thus, our findings suggest that instead of having the proposed detrimental effect on motivation, adaptive collaboration scripts even have the potential to increase perceived competence and, thus, intrinsic motivation.

Further, perceived social relatedness was not significantly affected by scripting. Descriptively, learners in the static ISS condition rated their social relatedness the lowest which is surprising, since, in contrast to the control condition, these learners did receive additional information from the simulated radiologist and, hence, interacted more with the agent than did the control group. Overall, the perceived social relatedness was rather low which could indicate that learners did not immerse in the collaboration as expected. It is, however, unclear, whether this finding is specific to the simulation and the text-based realization of the agent or due to a rather low social relatedness in the simulated situation itself.

4.3. Limitations

The presented study is not without limitations that need consideration when interpreting the results. A first limitation concerns the measure of evidence elicitation that suffered from a rather low internal consistency indicating that learners showed varying degrees of skills between cases. This is a rather common problem in knowledge-based domains such as medical education (Wimmers, Splinter, Hancock, & Schmidt, 2007). The simulation was developed by researchers with many years of experience in medicine, medical education, and educational psychology, and was positively evaluated in expert workshops, a pilot study, as well as a comprehensive validation study. Therefore, we are confident that the simulation has high external validity. Further, most learners scored rather high on the pretest and very high on the posttest, indicating that the respective tests may have been too easy to differentiate well between different levels of competence. The relatively low reliability and variance could have contributed to the non-significant effect of evidence elicitation in the posttest.

Further limitations concern the implementation of the static and adaptive ISS. So far, collaboration scripts were mostly used to scaffold collaborative problem-solving or collaborative co-construction of knowledge (e.g., Rummel & Spada, 2005; Stegmann, Weinberger, & Fischer, 2007). These collaboration scripts differ from the scripts used in this study as they scaffolded rather unidirectional elicitation and sharing processes (i.e., a subset of collaboration skills) which is mainly due to the agent-based realization of the study. Although this limits the comparability of our findings to other collaboration script studies, we are convinced that through careful development of the simulation and the collaboration script, we scaffold important collaboration skills and achieve a high standardization within the learning environment.

Beyond that, it is important to consider that the implementation of the collaboration scripts affects the learners' choices minimally which could explain the lack of influence on the perceived autonomy. Thus, these results may not generalize to more coercive collaboration scripts. Nevertheless, our findings demonstrate that criticizing collaboration scripts per se as coercive (Wise & Schwarz, 2017) does not reflect the range of possibilities for implementing collaboration scripts.

Moreover, it is important to mention that we used a text-based implementation of the computer agent. This means that the collaboration could have been perceived as low immersive since learners did not see or hear their collaboration partner. Yet, an agent-based collaboration (text-based or video-based) is per se more artificial than real human-to-human collaboration since spontaneous reactions typical for human-to-human interactions are very limited in human-to-agent interactions. The ultimate question that arises from the use of a text-based agent is, thus, whether the results generalize to the real professional collaborative situation. This question ultimately requires empirical testing. This is particularly the case since we investigated only some collaborative subskills of collaborative diagnostic reasoning (i.e., sharing and elicitation), but not others (e.g., negotiating). However, because of the careful development and empirical evaluation of the agent-based simulation as well as the collaboration scripts, we consider that the simulation has a certain level of external validity (see Radkowitsch et al., 2020). Beyond that we are convinced that the agent-based simulation offers a high degree of standardization which we consider important for the thorough empirical examination in the context of basic instructional research (Graesser et al., 2018). Particularly for the learning of very specific (sub-)skills, agents can provide meaningful learning tools. Form and text-based interaction is very close to clinical reality for the skills we have identified to require training. This even makes it possible to dispense with too much resource intense face-to-face communication and provide a high degree of standardization. So far, educational and psychological research has neglected the standardization of collaborative situations to large extents.

A final limitation concerns the statistical power of the analyses. Based on prior studies that reported moderate to large effects of collaboration scripts on learning to collaborate, we conducted the a priori power analyses based on a moderate effect. The effects found in the transfer tests were rather small and the posterior power for the effect of collaboration scripts on transfer performance of evidence sharing was 61%. Probably due to the low power, the robust analyses were not significant. In contrast to the ANCOVA, the robust analyses did not include the variation of the pretest variable which is why we rely and interpret findings of the ANCOVA. Yet, future research on the combined effect of different instructional means should assume small effects for the calculation of power and replicate the findings using larger samples.

4.4. Implications and further research

In this study, we advanced collaborative diagnostic reasoning, for which individual diagnostic activities (e.g., generation of evidence or hypotheses) and collaborative activities (e.g., sharing, negotiating) are necessary (e.g., OECD, 2017), by using agent-based simulations and collaboration scripts. We focused on sharing of evidence adjusted to a partner with different knowledge background and elicitation of evidence from such a partner in order to reduce the uncertainty within the own diagnostic reasoning processes. These skills are considered two important subskills of collaborative diagnostic reasoning (e.g., Tschan et al., 2009). Our results suggest that collaboration scripts have positive effects beyond learning with simulations, and although the adaptive collaboration script was not generally better than the static collaboration script, the adaptive script helped learners to internalize the collaboration script. Besides, the adaptive collaboration script had positive effects on the perceived competence as compared to static collaboration scripts. Taking these findings together suggests that although adaptive support requires more effort during development, its effect is relevant. Still, to get a clear picture when and how adaptive collaboration scripts are effective, systematization of research is necessary (Plass & Pawar, 2020; Rummel, Walker, & Aleven, 2016). Future research should systematically vary the mechanism of adaptivity (e.g., adaptive or fading out), the bases for decision (e.g., prior knowledge or performance in the process), and the skill targeted by the scaffold (e.g., elicitation or sharing). Beyond that it seems important to consider the extent to which learners are exposed to the treatment when using adaptive support. We considered this by conducting a treatment check. An alternative approach for future research could be to examine the effect of treatment exposition on learning. Agent-based collaboration could be a promising means to provide the necessary standardization for such analyses.

We used meta-knowledge prompts to explicitly guide learners' evidence sharing and elicitation processes in the context of collaborative diagnostic reasoning. Research and theory on collaboration scripts have so far focused on how engaging in and prompting specific collaborative activities affects learning (Radkowitsch et al., 2020). The role of meta-knowledge for engaging in collaborative processes was discussed instead in group awareness research (Engelmann & Hesse, 2011). Our findings show that collaboration scripts including meta-knowledge prompts indeed affect the learning of collaboration skills. Hence, a theory about collaboration scripts (Fischer et al., 2013) should address the role of knowledge about the collaboration partners explicitly.

Furthermore, we provided further counterevidence against the criticism that collaboration scripts were prone to undermine learners' agency (Wise & Schwarz, 2017). These results are in line with the findings of a meta-analysis (Radkowitsch et al., 2020). Our results provide more detailed insights as we analyzed differentiated effects on basic psychological needs. The results of our study suggest that if the negative effects of collaboration scripts on autonomy existed at all, the effects must be minimal. In contrast, adaptive scripts enhanced the perceived competence of learners. For generalizing our results, future research should focus on replicating these effects in different contexts with different types of collaboration scripts and investigate long-term effects on basic psychological needs, intrinsic motivation, and learning. Given the evidence provided by this and previous studies (Radkowitsch et al., 2020; Stegmann et al., 2011) and the moderate to large positive effects on learning to collaborate (Radkowitsch et al., 2020), we can recommend the use of collaboration scripts for learning to collaborate.

5. Conclusion

We investigated facilitative conditions for collaborative diagnostic reasoning in a standardized agent-based simulation in undergraduate medical education by using adaptive and static collaboration scripts. To date, agents have mainly been used to assess competences (e.g., OECD, 2017) or as pedagogical agents to support specific competences of individuals or groups of learners (e.g., intelligent tutoring systems, Steenbergen-Hu & Cooper, 2014). We showed that using an agent to simulate a collaboration partner is a suitable means to effectively facilitate the learning of collaborative competences without the confounding influence of variables related to group composition (Fransen, Weinberger, & Kirschner, 2013). For complex competences such as collaborative diagnostic reasoning, such agent-based simulation can provide an economical alternative to face-to-face team training and further allow to focus on essential but specific subskills. This seems important since results of this study suggest that knowledge about the collaboration partners and their roles, tasks, and responsibilities substantially affects collaboration. Furthermore, this study shows that combining simulations with adaptive instructional support helps learners to internalize complex skills without negatively affecting learners' basic psychological needs. Yet, adaptive support is no panacea, and systematizing research on adaptive support is necessary to better understand under which conditions adaptive support enhances the learning of collaboration.

We conclude that by complementing agent-based simulations with adaptive collaboration scripts, we identified conditions to effectively help medical students learn important aspects of collaborative diagnostic reasoning.

Author note

Ethical clearance was declared by the ethics committee of LMU Klinikum, Ludwig-Maximilians-Universität, prior to data collection.

Funding information

The research presented in this contribution was funded by a grant of the Deutsche Forschungsgemeinschaft (DFG, FOR 2385) to Frank Fischer, Martin R. Fischer, and Ralf Schmidmaier (Project 6 CoSiMed, FI 792/11-1).

CRediT authorship contribution statement

Anika Radkowitsch: Writing – original draft, Conceptualization, Resources, Methodology, Formal analysis, Project administration. **Michael Sailer:** Conceptualization, Supervision, Writing – review & editing. **Ralf Schmidmaier:** Funding acquisition, Resources, Conceptualization, Writing – review & editing. **Martin R. Fischer:** Funding acquisition, Resources, Conceptualization, Writing – review & editing. **Frank Fischer:** Conceptualization, Funding acquisition, Supervision, Resources, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no conflict of interest.

Acknowledgments

The authors would like to express their gratitude to Felix Behrmann, Dr. Hannah Gerstenkorn, Dr. Victoria Reitmeier, Dr. Alexander Sachs, and Dr. Hanna Zimmermann for contributing their medical expertise during the developement of the experimental material and to Felix Behrmann, Laura Brandl, and Andreas Wildner for their support before and during the collection of data.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.learninstruc.2021.101487.

References

Radkowitsch, A., Fischer, M. R., Schmidmaier, R., & Fischer, F. (2020). Learning to diagnose collaboratively: Validating a simulation for medical students. *GMS Journal of Medical Education, 37*(5). https://doi.org/10.3205/zma001344. Doc51.

Belland, B. R., Walker, A. E., Kim, N. J., & Lefler, M. (2017). Synthesizing results from empirical research on computer-based scaffolding in STEM education: A meta-analysis. *Review of Educational Research, 87*(2), 309–344. https://doi.org/10.3102/0034654316670999

Brady, A., Laoide, R.Ó., McCarthy, P., & McDermott, R. (2012). Discrepancy and error in radiology: Concepts, causes and consequences. *Ulster Medical Journal, 81*(1), 3. PMID: 23536732; PMCID: PMC3609674 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3609674/.

Cannon-Bowers, J. A., Salas, E., & Converse, S. (1993). Shared mental models in expert team decision making. In N. J. Catellan (Ed.), *Individual and group decision making: Current issues* (pp. 221–246). Hillsdale, NJ: Erlbaum.

Charlin, B., Boshuizen, H. P. A., Custers, E. J., & Feltovich, P. J. (2007). *Scripts and clinical reasoning. Medical Education* (Vol. 41). https://doi.org/10.1111/j.1365-2923.2007.02924.x

Charlin, B., Lubarsky, S., Millette, B., Crevier, F., Audétat, M. C., Charbonneau, A., et al. (2012). Clinical reasoning processes: Unravelling complexity through graphical representation. *Medical Education, 46*(5), 454–463. https://doi.org/10.1111/j.1365-2923.2012.04242.x

Chernikova, O., Heitzmann, N., Fink, M. C., Timothy, V., Seidel, T., & Fischer, F. (2020). Facilitating diagnostic competences in higher education - a meta-analysis in medical and teacher education. *Educational Psychology Review, 32*, 157–196. https://doi.org/10.1007/s10648-019-09492-2

Constantino-Gonzalez, M. A., Suthers, D. D., & de los Santos, J. G. E. (2003). Coaching web-based collaborative learning based on problem solution differences and participation. *International Journal of Artificial Intelligence in Education, 13*(2–4), 263–299.

Cook, D. A. (2014). How much evidence does it take? A cumulative meta-analysis of outcomes of simulation-based education. *Medical Education, 48*(8), 750–760. https://doi.org/10.1111/medu.12473

Davies, S., George, A., Macallister, A., Barton, H., Youssef, A., Boyle, L., et al. (2018). "It's all in the history": A service evaluation of the quality of radiological requests in acute imaging. *Radiography, 24*(3), 252–256. https://doi.org/10.1016/j.radi.2018.03.005

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York, NY: Plenum.

Deci, E. L., & Vansteenkiste, M. (2004). Self-determination theory and basic need satisfaction: Understanding human development in positive psychology. *Ricerche di Psicologia, 27*(1), 23–40.

Dillenbourg, P. (2002). Over-scripting CSCL: The risks of blending collaborative learning with instructional design. In P. A. Kirschner (Ed.), *Three worlds of CSCL. Can we support CSCL?* (pp. 61–91). Netherlands: Open Universiteit Nederland. Heerlen.

Diziol, D., Walker, E., Rummel, N., & Koedinger, K. R. (2010). Using intelligent tutor technology to implement adaptive support for student collaboration. *Educational Psychology Review, 22*(1), 89–102. https://doi.org/10.1007/s10648-009-9116-9

Engelmann, T., & Hesse, F. W. (2011). Fostering sharing of unshared knowledge by having access to the collaborators' meta-knowledge structures. *Computers in Human Behavior, 27*, 2078–2087. https://doi.org/10.1016/j.chb.2011.06.002

Feltovich, P. J., & Barrows, H. S. (1984). Issues of generality in medical problem solving. *Tutorials in problem-based learning: A new direction in teaching the health professions* (pp. 128–142). Assen, NL: Van Gorcum.

Fiore, S. M., Rosen, M. A., Smith-Jentsch, K. A., Salas, E., Letsky, M., & Warner, N. (2010). Toward an understanding of macrocognition in teams: Predicting processes in complex collaborative contexts. *Human Factors, 52*, 203–224. https://doi.org/10.1177/0018720810369807

Fischer, F., Kollar, I., Stegmann, K., & Wecker, C. (2013). Toward a script theory of guidance in computer-supported collaborative learning. *Educational Psychologist, 48*, 56–66. https://doi.org/10.1080/00461520.2012.748005

Fransen, J., Weinberger, A., & Kirschner, P. A. (2013). Team effectiveness and team development in CSCL. *Educational Psychologist, 48*(1), 9–24. https://doi.org/10.1080/00461520.2012.747947

Gegenfurtner, A., Quesada-Pallarès, C., & Knogler, M. (2014). Digital simulation-based training: A meta-analysis. *British Journal of Educational Technology, 45*(6), 1097–1114. https://doi.org/10.1111/bjet.12188

Graesser, A. C., Fiore, S. M., Greiff, S., Andrews-Todd, J., Foltz, P. W., & Hesse, F. (2018). Advancing the science of collaborative problem solving. *Psychological Science in the Public Interest, 19*(2), 59–92. https://doi.org/10.1177/1529100618808244

Heitzmann, N., Seidel, T., Opitz, A., Hetmanek, A., Wecker, C., Fischer, M., et al. (2019). Facilitating diagnostic competences in simulations: A conceptual framework and a research agenda for medical and teacher education. *Frontline Learning Research, 7*(4), 1–24. https://doi.org/10.14786/flr.v7i4.384

Hesse, F., Care, E., Buder, J., Sassenberg, K., & Griffin, P. (2015). A framework for teachable collaborative problem solving skills. In P. Griffin, & E. Care (Eds.), *Assessment and teaching of 21st century skills: Methods and approach* (pp. 37–56). Springer.

Hinsz, V. B., Tindale, R. S., & Vollrath, D. A. (1997). The emerging conceptualization of groups as information processors. *Psychological Bulletin, 121*(1), 43–64. https://doi.org/10.1037/0033-2909.121.1.43

Järvelä, S., & Hadwin, A. F. (2013). New frontiers: Regulating learning in CSCL. *Educational Psychologist, 48*(1), 25–39. https://doi.org/10.1080/00461520.2012.748006

Karakostas, A., & Demetriadis, S. (2011). Enhancing collaborative learning through dynamic forms of support: The impact of an adaptive domain-specific support strategy. *Journal of Computer Assisted Learning, 27*(3), 243–258. https://doi.org/10.1111/j.1365-2729.2010.00388.x

Kiesewetter, J., Fischer, F., & Fischer, M. R. (2017). Collaborative clinical reasoning – a systematic review of empirical studies. *Journal of Continuing Education in the Health Professions, 37*(2), 123–128. https://doi.org/10.1097/ceh.0000000000000158

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. https://doi.org/10.1207/s15326985ep4102_1

Kirschner, P. A., Sweller, J., Kirschner, F., & Zambrano, J. (2018). From cognitive load theory to collaborative cognitive load theoy. *International Journal of Computer-Supported Collaborative Learning, 13*, 213–233. https://doi.org/10.1007/s11412-018-9277-y

Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science, 12*, 1–48. https://doi.org/10.1207/s15516709cog1201_1

Klein, M., Otto, B., Fischer, M. R., & Stark, R. (2019). Fostering medical students' clinical reasoning by learning from errors in clinical case vignettes: Effects and conditions of additional prompting procedures to foster self-explanations. *Advances in Health Sciences Education: Theory and Practice, 24*(2), 331–351. https://doi.org/10.1007/s10459-018-09870-5

Kolodner, J. L. (1992). An introduction to case-based reasoning. *Artificial Intelligence Review, 6*(1), 3–34.

Larson, J. R., Christensen, C., Franz, T. M., & Abbot, A. S. (1998). Diagnosing groups: The pooling, management, and impact of shared and unshared case information in team-based medical decision making. *Journal of Personality and Social Psychology, 75*(1), 93–108. https://doi.org/10.1037/0022-3514.75.1.93

Liu, L., Hao, J., von Davier, A. A., Kyllonen, P., & Zapata-Rivera, D. (2015). A tough nut to crack: Measuring collaborative problem solving. In Y. Rosen, S. Ferrara, & M. Mosharraf (Eds.), *Handbook of research on computational tools for real-world skill development* (pp. 344–359). Hershey, PA: IGI Global.

Meier, A., Spada, H., & Rummel, N. (2007). A rating scheme for assessing the quality of computer-supported collaboration processes. *International Journal of Computer-Supported Collaborative Learning, 2*(1), 63–86.

Nickerson, R.-S. (1988). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2*(2), 175–220. https://doi.org/10.1037/1089-2680.2.2.175

Noroozi, O., Biemans, H. J. A., Weinberger, A., Mulder, M., & Chizari, M. (2013). Scripting for construction of a transactive memory system in multidisciplinary CSCL environments. *Learning and Instruction, 25*, 1–12. https://doi.org/10.1016/j.learninstruc.2012.10.002

Nye, B. D., Graesser, A. C., & Hu, X. (2014). AutoTutor and family: A review of 17 years of natural language tutoring. *International Journal of Artificial Intelligence in Education, 24*(4), 427–469. https://doi.org/10.1007/s40593-014-0029-5

OECD. (2017). *PISA 2015 Assessment and analytical framework: Science, reading, mathematic, financial literacy and collaborative problem solving, revised edition*. Paris: France. OECD Publishing.

Patel, V. L., Kaufman, D. R., & Arocha, J. F. (2002). Emerging paradigms of cognition in medical decision-making. *Journal of Biomedical Informatics, 35*, 52–75.

Pea, R. D. (2004). The social and technological dimensions of scaffolding and related theoretical concepts for learning, education, and human activity. *The Journal of the Learning Sciences, 13*(3), 423–451. https://doi.org/10.1207/s15327809jls1303_6

Plass, J. L., & Pawar, S. (2020). Toward a taxonomy of adaptivity for learning. *Journal of Research on Technology in Education, 52*(3), 275–300. https://doi.org/10.1080/15391523.2020.1719943

R Core Team. (2020). R: A language and environment for statistical computing. In *R foundation for statistical computing*Vienna: Austria. https://www.R-project.org/.

Radkowitsch, A., Vogel, F., & Fischer, F. (2020). Good for learning, bad for motivation? A meta-analysis on the effects of computer-supported collaboration scripts. *International Journal of Computer-Supported Collaborative Learning, 15*(1), 5–47. https://doi.org/10.1007/s11412-020-09316-4

Roschelle, J., & Teasley, S. D. (1995). The construction of shared knowledge in collaborative problem solving. In C. Oḿalley (Ed.), *Computer-supported collaborative learning* (pp. 69–97). Berlin, Germany: Springer.

Rosen, Y. (2015). Computer-based assessment of collaborative problem solving: Exploring the feasibility of human-to-agent approach. *International Journal of Artificial Intelligence in Education, 25*(3), 380–406. https://doi.org/10.1007/s40593-015-0042-3

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*, 201–241. https://doi.org/10.1207/s15327809jls1402_2

Rummel, N., Walker, E., & Aleven, V. (2016). Different futures of adaptive collaborative learning support. *International Journal of Artificial Intelligence in Education, 26*(2), 784–795. https://doi.org/10.1007/s40593-016-0102-3

Sailer, M., Hense, J. U., Mayr, S. K., & Mandl, H. (2017). How gamification motivates: An experimental study of the effects of specific game design elements on psychological need satisfaction. *Computers in Human Behavior, 69*, 371–380.

Schwaighofer, M., Vogel, F., Kollar, I., Ufer, S., Strohmaier, A., Terwedow, I., et al. (2017). How to combine collaboration scripts and heuristic worked examples to foster mathematical argumentation – when working memory matters. *International Journal of Computer-Supported Collaborative Learning, 12*(3), 281–305. https://doi.org/10.1007/s11412-017-9260-z

Siebeck, M., Schwald, B., Frey, C., Röding, S., Stegmann, K., & Fischer, F. (2011). Teaching the rectal examination with simulations: Effects on knowledge acquisition and inhibition. *Medical Education, 45*, 1025–1031. https://doi.org/10.1111/j.1365-2923.2011.04005.x

Simmons, B. (2010). Clinical reasoning: Concept analysis. *Journal of Advanced Nursing, 66*(5), 1151–1158. https://doi.org/10.1111/j.1365-2648.2010.05262.x

Stadler, M., Herborn, K., Mustafić, M., & Greiff, S. (2019). Computer-based collaborative problem solving in PISA 2015 and the role of personality. *Journal of Intelligence, 7*(3), 15. https://doi.org/10.3390/jintelligence7030015

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. https://doi.org/10.1016/j.learninstruc.2009.10.001

Steenbergen-Hu, S., & Cooper, H. (2014). A meta-analysis of the effectiveness of intelligent tutoring systems on college students' academic learning. *Journal of Educational Psychology, 106*(2), 331. https://doi.org/10.1037/a0034752

Stegmann, K., Mu, J., Gehlen-Baum, V., & Fischer, F. (2011). The myth of over-scripting: Can novices be supported too much?. In H. Spada, G. Stahl, N. Miyake, & N. Law (Eds.), *Connecting computer-supported collaborative learning to policy and practice: CSCL2011 conference proceedings. Volume I. Hong-kong, China: International society of the learning sciences*.

Stegmann, K., Weinberger, A., & Fischer, F. (2007). Facilitating argumentative knowledge construction with computer-supported collaboration scripts. *International Journal of Computer-Supported Collaborative Learning, 2*, 421–447. https://doi.org/10.1007/s11412-007-9028-y

Sun, C., Shute, V. J., Stewart, A., Yonehiro, J., Duran, N., & D'Mello, S. (2020). Towards a generalized competency model of collaborative problem solving. *Computers & Education, 143*. https://doi.org/10.1016/j.compedu.2019.103672

Sweller, J., van Merriënboer, J. J., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review, 31*, 261–292. https://doi.org/10.1007/s10648-019-09465-5

Tchounikine, P., Rummel, N., & McLaren, B. M. (2010). Computer supported collaborative learning and intelligent tutoring systems. In R. Nkambou, J. Bourdeau, & R. Mizoguchi (Eds.), *Advances in intelligent tutoring systems* (pp. 447–463). Berlin, Heidelberg: Springer. Berlin Heidelberg.

Tschan, F., Semmer, N. K., Gurtner, A., Bizzari, L., Spychiger, M., Breuer, M., et al. (2009). Explicit reasoning, confirmation bias, and illusory transactive memory: A simulation study of group medical decision making. *Small Group Research, 40*(3), 271–300. https://doi.org/10.1177/1046496409332928

Tsovaltzi, D., Rummel, N., McLaren, B. M., Pinkwart, N., Scheuer, O., Harrer, A., et al. (2010). Extending a virtual chemistry laboratory with a collaboration script to promote conceptual learning. *International Journal of Technology Enhanced Learning, 2*(1–2), 91–110. https://doi.org/10.1504/IJTEL

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Wecker, C., & Fischer, F. (2011). From guided to self-regulated performance of domain-general skills: The role of peer monitoring during the fading of instructional scripts. *Learning and Instruction, 21*(6), 746–756. https://doi.org/10.1016/j.learninstruc.2011.05.001

Wegner, D. M. (1987). Transactive memory: A contemporary analysis of the group mind. In B. Mullen, & G. R. Goethals (Eds.), *Theories of group behavior* (pp. 185–208). New York, NY: Springer.

Wimmers, P. F., Splinter, T. A., Hancock, G. R., & Schmidt, H. G. (2007). Clinical competence: General ability or case-specific? *Advances in Health Sciences Education, 12*(3), 299–314. https://doi.org/10.1007/s10459-006-9002-x

Wise, A. F., & Schwarz, B. B. (2017). Visions of CSCL: Eight provocations for the future of the field. *International Journal of Computer-Supported Collaborative Learning, 12*(4), 423–467. https://doi.org/10.1007/s11412-017-9267-5

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry, 17*(2), 89–100. https://doi.org/10.1111/j.1469-7610.1976.tb00381.x

5 General Discussion

Anika Radkowitsch

This thesis pursued two main goals: (1) to advance our understanding of collaborative diagnostic reasoning and (2) to identify conditions under which facilitating collaborative diagnostic reasoning is effective when learning with agent-based simulations and collaboration scripts. These goals were addressed as follows: first, in collaboration with medical educators, physicians, psychologists, and software engineers, I developed an agent-based simulation. Second, I delineated the CDR model to describe collaborative diagnostic reasoning processes. Third, I conducted three studies with the goal to identify effective learning conditions. While the first study focused on modelling collaborative diagnostic reasoning in agent-based simulations, the second and third study focused on its facilitation with collaboration scripts. By doing so, the first two studies provide the general foundations for the third study, which more directly addressed the presented goals. Below, I will shortly summarize and interpret the findings of the three studies and discuss them in the light of theories and their practical relevance. I then conclude the thesis with a discussion of the limitations of the presented studies and some suggestions for further research.

5.1 Summary of the Results

In *Study 1* (Radkowitsch, Fischer, Schmidmaier, & Fischer, accepted), I constructed a validity argument for the use of the simulation as assessment instrument in the context of research in educational psychology following suggestions by Cook and Hatala (2016). To this end, the performance of medical students with low and high prior knowledge was compared to that of internists with at least three years of working experience. The comparison referred to the participants' information sharing skills, their diagnostic accuracy, diagnostic efficiency, and their intrinsic cognitive load when working with the agent-based simulation. Further, internists rated the perceived authenticity of the simulation and the collaboration. The study showed that our proposed measures differentiate between different levels of prior knowledge. The largest differences on all measures were observed between low prior knowledge students and the other two groups. Differences between all three groups were largest for the measures of diagnostic efficiency and intrinsic cognitive load. This is in line with theories about the development of diagnostic reasoning skills that assume that low prior knowledge students are also able to diagnose accurately, yet their diagnostic reasoning processes are more effortful and slower. In contrast, diagnostic reasoning processes of more experienced diagnosticians are faster and less effortful (Charlin et al., 2007). Thus, medical expertise becomes evident in the speed of the diagnostic process and – at least for routine cases such as those presented in our simulation – in how effortful this process is, i.e. how much cognitive load is produced (Schmidt & Rikers, 2007). Therefore, the evaluation with the simulation produced exactly those differences between prior knowledge groups that one could expect to find in real professional requirement situations. These findings concur with the result that internists rated the simulation and the collaborative process mostly as authentic or rather authentic. At the same time, the validation study allowed for the identification of some weaknesses of the simulation. One such weakness concerned the operationalization of collaborative diagnostic competences. The simulation focuses on information sharing skills during collaborative evidence generation. Information sharing skills were operationalized as the ability to sufficiently justify a radiologic test request, by that using an overall measure for collaborative diagnostic quality. However, the validation study finds low internal consistency for this measure indicating that across patient cases, the participants in the study showed varying information sharing skills. A plausible explanation is that the measure was not one-dimensional, meaning that the information sharing skill as operationalized in this study could have covered more than one subskill. This could, in principle, reduce consistency (N. Schmitt, 1996). As a consequence, the measure for information sharing skills in the third study got split into two related subskills, namely sharing and elicitation. By that, I was able to get a more nuanced picture of collaborative diagnostic reasoning. In light of the results from Study 3, this decision proved useful since internal consistency values of the new measures were clearly higher. Overall, however, study showed that the simulation we developed in principle can be considered as valid and allowed us to spot weaknesses, which we addressed before using the simulation for experimental studies.

In *Study 2* (Radkowitsch, Vogel, & Fischer, 2020), I investigated the effect of collaboration scripts on learning to collaborate and motivation by synthesizing prior research by means of a meta-analysis. The analyses were based on a sample of 56 studies that were identified during two comprehensive literature searches. Included were all published and peer-reviewed studies that tested effects of a collaboration script against unstructured learning and provided data on a posttest. The study allowed for a number of interesting conclusions: Firstly, it showed moderate positive effects of collaboration scripts on learning to collaborate. These results confirm findings of a prior meta-analysis (Vogel et al., 2017) and could be seen as an additional indicator that collaboration scripts could be useful tools to also enhance the learning of collaborative diagnostic reasoning. When only analyzing the effect on information sharing skills (including elicitation and sharing), the effects were moderate to large as well (Radkowitsch et al., 2020).Yet, there was large inconsistency between studies and the study failed to identify moderators that explained a significant amount of between study variance. Descriptively, collaboration scripts that facilitated a single collaborative activity were found to be more effective when compared to collaboration scripts that prompted two or more. This indicates that differentiated scaffolding as suggested by Tabak (2004) is difficult to achieve and that combining different collaborative activities hinders practice of a single collaboration skill.

This study further addressed the question of whether collaboration scripts negatively affect learners' motivation. Only very few studies in the sample provided data on this question. The summary effect was non-significant and close to zero. This is a first indicator that, despite the widespread criticism (Wise & Schwarz, 2017), collaboration scripts do not seem to have a negative effect on learners' motivation. Despite consistently positive effects of collaboration scripts on learning, the criticism is persistent and therefore needs thorough empirical examination. Thus, I took up this lack of empirical data and directly addressed the criticism (Dillenbourg, 2002; Wise & Schwarz, 2017) in Study 3.

Finally, *Study 3* (Radkowitsch, Sailer, Schmidmaier, Fischer, & Fischer, submitted) combined the approaches of both prior studies by investigating the effect of two types of collaboration scripts – an adaptive and a static collaboration script – on learning of collaborative diagnostic reasoning and on basic psychological needs (Deci & Ryan, 1985) when learning with agent-based simulations. For that, 160 medical students learnt either with an adaptive, a static, or without any collaboration script. All medical students first solved a pretest patient case, then solved four patient cases according to their experimental condition, then rated their perceived basic psychological needs and extraneous cognitive load, and finally solved a patient case that served as posttest. The results of this study show that the static collaboration script had positive effects on evidence elicitation performance and evidence sharing performance. The adaptive collaboration script increased evidence sharing performance. Examining the transfer performance of evidence elicitation in the posttest showed that irrespective of the conditions most students scored high in the posttest. This suggests that for learning evidence elicitation trial and error strategies were sufficient and no extra support in the form of collaboration scripts was needed. Concerning the transfer performance of evidence sharing, only learners who were supported by the adaptive collaboration script outperformed the control group. This indicates that only the adaptive collaboration scripts were able to support internalization. With respect to the effect of static and adaptive collaboration scripts on satisfaction of basic psychological needs, Study 3 yielded no significant effects on perceived autonomy and perceived social relatedness. However, learners supported with the adaptive collaboration script showed a higher perceived competence and a lower extraneous cognitive load than learners supported with the static collaboration script. For both collaboration scripts, the differences to the control group were not significant. These findings align well with the findings from Study 2, which also showed no significant effect of collaboration scripts on motivation when compared to unstructured collaboration. Overall, these findings suggest that instead of affecting perceived autonomy, collaboration scripts seem to affect perceived competence and cognitive load (cf. Wise & Schwarz, 2017). For learners to benefit from these effects, adaptive collaboration scripts should be used. In all, these findings clearly indicate that supporting learners with adaptive collaboration scripts is beneficial for advancing collaborative diagnostic reasoning.

5.2 Theoretical Implications for the Understanding of Collaborative Diagnostic Reasoning

One goal of this thesis was to deepen the theoretical underpinning of collaborative diagnostic reasoning. A first essential step towards this goal was to develop the CDR model (see Section 1.2.3) based on prior literature (see Radkowitsch, Sailer, et al., accepted). The CDR model served as theoretical underpinning for researching the facilitation of collaborative diagnostic reasoning and allowed some conclusions on collaborative diagnostic reasoning processes. A first conclusion concerns the collaborative diagnostic activities proposed in the model. As described above, the first general measure used for the information sharing quality suffered from low internal consistency, probably due to multidimensionality (N. Schmitt, 1996). Thus, the measure was split into a sharing and elicitation subskill. The analyses of study 3 show that sharing and elicitation of evidence indeed are two distinct subskills that correlate only poorly with each other. These findings indicate that – as proposed by the CDR model – sharing and elicitation of evidence are two distinct collaborative diagnostic activities with two distinct underlying collaboration skills (see Fischer et al., 2002).

A second conclusion concerns the assumption that meta-knowledge, such as knowledge about the collaboration partners' task, roles, and responsibilities, influences the quality of collaborative diagnostic activities (Engelmann & Hesse, 2011; Mathieu et al., 2000). In Study 3, I tested this assumption indirectly by supporting the learners' meta-knowledge with collaboration scripts. Providing learners with meta-knowledge increased their elicitation and sharing performance and, in parts increased their performance in a transfer test. These findings suggest that, at least in the investigated setting and for the investigated collaborative diagnostic activities, knowledge about the collaboration partners' task, roles, and responsibilities indeed affects how diagnosticians engage in collaborative diagnostic activities. To what extent these relations exist between meta-knowledge and other collaborative diagnostic activities, for instance negotiation of hypotheses, remains an open question that should be addressed in future research.

The finding that meta-knowledge affects collaborative diagnostic activities further relates to the question to what extent collaboration during collaborative diagnostic reasoning is based on domain-specific or domain-general skills. In the CDR model, I assume that both domain-general social skills and domain-specific meta-knowledge influence collaborative diagnostic activities. However, so far there is little evidence on domain-generality or domain-specificity of collaboration skills. Earlier models on collaborative problem solving (e.g., Hesse et al., 2015; Liu et al., 2015) do not describe domain-specific components. First evidence for the domain-specificity of collaboration skills comes from Kiesewetter, Fischer, and Fischer (2016). They showed collaboration experts and novices pictures of typical collaborative situations of their own and of another domain. Collaboration experts remembered more activity-related information from the pictures in their own domain, but not in another domain than did novices. These findings suggest that collaboration skills are – at least to some extent – domain-specific skills and further align with findings from Study 3 showing that domain-specific meta-knowledge affects collaborative diagnostic activities. In the context of scientific reasoning and argumentation, Hetmanek et al. (2018) assumed that the influence of domain-general and domain-specific knowledge and skills shift with increasing expertise. That means that domain-general skills supplement a lack of domain-specific skills, but with increasing domain-specific skills, the influence of domain-general skills decreases. Albeit further empirical evidence on this assumption is needed, it seems plausible for collaborative diagnostic reasoning as well and should be tested in the future.

Overall, the data collected in this thesis directly or indirectly supports three assumptions proposed in the CDR model. Firstly, there are at least two distinct collaborative activities that are important in at least some collaborative diagnostic situations, namely elicitation and sharing. Secondly, these collaborative diagnostic activities are indeed affected by knowledge about the collaboration partners' task, roles, and responsibilities. These findings, finally, imply that collaboration during collaborative diagnostic reasoning depends at least to some extent on domain-specific skills. These assumptions were tested in the context of collaborative evidence generation between internists and radiologists in medicine. It seems plausible that assumed relations generalize to other contexts in which two or more diagnosticians collaboratively generate evidence. Yet, so far, generalizability to other contexts has not been tested empirically. Thus, testing whether these findings transfer to other collaborative diagnostic situations within medicine or outside medicine represents an attractive avenue for future research.

5.3 Theoretical Implications for Facilitating Collaborative Diagnostic Reasoning with Collaboration Scripts in Agent-Based Simulations

A further goal of this thesis was to identify conditions under which collaborative diagnostic reasoning can be facilitated effectively when learning with agent-based simulations. A first step towards this goal was to assess whether collaboration scripts in general and within agent-based simulations are an effective instructional support. To address this question I first conducted a meta-analysis on the effectiveness of collaboration scripts in computer-supported learning environments (Study 2). Although the included studies varied considerable in their effectiveness, the synthesized effect shows that collaboration scripts do advance the learning of collaboration skills (Radkowitsch et al., 2020). These findings are in line with prior synthesized findings (Vogel et al., 2017) and provide further support for the script theory of guidance, which suggests that collaboration scripts induce beneficial collaborative processes that support adjusting or developing functional internal collaboration scripts (Fischer et al., 2013). Findings from Study 3 allow – at least to some extent - for a generalization of these findings to learning with agent-based simulations. In agent-based simulations, collaboration scripts were effective for advancing performance of collaborative diagnostic reasoning when being scaffolded. However, for the easier collaborative diagnostic activity of evidence elicitation, using the agent-based simulation for trial and error strategies might have been sufficient for developing or adjusting the internal collaboration scripts. This rendered additional external collaboration scripts unnecessary. For the more complex skill of evidence sharing such trial and error strategies probably were insufficient. In this case, further instructional support in form of collaboration scripts was necessary.

Study 2 and 3 further identified conditions under which collaboration scripts are particularly effective. In Study 3, only adaptive collaboration scripts supported internalization of the collaboration script, whereas in Study 2 also non-adaptive collaboration scripts were effective. It seems that the static collaboration script in combination with the agent-based simulation posed an excessive cognitive load on learners and, consequently, hindered the script internalization. Similar results were presented in a study by Schwaighofer et al. (2017) who found that fading of instructional support is important when combining collaboration scripts with worked examples, particularly for learners' with low working memory capacity. Thus, it seems that combining instructional support or scaffolds also binds cognitive capacity. This makes fading or adaptive approaches necessary for learning. This assumption aligns well with findings from Study 2 indicating that collaboration scripts that address several collaboration skills have smaller effects than those focusing on a single collaboration skill (Radkowitsch et al., 2020). Besides, when learning with static collaboration scripts, learners could have relied on the presence of the static collaboration script using it as “distributed intelligence” (Pea, 2004, p. 431) and, thus, have failed to internalize the collaboration script. Therefore, collaboration scripts can be cognitively demanding for what reason adaptivity seems beneficial when combining agent-based simulations with collaboration scripts for supporting complex skills. By that, agent-based simulations and collaboration scripts can synergize their effects and provide instructional support beyond the effects of agent-based simulations alone (Tabak, 2004). Nevertheless, adaptive collaboration scripts were not consistently more beneficial than static collaboration scripts and there are several different ways to adapt instructional support of which only one type was investigated. This suggests that further research should systematically investigate the effectiveness of different types of adaptivity (Rummel, Walker, & Aleven, 2016) by varying the mechanisms of adaptivity (e.g., fading or adaptivity), the targeted skill (e.g., elicitation or sharing), and the basis for decision (e.g., performance or prior knowledge). Agent-based simulations could provide the standardization necessary for this endeavor.

In Study 2 and 3, I assessed whether and under which conditions collaboration scripts might negatively affect learners’ motivation and, thereby, could have negative effects on learning in the long run. According to this criticism, collaboration scripts are prone to disturb natural interactions, undermining learners’ agency and, therefore, reducing learners’ motivation (Dillenbourg, 2002; Wise & Schwarz, 2017). The findings of Study 2 and 3 suggest that there is little evidence in support of such criticism. Study 2 reveals that to date few studies examined the effect of collaboration scripts on motivation. Combined, these studies yielded a non-significant small *positive* effect on motivation. To address a possible compensating effect with collaboration scripts reducing perceived autonomy on the one hand (Wise & Schwarz, 2017), but increasing perceived competence on the other hand (Weinberger, Kollar, Dimitriadis, Mäkitalo-Siegl, & Fischer, 2009), I examined the effect of collaboration scripts on basic psychological needs (Deci & Ryan, 1985) in Study 3. Generally, collaboration scripts did not significantly affect satisfaction of basic psychological needs when compared to unstructured learning. Yet, learners supported with adaptive collaboration scripts reported significantly higher perceived competence as compared to learners supported with static collaboration scripts. The collaboration scripts did not significantly differ with respect to perceived autonomy or perceived social relatedness. These findings do neither support the compensation hypothesis (Radkowitsch et al., 2020) nor the assumption of collaboration scripts negatively affecting learners' autonomy (Wise & Schwarz, 2017). It seems, however, that adaptive collaboration scripts challenged learners' at their zone of proximal development (Vygotsky, 1978) and, thereby, enabled them to engage in high quality collaborative diagnostic activities and affected positively the learners' perceived competence. By contrast, for learners supported with the static collaboration script identifying relevant collaborative diagnostic activities might have been more difficult as indicated by the increased extraneous cognitive load that could have reduced their perceived competence. An explanation for why collaboration scripts did not affect perceived autonomy is that their level of coerciveness was low. It is, thus, not clear, whether high coercive collaboration scripts also only affect perceived competence but not the feeling of autonomy. A closer empirically examination of the effects of collaboration scripts on basic psychological needs by varying the degree of coercion of collaboration scripts, therefore, seems to be a particularly fruitful avenue for future research. Overall, the findings of the present studies suggest that scripting per se does not necessarily come with a negative effect on intrinsic motivation, but that adaptive collaboration scripts seem to be more beneficial for intrinsic motivation than static collaboration scripts and are, therefore, expected to be more beneficial for learning in the long-term.

A further implication of this thesis concerns the use of meta-knowledge for scaffolding collaborative diagnostic reasoning. To date, knowledge about the collaboration partners' tasks, roles, and responsibilities has mainly been investigated in the context of group awareness research (e.g., Engelmann & Hesse, 2011; Schnaubert & Bodemer, 2019) and was largely ignored by the scripting community. The collaboration script used in Study 3 showed, however, that meta-knowledge is essential for successful collaborative processes and that collaboration scripts are a useful tool to provide this knowledge. The script theory of guidance (Fischer et al., 2013) acknowledges this relevance by proposing a role component and suggesting the "internal script configuration principle" (Fischer et al., 2013, p. 58), which states that a learners' goals and the perceived situational characteristics configure the internal collaboration scripts. Yet, this theory could be further improved by making the importance of knowledge about the collaboration partners' roles and knowledge an explicit component. The present findings also emphasize that the activation of an internal collaboration script depends on the collaboration partners. Thus, scaffolds for collaboration should be adapted to the collaboration partner. In this light, standardizing collaboration partners, for instance by using agent-based simulations, seems vital when investigating the effectiveness of collaboration scripts.

Overall, the findings presented in this thesis imply that for learning complex subskills of collaborative diagnostic reasoning, the use of collaboration scripts containing meta-knowledge is beneficial. While the effects of adaptivity during the learning process were not as clear as expected, the findings do suggest that adaptive collaboration scripts better supported the internalization of collaborative diagnostic reasoning. Further, adaptivity had positive effects on perceived competence but no significant effects on perceived autonomy, which becomes more important when considering long-term effects of learning with collaboration scripts. It became also evident from Studies 2 and 3 that collaboration scripts can also affect cognitive load when combined with other instructional support, which can be reduced by providing focused or adapted support. Thus, considering both affective and cognitive outcomes suggests that for scaffolding collaborative diagnostic reasoning when learning with agent-based simulations, collaboration scripts are suitable with an advantage of adaptive over static collaboration scripts. Future research should, however, investigate, whether positive effects on perceived competence and learning are preserved when using more coercive collaboration scripts. Agent-based simulation could offer the necessary standardization to investigate such effects.

5.4 Practical Implications

The insights generated by this thesis have some relevant implications for medical education. Primarily, this thesis highlighted the relevance of collaboration for diagnosing in many situations of the physicians' routines. Although medical educators have recognized that the physicians' role as a team member is essential (e.g., MFT Medizinischer Fakultätentag der Bundesrepublik Deutschland e. V., 2015), medical education lacks a systematic facilitation of collaborative diagnostic reasoning. When collaboration is the target of a training, collaboration skills are often conceptualized as domain-general skills (i.e., so called non-technical skills, Flin & Maran, 2004) and in broad categories such as communication skills or leadership skills (Michinov, Olivier-Chiron, Rusch, & Chiron, 2008). Such perspectives ignore the relevance of profession and situation specific meta-knowledge as emphasized in this thesis. This new perspective implies that during medical education, medical students should not only diagnose virtual patients individually, but in collaboration with physicians of different professional backgrounds. By that, medical students might not only be enabled to develop and improve illness scripts (Custers, 2015), but also to cultivate differentiated internal collaboration scripts (Fischer et al., 2013).

Furthermore, this thesis showed that agent-based simulations are a suitable means to facilitate collaborative diagnostic reasoning and the development of collaboration scripts. Therefore, they could be a beneficial supplement to traditional teaching (Graesser et al., 2017). Via agent-based simulations, medical students could quickly get in contact with collaborative diagnostic situations and train specific subskills deliberately in a risk-free learning environment (Ericsson, 2004; Schmidt & Rikers, 2007; Schuwirth & Van der Vleuten, 2003). When developing agent-based simulations, medical educators should identify key collaborative situations that are particularly difficult or occur rarely but are overly critical. With increasing expertise, the complexity of the simulations could be increased continuously until learning with full-scale team trainings becomes feasible (e.g., Gardner & Ahmed, 2014).

Finally, the complexity of learning environments can be reduced by using scaffolds (e.g., Vogel et al., 2017). This thesis shows that adaptive collaboration scripts can support the internalization of difficult subskills of collaborative diagnostic reasoning without hampering basic psychological needs. For designing effective adaptive collaboration scripts, the main challenge for medical educators is to model collaborative processes and to define successful collaboration (see Rosen, 2015). Focusing on specific situations and specific subskills as in the agent-based simulation used in this thesis could help to develop externally valid collaboration scripts.

Overall, the findings of this thesis suggest that providing early opportunities for medical students to engage in collaborative diagnostic reasoning seems beneficial. For that, agent-based simulation could provide a useful standardization that allows to evoke particular difficult or rare collaborative situations (Rosen, 2015) and to learn diagnosing in different group constellations. Moreover, for the learning of difficult subskills, providing scaffolding such as adaptive collaboration scripts seems to be an effective instructional mean. At the same time it should be kept in mind that modelling collaborative diagnostic reasoning requires a careful analysis of the underlying skills.

5.5 Limitations

This thesis is not without limitations that in some respect might lower its generalizability. The first limitation concerns the use of agent-based simulations. Although the use of agent-based simulations entails many advantages for investigating collaborative processes, it may run into danger of limiting validity of collaborative processes. Agent-based simulations reduce the complexity of collaboration since possible collaborative activities must be defined a priori, particularly if the simulation is not based on natural language processing (Graesser et al., 2017; Herborn et al., 2018). Yet, modelling collaborative processes and defining their quality is difficult and a deep understanding of collaborative processes is necessary to authentically model these processes (OECD, 2017). Thus, important collaborative processes could be omitted or represented in a flawed way if the simulation is not designed and validated carefully. I addressed this danger by focusing on a specific collaborative diagnostic situation that practitioners and prior research (Davies et al., 2018) highlighted as particularly difficult. The simulation was developed in collaboration with medical experts and validity evidence was collected in a comprehensive validation study (see Study 1). Thus, the processes represented in the present simulation are likely to be authentic. However, the validity argument of simulations could be improved even further if collaborative processes during learning with the agent-based simulation were compared to the processes during human-to-agent version of the simulation (see Herborn et al., 2018).

Although the present focus on a specific situation allowed to reduce the complexity of the situation and, thus, facilitated the production of a valid model of that situation in the simulation, this narrow focus may reduce generalizability of the results. The present thesis focused on a specific situation, which is the collaborative generation of radiologic evidence between internists and radiologist with a focus on evidence elicitation and sharing processes. These collaborative diagnostic processes were, however, not chosen randomly but based on prior research showing that sharing and elicitation of evidence are particularly difficult for physicians (e.g., Davies et al., 2018; Larson, Christensen, Abbot, & Franz, 1999; Tschan et al., 2009). Yet, whether our results generalize to other diagnostic situations or other domains is still subject to further investigation. Thus, future research should transfer our results to provide evidence for their generalizability.

Further limitations concern the intervention investigated in Study 3 from which conclusions on the facilitation of collaborative diagnostic reasoning were drawn. Firstly, the intervention period lasted about 1.5 hours during which medical students worked on four patient cases. Thus, the intervention was rather short. Providing learners with the opportunity to engage in collaborative diagnostic processes and supporting them with collaboration scripts serves the goal to enable medical students to developing differentiated illness scripts (Custers, 2015) and internal collaboration scripts (Fischer et al., 2013). It seems reasonable to object that after solving four patient cases, medical students are a long way from having developed differentiated illness scripts or internal collaboration scripts. Yet, the results of the study indicate that even after such short-term interventions, medical students improved their collaborative diagnostic reasoning performance. Nevertheless, future research should investigate how internal collaboration scripts and illness scripts develop in long-term interventions when learning with agent-based simulations and collaboration scripts.

Finally, I investigated how collaboration scripts affect basic psychological needs in order to address the criticism of collaboration scripts undermining learners' agency (Wise & Schwarz, 2017). In Study 2, the included studies examined different conceptualizations of motivation such as expectancy-value models or motivation to participate. Thus, these studies did not directly address the criticism of collaboration scripts undermining learners' agency. In Study 3, the criticism was assessed more directly by assessing satisfaction of basic psychological needs. However, the collaboration scripts used in Study 3 featured only low levels of coerciveness since they did not limit learners' choices during the collaborative diagnostic process. This implies that the effect for low coercive scripts could be smaller than previously assumed and, therefore, a lack of statistical power could have affected the findings. This may limit generalizability of the findings to low coercive collaboration scripts, which are, nevertheless, widely-used in research in educational psychology and are, therefore, of particular relevance. In all, this thesis highlighted that to date, there is little evidence directly addressing the criticism and future research should examine the effect of different collaboration scripts on intrinsic motivation or basic psychological needs in different contexts and with varying levels of script coerciveness (Radkowitsch et al., 2020).

5.6 Directions for Future Research

The main focus of this thesis was the investigation of collaborative diagnostic reasoning and means to facilitate it when learning with agent-based simulations. Besides providing answers to the posed questions, the findings also highlight some promising directions for future research. Firstly, although the findings allow for some conclusions about the proposed CDR model (Radkowitsch, Sailer, et al., accepted), a comprehensive validation of the model remains desirable. In Radkowitsch, Sailer, et al. (accepted), several testable assumptions that could guide the validation process of the model were already proposed. A next step in the validation of the CDR model could shift the focus to individual diagnostic activities, and how they are affected by the collaboration partners' collaborative diagnostic activities. The idea that collaboration affects cognitive processes interactively is not new and is one explanation for the effectiveness of collaborative learning (Chi & Wylie, 2014; King, 2007; Teasley, 1997). Yet, a thorough understanding of how diagnosticians' influence each other's diagnostic processes is still lacking. The CDR assumes, for instance, that evidence and hypotheses introduced by the collaboration partner may affect a diagnostician's diagnostic reasoning processes. This influence might depend on situational characteristics as well as on group constellations. For example, prior research has shown that the perceived expertise of a collaboration partner affects the extent to which an individual relies on their contribution (Andrews & Rapp, 2014). Thus, future research should identify under which conditions and how a collaboration partners' contributions affect the own collaborative diagnostic activities. This research could then inform researchers and educators, for instance, on how to adapt scaffolds to specific group constellations.

Moreover, the CDR model assumes that domain-general skills as well as domain-specific knowledge and skills influence collaborative diagnostic reasoning. There is clear evidence that individual diagnostic reasoning is largely dependent on domain-specific knowledge since diagnostic reasoning performance correlates poorly between different illnesses and clinical domains (Norman, 2005; Schuwirth & Van der Vleuten, 2003; Wimmers et al., 2007). However, there is also evidence that diagnostic reasoning (Wimmers et al., 2007) and complex problem solving (Stadler et al., 2015) in general depend on domain-general abilities. The same applies to collaboration skills. On the one hand, Study 3 shows that meta-knowledge about the collaboration partners' responsibilities, tasks, and roles affects collaborative diagnostic processes. This indicates that there is a domain-specific professional collaboration knowledge that affects collaboration. On the other hand, general social factors such as personality also affect collaborative problem solving performance (Stadler et al., 2019). Thus, future research should examine the influence of domain-general cognitive and social skills and how this influence develops with growing expertise (see Hetmanek et al., 2018). Such research could prove very valuable for informing researchers and educators on how to design scaffolds on prior social and cognitive characteristics.

A further direction for future research could lie in examining the transferability of the CDR model and the findings of the presented studies to other diagnostic situations and to other domains. Since this thesis focused on sharing and elicitation processes in the medical context, future research should examine the relations assumed in the CDR focusing on different collaborative diagnostic activities such negotiating hypotheses. To understand whether the identified relations are specific to the situation investigated, the findings should be replicated in different collaborative diagnostic situations such as collaboration between internists and surgeons, or in other domains such as collaborative diagnosing between teachers (see Heitzmann et al., 2019). Compared to other disciplines, medicine is a highly standardized domain in which many diagnostic procedures follow standardized guidelines. This standardization of processes could also affect collaborative diagnostic processes and therefore limit the transferability to other domains. However, a direct comparison between domains could help to provide a more thorough understanding of collaborative diagnostic reasoning in different domains and contribute further to the validation of the assumptions underlying the CDR model.

Finally, future research should further investigate the conditions under which adaptive collaboration scripts are effective. The findings in Study 3 suggest that adaptive collaboration scripts are not consistently more beneficial than static collaboration scripts although they support medical students to internalize complex subskills and led to a higher perceived competence when compared to static support. Although there is much research on adaptive support in individual learning settings, for instance, in form of intelligent tutoring systems (e.g., Steenbergen-Hu & Cooper, 2014), research on adaptivity in collaborative learning contexts (e.g., Olsen, Aleven, & Rummel, 2017) and particularly on adaptive collaboration scripts remains scarce. In research on collaboration scripts, adaptivity is mostly realized in form of static fading (e.g., Wecker & Fischer, 2011). The findings in Study 3 suggest, however, that adaptivity is more complex and that systematizing research is necessary (Rummel et al., 2016). Thus, future research should systematically vary the bases for decision for adaptivity (i.e., to what variable is the collaboration script adapted), the mechanism of adaptivity (i.e., fading or just-in-time adaptivity), and the skill targeted by the scaffold.

6 Conclusion

Anika Radkowitsch

For the daily routines of many professionals, collaborative diagnostic reasoning is essential. A lack of competences and failures can have severe consequences, for instance for the patients' well-being in the context of medicine (e.g., Davies et al., 2018). Therefore, the present thesis focused on the understanding and facilitating collaborative diagnostic reasoning. Collaborative diagnostic reasoning is a demanding and complex competence that requires individual and collaborative diagnostic activities that influence each other and are influenced by knowledge about collaboration partners and situational characteristics (Radkowitsch, Sailer, et al., accepted). The present thesis introduced an agent-based simulation that offers opportunities for repeated practice in standardized settings, easily can evoke situations that need specific practice (e.g., Rosen, 2015), and at the same time allows implementing additional instructional support for particularly difficult subskills. The findings of the reported studies illustrated how learning only with the agent-based simulation can suffice for facilitating some subskills (e.g., evidence elicitation). However, facilitating other more difficult subskills requires additional instructional support (e.g., evidence sharing), for instance, by using collaboration scripts. Thus, a first step for the effective facilitation of collaborative diagnostic reasoning is to develop agent-based simulations and to identify subskills that require further instructional support to be mastered. This thesis shows that collaboration scripts can be an effective scaffold for advancing collaboration skills, but may also pose extraneous cognitive load on learners and affect their perceived competence when implemented in agent-based simulations. Adapting collaboration scripts to the learners' performance seemed to address these issues. Importantly, adaptive collaboration scripts supported the internalization of collaborative diagnostic activities and had positive effects on affective measures, particularly on perceived competence. Thus, using adaptive collaboration scripts focusing on specific subskills for facilitating collaborative diagnostic competences when learning with agent-based simulations seems to provide effective learning conditions and could be effective means to exploit the full potential of agent-based simulations.

In all, the thesis highlights the importance and complexity of learning to diagnose collaboratively and emphasized the potential of agent-based simulations. Agent-based simulations provide beneficial opportunities for practice and at the same time allow the standardization of dynamic collaborative processes. By systematically implementing agent-based simulations complemented with adaptive collaboration scripts in higher education, students could benefit from early opportunities of practice and, by that, become more proficient collaborators in the future.

7 References

Abele, S. (2018). Diagnostic problem-solving process in professional contexts: Theory and empirical investigation in the context of car mechatronics using computer-generated log-files. *Vocations and Learning, 11*(1), 133-159. doi:10.1007/s12186-017-9183-x

Andrews-Todd, J., & Forsyth, C.-M. (2018). Exploring social and cognitive dimensions of collaborative problem solving in an open online simulation-based task. *Computers in Human Behavior*. doi:10.1016/j.chb.2018.10.025

Andrews, J. J., & Rapp, D. N. (2014). Partner characteristics and social contagion: Does group composition matter? *Applied Cognitive Psychology, 28*(4), 505-517. doi:10.1002/acp.3024

Barrows, H. S., & Feltovich, P. J. (1987). The clinical reasoning process. *Medical Education, 21*(2), 86-91. doi:10.1111/j.1365-2923.1987.tb00671.x

Belland, B. R., Walker, A. E., Kim, N. J., & Lefler, M. (2017). Synthesizing results from empirical research on computer-based scaffolding in STEM education: A meta-analysis. *Review of Educational Research, 87*(2), 309–344. doi:10.3102/0034654316670999

Bichler, S., Schwaighofer, M., Stadler, M., Bühner, M., Greiff, S., & Fischer, F. (2019). How working memory capacity and shifting matter for learning with worked examples—A replication study. *Journal of Educational Psychology*. doi:10.1037/edu0000433

Boshuizen, H. P., Gruber, H., & Strasser, J. (2020). Knowledge restructuring through case processing: The key to generalise expertise development theory across domains? *Educational Research Review*, 100310. doi:10.1016/j.edurev.2020.100310

Boshuizen, H. P., & Schmidt, H. G. (1992). On the role of biomedical knowledge in clinical reasoning by experts, intermediates and novices. *Cognitive Science, 16*(2), 153-184. doi:10.1016/0364-0213(92)90022-M

Bradley, P. (2006). The history of simulation in medical education and possible future directions. *Medical Education, 40*(3), 254-262. doi:10.1111/j.1365-2929.2006.02394.x

Brady, A., Laoide, R. Ó., McCarthy, P., & McDermott, R. (2012). Discrepancy and error in radiology: Concepts, causes and consequences. *Ulster Medical Journal, 81*(1), 3.

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*(1), 32-42. doi:10.3102/0013189X018001032

Cannon-Bowers, J. A., Salas, E., & Converse, S. (1993). Shared mental models in expert team decision making. In N. J. Catellan (Ed.), *Individual and group decision making: Current issues* (pp. 221-246). Hillsdale, NJ: Erlbaum.

Cannon-Bowers, J. A., Tannenbaum, S. I., Salas, E., & Volpe, C. E. (1995). Defining competencies and establishing team training requirements. In R. Guzzo & E. Salas (Eds.), *Team effectiveness and decision making in organizations* (pp. 333-380). San Francisco, CA: Jossey-Bass.

Chaiken, S., & Trope, Y. (1999). *Dual-Process Theories in Social Psychology*. New York, NY: Guilford Press.

Charlin, B., Boshuizen, H. P. A., Custers, E. J., & Feltovich, P. J. (2007). Scripts and clinical reasoning. *Medical Education, 41*(12). doi:10.1111/j.1365-2923.2007.02924.x

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. doi:10.1111/j.1365-2923.2012.04242.x

Chernikova, O., Heitzmann, N., Fink, M. C., Timothy, V., Seidel, T., & Fischer, F. (2020). Facilitating diagnostic competences in higher education - A meta-analysis in medical and teacher education. *Educational Psychology Review, 32*, 157–196. doi:10.1007/s10648-019-09492-2

Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist, 49*, 219-243. doi:10.1080/00461520.2014.965823

Christensen, C., Larson, J. R., Abbott, A., Ardolino, A., Franz, T., & Pfeiffer, C. (2000). Decision making of clinical teams: Communication patterns and diagnostic error. *Medical Decision Making, 20*(1), 45-50. doi:10.1177/0272989X0002000106

Claus, A. M., & Wiese, B. S. (2019). Development and test of a model of interdisciplinary competencies. *European Journal of Work and Organizational Psychology, 28*(2), 191-205. doi:10.1080/1359432X.2019.1567491

Cook, D. A. (2014). How much evidence does it take? A cumulative meta-analysis of outcomes of simulation-based education. *Medical Education, 48*(8), 750-760. doi:10.1111/medu.12473

Cook, D. A., Erwin, P. J., & Triola, M. M. (2010). Computerized virtual patients in health professions education: A systematic review and meta-analysis. *Academic Medicine, 85*(10), 1589-1602. doi:10.1097/ACM.0b013e3181edfe13

Cook, D. A., & Hatala, R. (2016). Validation of educational assessments: A primer for simulation and beyond. *Advances in Simulation, 1*, 31.

Cook, D. A., Hatala, R., Brydges, R., Zendejas, B., Szostek, J. H., Wang, A. T., . . . Hamstra, S. J. (2011). Technology-enhanced simulation for health professions education: A systematic review and meta-analysis. *JAMA, 306*(9), 978-988.

Custers, E. J. (2015). Thirty years of illness scripts: Theoretical origins and practical applications. *Medical Teacher, 37*(5), 457-462. doi:10.3109/0142159X.2014.956052

Custers, E. J., Boshuizen, H. P., & Schmidt, H. G. (1996). The influence of medical expertise, case typicality, and illness script component on case processing and disease probability estimates. *Memory & cognition, 24*(3), 384-399. doi:10.3758/BF03213301

Davies, S., George, A., Macallister, A., Barton, H., Youssef, A., Boyle, L., & Sequeiros, I. (2018). "It's all in the history": A service evaluation of the quality of radiological requests in acute imaging. *Radiography (London, England : 1995), 24*(3), 252–256. doi:10.1016/j.radi.2018.03.005

De Coninck, K., Valcke, M., Ophalvens, I., & Vanderlinde, R. (2019). Bridging the theory-practice gap in teacher education: The design and construction of simulation-based learning environments. In *Kohärenz in der Lehrerbildung* (pp. 263-280): Springer.

DeChurch, L. A., & Mesmer-Magnus, J. R. (2010). The cognitive underpinnings of effective teamwork: A meta-analysis. *Journal of applied psychology, 95*(1), 32-53. doi:10.1037/a0017328

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*: New York, NY: Plenum.

Demetriadis, S., Egerter, T., Hanisch, F., & Fischer, F. (2011). Peer review-based scripted collaboration to support domain-specific and domain-general knowledge acquisition in computer science. *Computer Science Education, 21*, 29-56. doi:10.1080/08993408.2010.539069

Dillenbourg, P. (2002). Over-scripting CSCL: The risks of blending collaborative learning with instructional design. In P. A. Kirschner (Ed.), *Three worlds of CSCL. Can we support CSCL?* (pp. 61-91). Heerlen, Netherlands: Open Universiteit Nederland.

Dillenbourg, P., Baker, M. J., Blaye, A., & O'Malley, C. (1996). The evolution of research on collaborative learning. In H. Spada & P. Reimann (Eds.), *Learning in Humans and Machine: Towards an Interdisciplinary Learning Science* (pp. 189-211). Oxford, GB: Elsevier.

Dillenbourg, P., & Hong, F. (2008). The mechanics of CSCL macro scripts. *International Journal of Computer-Supported Collaborative Learning, 3*(1), 5-23. doi:10.1007/s11412-007-9033-1

Elstein, A. S., & Schwarz, A. (2002). Clinical problem solving and diagnostic decision making: Selective review of the cognitive literature. *Bmj, 324*(7339), 729-732. doi:10.1136/bmj.324.7339.729

Elstein, A. S., Shulman, L. S., & Sprafka, S. A. (1978). *Medical problem solving: An analysis of clinical reasoning*. Cambridge, MA: Harvard University Press.

Engelmann, T., & Hesse, F. W. (2010). How digital concept maps about the collaborators' knowledge and information influence computer-supported collaborative problem solving. *International Journal of Computer-Supported Collaborative Learning, 5*(3), 299-319. doi:10.1007/s11412-010-9089-1

Engelmann, T., & Hesse, F. W. (2011). Fostering sharing of unshared knowledge by having access to the collaborators' meta-knowledge structures. *Computers in Human Behavior, 27*, 2078–2087. doi:10.1016/j.chb.2011.06.002

Ericsson, K. A. (2004). Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Academic Medicine, 79*(10), 70-81.

Eva, K. W. (2005). What every teacher needs to know about clinical reasoning. *Medical Education, 39*(1), 98-106. doi:10.1111/j.1365-2929.2004.01972.x

Eva, K. W., Hatala, R. M., LeBlanc, V. R., & Brooks, L. R. (2007). Teaching from the clinical reasoning literature: Combined reasoning strategies help novice diagnosticians overcome misleading information. *Medical Education, 41*(12), 1152-1158. doi:10.1111/j.1365-2923.2007.02923.x

Feltovich, P. J., & Barrows, H. S. (1984). Issues of Generality in Medical Problem Solving. In *Tutorials in Problem-Based Learning: A new Direction in Teaching the Health Professions* (pp. 128-142). Assen, NL: Van Gorcum.

Fiore, S. M., Rosen, M. A., Smith-Jentsch, K. A., Salas, E., Letsky, M., & Warner, N. (2010). Toward an understanding of macrocognition in teams: Predicting processes in complex collaborative contexts. *Human Factors, 52*, 203-224. doi:10.1177/0018720810369807

Fischer, F., Bruhn, J., Gräsel, C., & Mandl, H. (2002). Fostering collaborative knowledge construction with visualization tools. *Learning and Instruction, 12*(2), 213-232. doi:10.1016/S0959-4752(01)00005-6

Fischer, F., Kollar, I., Stegmann, K., & Wecker, C. (2013). Toward a script theory of guidance in computer-supported collaborative learning. *Educational Psychologist, 48*, 56-66. doi:10.1080/00461520.2012.748005

Fischer, F., Kollar, I., Ufer, S., Sodian, B., Hussmann, H., Pekrun, R., . . . Eberle, J. (2014). Scientific reasoning and argumentation: Advancing an interdisciplinary research agenda in education. *Frontline Learning Research*(5), 28–45. doi:10.14786/flr.v2i3.96

Flin, R., & Maran, N. (2004). Identifying and training non-technical skills for teams in acute medicine. *BMJ Quality & Safety, 13*(suppl 1), i80-i84.

Gardner, A. K., & Ahmed, R. A. (2014). Transforming trauma teams through transactive memory: Can simulation enhance performance? *Simulation & Gaming, 45*(3), 356-370. doi:10.1177/1046878114547836

Gräbner, C., Heimberger, P., Kapeller, J., & Schütz, B. (2020). Is Europe Disintegrating? Macroeconomic Divergence, Structural Polarisation, Trade and Fragility. *Cambridge Journal of Economics*, bez059. doi:10.1093/cje/bez059

Graesser, A. C., Fiore, S. M., Greiff, S., Andrews-Todd, J., Foltz, P. W., & Hesse, F. (2018). Advancing the science of collaborative problem solving. *Psychological Science in the Public Interest, 19*(2), 59–92. doi:10.1177/1529100618808244

Graesser, A. C., Kuo, B.-C., & Liao, C. h.-H. (2017). Complex Problem Solving in Assessments of Collaborative Problem Solving. *Journal of Intelligence, 5*(2). doi:10.3390/jintelligence5020010

Griffin, P., & Care, E. (Eds.). (2015). *Assessment and Teaching of 21st Century Skills: Methods and Approach*. Dordrecht, NL: Springer

Hao, J., Liu, L., von Davier, A., & Kyllonen, P. (2015). Assessing collaborative problem solving with simulation based tasks. In O. Lindwall, P. Häkkinen, T. Koschman, P. Tchounikine, & S. Ludvigsen (Eds.), *Exploring the Material Conditions of Learning: The Computer Supported Collaborative Learning (CSCL) Conference 2015, Volume 2*. Gothenburg, Sweden: International Society of the Learning Sciences.

Hays, R. T., Jacobs, J. W., Prince, C., & Salas, E. (1992). Flight simulator training effectiveness: A meta-analysis. *Military Psychology, 4*(2), 63-74.

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. doi:10.1111/medu.12778

Heitzmann, N., Seidel, T., Opitz, A., Hetmanek, A., Wecker, C., Fischer, M., . . . Fischer, F. (2019). Facilitating diagnostic competences in simulations: A conceptual framework and a research agenda for medical and teacher education. *Frontline Learning Research, 7*(4), 1-24. doi:10.14786/flr.v7i4.384

Herborn, K., Stadler, M., Mustafić, M., & Greiff, S. (2018). The assessment of collaborative problem solving in PISA 2015: Can computer agents replace humans? *Computers in Human Behavior, 104*, 105624. doi:10.1016/j.chb.2018.07.035

Hesse, F., Care, E., Buder, J., Sassenberg, K., & Griffin, P. (2015). A Framework for Teachable Collaborative Problem Solving Skills. In P. Griffin & E. Care (Eds.), *Assessment and Teaching of 21st Century Skills: Methods and Approach* (pp. 37-56): Springer.

Hetmanek, A., Engelmann, K., Opitz, A., & Fischer, F. (2018). Beyond Intelligence and Domain Knowledge: Scientific Reasoning and Argumentation as a Set of Cross-Domain Skills. In F. Fischer (Ed.), *Scientific reasoning and argumentation* (pp. 203–226). New York, NY: Routledge.

Hidalgo, C. A., Klinger, B., Barabási, A.-L., & Hausmann, R. (2007). The product space conditions the development of nations. *science, 317*(5837), 482-487. doi:10.1126/science.1144581

Hinsz, V. B., Tindale, R. S., & Vollrath, D. A. (1997). The emerging conceptualization of groups as information processors. *Psychological Bulletin, 121*(1), 43-64. doi:10.1037/0033-2909.121.1.43

Hodges, B., McNaughton, N., & Tiberius, R. (1999). OSCE Checklists Do Not Capture Increasing. *Acad. Med, 74*, 1129-1134.

Hutchins, E. (1995). *Cognition in the Wild*. Cambridge, GB: MIT press.

Järvelä, S., & Hadwin, A. F. (2013). New frontiers: Regulating learning in CSCL. *Educational Psychologist, 48*(1), 25-39. doi:10.1080/00461520.2012.748006

Jonassen, D. H. (2000). Toward a design theory of problem solving. *Educational Technology Research and Development, 48*(4), 63-85. doi:10.1007/BF02300500

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

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge, UK: Cambridge University Press.

Kane, M. (2006). Validation. In R. Brennan (Ed.), *Educational measurement* (pp. 17-64). Westport, CT: Praeger.

Kaufman, D., & Ireland, A. (2016). Enhancing teacher education with simulations. *TechTrends, 60*(3), 260-267. doi:10.1007/s11528-016-0049-0

Kiesewetter, J., Fischer, F., & Fischer, M. R. (2016). Collaboration expertise in medicine - No evidence for cross-domain application from a memory retrieval study. *PLoS ONE, 11*(2), e0148754. doi:10.1371/journal.pone.0148754

King, A. (2007). Scripting collaborative learning processes: A cognitive perspective. In F. Fischer, I. Kollar, H. Mandl, & J. M. Haake (Eds.), *Scripting Computer-Supported Collaborative Learning: Cognitive, Computational and Educational Perspectives* (pp. 13-37): Bosten, MA: Springer US.

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. doi:10.1207/s15326985ep4102_1

Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science, 12*, 1–48. doi:10.1207/s15516709cog1201_1

Kollar, I., Wecker, C., & Fischer, F. (2018). Scaffolding and scripting (computer-supported) collaborative learning. In F. Fischer, C. E. Hmelo-Silver, S. R. Goldman, & P. Reimann (Eds.), *International Handbook of the Learning Sciences* (pp. 340-350). New York, NY: Routledge.

Kolodner, J. L. (1992). An introduction to case-based reasoning. *Artificial Intelligence Review, 6*(1), 3-34.

Kozlov, M. D., & Große, C. S. (2016). Online collaborative learning in dyads: Effects of knowledge distribution and awareness. *Computers in Human Behavior, 59*, 389-401. doi:10.1016/j.chb.2016.01.043

Kozlowski, S. W. (2018). Enhancing the effectiveness of work groups and teams: a reflection. *Perspectives on Psychological Science, 13*(2), 205-212. doi:10.1177/1745691617697078

Larson, J. R., Christensen, C., Abbot, A. S., & Franz, T. M. (1999). Diagnosing Groups: Charting the Flow of Information in Medical Decision-Making Teams. *Journal of Personality and Social Psychology, 71*(2), 315–330.

Larson, J. R., Christensen, C., Franz, T. M., & Abbot, A. S. (1998). Diagnosing Groups: The Pooling, Management, and Impact of Shared and Unshared Case Information in Team-Based Medical Decision Making. *Journal of Personality and Social Psychology, 75*(1), 93–108. doi:10.1037/0022-3514.75.1.93

Liu, L., Hao, J., von Davier, A. A., Kyllonen, P., & Zapata-Rivera, D. (2015). A Tough Nut to Crack: Measuring Collaborative Problem Solving. In Y. Rosen, S. Ferrara, & M.

Mosharraf (Eds.), *Handbook of Research on Computational Tools for Real-World Skill Development* (pp. 344-359). Hershey, PA: IGI Global.

Lubarsky, S., Dory, V., Audétat, M.-C., Custers, E., & Charlin, B. (2015). Using script theory to cultivate illness script formation and clinical reasoning in health professions education. *Canadian medical education journal, 6*(2), e61.

Mamede, S., & Schmidt, H. G. (2017). Reflection in medical diagnosis: A literature review. *Health Professions Education, 3*(1), 15–25. doi:10.1016/j.hpe.2017.01.003

Mathieu, J. E., Heffner, T. S., Goodwin, G. F., Salas, E., & Cannon-Bowers, J. A. (2000). The influence of shared mental models on team process and performance. *Journal of applied psychology, 85*(2), 273. doi:10.1037/0021-9010.85.2.273

Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? *American psychologist, 59*(1), 14. doi:10.1037/0003-066X.59.1.14

Meier, A., Spada, H., & Rummel, N. (2007). A rating scheme for assessing the quality of computer-supported collaboration processes. *International Journal of Computer-Supported Collaborative Learning, 2*(1), 63–86.

Mende, S., Proske, A., Körndle, H., & Narciss, S. (2017). Who benefits from a low versus high guidance CSCL script and why? *Instructional Science, 45*(4), 439-468.

MFT Medizinischer Fakultätentag der Bundesrepublik Deutschland e. V. (2015). Nationaler Kompetenzbasierter Lernzielkatalog Medizin (NKLM). Retrieved from http://www.nklm.de/kataloge/nklm/lernziel/uebersicht

Michinov, E., Olivier-Chiron, E., Rusch, E., & Chiron, B. (2008). Influence of transactive memory on perceived performance, job satisfaction and identification in anaesthesia teams. *British journal of anaesthesia, 100*(3), 327-332. doi:10.1093/bja/aem404

Mohammed, S., & Dumville, B. C. (2001). Team mental models in a team knowledge framework: Expanding theory and measurement across disciplinary boundaries. *Journal of Organizational Behavior, 22*(2), 89-106. doi:10.1002/job.86

Nickerson, R.-S. (1988). Confirmation Bias: A Ubiquitous Phenomenon in Many Guises. *Review of General Psychology, 2*(2), 175–220. doi:10.1037/1089-2680.2.2.175

Norman, G. (2005). Research in clinical reasoning: Past history and current trends. *Medical Education, 39*(4), 418–427. doi:10.1111/j.1365-2929.2005.02127.x

Norman, G., Young, M., & Brooks, L. (2007). Non-analytical models of clinical reasoning: The role of experience. *Medical Education, 41*(12), 1140-1145. doi:10.1111/j.1365-2923.2007.02914.x

Noroozi, O., Teasley, S. D., Biemans, H. J. A., Weinberger, A., & Mulder, M. (2013). Facilitating learning in multidisciplinary groups with transactive CSCL scripts. *International Journal of Computer-Supported Collaborative Learning, 8*, 189-223. doi:10.1007/s11412-012-9162-z

Nye, B. D., Graesser, A. C., & Hu, X. (2014). AutoTutor and family: A review of 17 years of natural language tutoring. *International Journal of Artificial Intelligence in Education, 24*(4), 427-469. doi:10.1007/s40593-014-0029-5

OECD. (2017). *PISA 2015 Assessment and analytical framework: Science, reading, mathematic, financial literacy and collaborative problem solving, revised edition*. Paris, France: OECD Publishing.

Olsen, J., Aleven, V., & Rummel, N. (2017). Statistically modeling individual students' learning over successive collaborative practice opportunities. *Journal of Educational Measurement, 54*(1), 123-138. doi:10.1111/jedm.12137

Patel, V. L., Groen, G. J., & Arocha, J. F. (1990). Medical expertise as a function of task difficulty. *Memory & cognition, 18*(4), 394-406. doi:10.3758/BF03197128

Patel, V. L., Kaufman, D. R., & Arocha, J. F. (2002). Emerging paradigms of cognition in medical decision-making. *Journal of Biomedical Informatics*(35), 52–75.

Pea, R. D. (2004). The social and technological dimensions of scaffolding and related theoretical concepts for learning, education, and human activity. *Journal of the Learning Sciences, 13*(3), 423-451. doi:10.1207/s15327809jls1303_6

Peltokorpi, V. (2008). Transactive memory systems. *Review of General Psychology, 12*(4), 378-394. doi:10.1037/1089-2680.12.4.378

Peterson, A. T., & Roseth, C. J. (2016). Effects of four CSCL strategies for enhancing online discussion forums: Social interdependence, summarizing, scripts, and synchronicity. *International Journal of Educational Research, 76*, 147-161. doi:10.1016/j.ijer.2015.04.009

Radkowitsch, A., Fischer, M. R., Schmidmaier, R., & Fischer, F. (accepted). Learning to diagnose collaboratively: Validating a simulation for medical students. *Journal for Medical Education*.

Radkowitsch, A., Sailer, M., Fischer, M. R., Schmidmaier, R., & Fischer, F. (accepted). Diagnosing collaboratively: A theoretical model and a simulation-based learning environment. In F. Fischer & A. Opitz (Eds.), *Learning to diagnose with simulations - Examples from teacher education and medical education*. New York, NY: Springer Briefs in Education Series.

Radkowitsch, A., Sailer, M., Schmidmaier, R., Fischer, M. R., & Fischer, F. (submitted). Learning to diagnose collaboratively – Effects of adaptive collaboration scripts in agent-based medical simulations. *Learning and Instruction*.

Radkowitsch, A., Vogel, F., & Fischer, F. (2020). Good for learning, bad for motivation? A meta-analysis on the effects of computer-supported collaboration scripts. *International Journal of Computer-Supported Collaborative Learning, 15*(1), 5-47. doi:10.1007/s11412-020-09316-4

Rikers, R. M., Schmidt, H. G., & Boshuizen, H. P. (2000). Knowledge encapsulation and the intermediate effect. *Contemporary Educational Psychology, 25*(2), 150-166. doi:10.1006/ceps.1998.1000

Roschelle, J., & Teasley, S. D. (1995). The Construction of Shared Knowledge in Collaborative Problem Solving. In C. Oḿalley (Ed.), *Computer-Supported Collaborative Learning* (pp. 69-97). Berlin, Germany: Springer.

Rosen, Y. (2015). Computer-based assessment of collaborative problem solving: Exploring the feasibility of human-to-agent approach. *International Journal of Artificial Intelligence in Education, 25*(3), 380-406. doi:10.1007/s40593-015-0042-3

Rummel, N., Spada, H., & Hauser, S. (2009). Learning to collaborate while being scripted or by observing a model. *International Journal of Computer-Supported Collaborative Learning, 4*, 69-92. doi:10.1007/s11412-008-9054-4

Rummel, N., Walker, E., & Aleven, V. (2016). Different futures of adaptive collaborative learning support. *International Journal of Artificial Intelligence in Education, 26*(2), 784-795. doi:10.1007/s40593-016-0102-3

Schank, R. C. (1999). *Dynamic Memory Revisited*: Cambridge, UK: Cambridge University Press.

Schmidt, H. G., & Boshuizen, H. P. (1993). On acquiring expertise in medicine. *Educational Psychology Review, 5*(3), 205-221. doi:10.1007/BF01323044

Schmidt, H. G., & Rikers, R. M. J. P. (2007). How expertise develops in medicine: Knowledge encapsulation and illness script formation. *Medical Education, 41*(12), 1133–1139. doi:10.1111/j.1365-2923.2007.02915.x

Schmitt, L. J., & Weinberger, A. (2019). Fourth graders’ dyadic learning on multi-touch interfaces—versatile effects of verbalization prompts. *Educational Technology Research and Development, 67*(3), 519-539. doi:10.1007/s11423-018-9619-5

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological assessment, 8*(4), 350-353. doi:10.1037/1040-3590.8.4.350

Schnaubert, L., & Bodemer, D. (2019). Providing different types of group awareness information to guide collaborative learning. *International Journal of Computer-Supported Collaborative Learning, 14*, 7–51. doi:10.1007/s11412-018-9293-y

Schuwirth, L. W., & Van der Vleuten, C. P. (2003). The use of clinical simulations in assessment. *Medical Education, 37*, 65-71. doi:10.1046/j.1365-2923.37.s1.8.x

Schwaighofer, M., Vogel, F., Kollar, I., Ufer, S., Strohmaier, A., Terwedow, I., . . . Fischer, F. (2017). How to combine collaboration scripts and heuristic worked examples to foster mathematical argumentation – When working memory matters. *International Journal of Computer-Supported Collaborative Learning, 12*(3), 281–305. doi:10.1007/s11412-017-9260-z

Sherbino, J., Dore, K. L., Wood, T. J., Young, M. E., Gaissmaier, W., Kreuger, S., & Norman, G. R. (2012). The relationship between response time and diagnostic accuracy. *Academic Medicine, 87*(6), 785-791. doi:10.1097/ACM.0b013e318253acbd

Siebeck, M., Schwald, B., Frey, C., Röding, S., Stegmann, K., & Fischer, F. (2011). Teaching the rectal examination with simulations: Effects on knowledge acquisition and inhibition. *Medical Education, 45*, 1025-1031. doi:10.1111/j.1365-2923.2011.04005.x

Simmons, B. (2010). Clinical reasoning: Concept analysis. *Journal of Advanced Nursing, 66*(5), 1151-1158. doi:10.1111/j.1365-2648.2010.05262.x

Smith, R. (2010). The long history of gaming in military training. *Simulation & Gaming, 41*(1), 6-19.

Stadler, M., Becker, N., Gödker, M., Leutner, D., & Greiff, S. (2015). Complex problem solving and intelligence: A meta-analysis. *Intelligence, 53*, 92-101. doi:10.1016/j.intell.2015.09.005

Stadler, M., Herborn, K., Mustafić, M., & Greiff, S. (2019). Computer-based collaborative problem solving in PISA 2015 and the role of personality. *Journal of Intelligence, 7*(3), 15. doi:10.3390/jintelligence7030015

Stasser, G., & Titus, W. (1985). Pooling of unshared information in group decision making: Biased information sampling during discussion. *Journal of Personality and Social Psychology, 48*(6), 1467-1467. doi:10.1037/0022-3514.48.6.1467

Stasser, G., & Titus, W. (1987). Effects of information load and percentage of shared information on the dissemination of unshared information during group discussion. *Journal of Personality and Social Psychology, 53*(1), 81-81.

Stasser, G., Vaughan, S. I., & Stewart, D. D. (2000). Pooling unshared information: The benefits of knowing how access to information is distributed among group members.

Organizational behavior and human decision processes, 82(1), 102-116. doi:10.1006/obhd.2000.2890

Steenbergen-Hu, S., & Cooper, H. (2014). A meta-analysis of the effectiveness of intelligent tutoring systems on college students' academic learning. *Journal of Educational Psychology, 106*(2), 331. doi:10.1037/a0034752

Stegmann, K., Mu, J., Gehlen-Baum, V., & Fischer, F. (2011). The myth of over-scripting: Can novices be supported too much? In H. Spada, G. Stahl, N. Miyake, & N. Law (Eds.), *Connecting Computer-Supported Collaborative Learning to Policy and Practice: CSCL2011 Conference Proceedings. Volume I*. Hong-Kong, China: International Society of the Learning Sciences.

Stegmann, K., Weinberger, A., & Fischer, F. (2007). Facilitating argumentative knowledge construction with computer-supported collaboration scripts. *International Journal of Computer-Supported Collaborative Learning, 2*, 421-447. doi:10.1007/s11412-007-9028-y

Sun, C., Shute, V. J., Stewart, A., Yonehiro, J., Duran, N., & D'Mello, S. (2020). Towards a generalized competency model of collaborative problem solving. *Computers & Education, 143*, 103672. doi:10.1016/j.compedu.2019.103672

Tabak, I. (2004). Synergy: A complement to emerging patterns of distributed scaffolding. *The Journal of the Learning Sciences, 13*, 305-335. doi:10.1016/j.chb.2018.12.010

Tabak, I., & Kyza, E. A. (2018). Research on scaffolding in the learning sciences: A methodological perspective. In F. Fischer, C. E. Hmelo-Silver, S. R. Goldman, & P. Reimann (Eds.), *International Handbook of the Learning Sciences* (pp. 191-200). New York, NY: Routledge.

Teasley, S. D. (1997). Talking about reasoning: How important is the peer in peer collaboration? In L. R. Resnick, R. Säljö, C. Pontecorvo, & B. Burge (Eds.), *Discourse, tools and reasoning: Essays on situated cognition* (pp. 361-384). Berlin, Germany: Springer.

Tschan, F., Semmer, N. K., Gurtner, A., Bizzari, L., Spychiger, M., Breuer, M., & Marsch, S. U. (2009). Explicit reasoning, confirmation bias, and illusory transactive memory: A simulation study of group medical decision making. *Small Group Research, 40*(3), 271-300. doi:10.1177/1046496409332928

Van Joolingen, W. R., & De Jong, T. (1997). An extended dual search space model of scientific discovery learning. *Instructional Science, 25*(5), 307-346. doi:10.1023/A:1002993406499

Van Schaik, P., Flynn, D., Van Wersch, A., Douglass, A., & Cann, P. (2005). Influence of illness script components and medical practice on medical decision making. *Journal of experimental psychology: Applied, 11*(3), 187-199.

Vogel, F., Wecker, C., Kollar, I., & Fischer, F. (2017). Socio-cognitive scaffolding with computer-supported collaboration scripts: A meta-analysis. *Educational Psychology Review, 29*, 477 - 511. doi:10.1007/s10648-016-9361-7

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*: Cambridge, MA: Harvard University Press.

Wecker, C., & Fischer, F. (2011). From guided to self-regulated performance of domain-general skills: The role of peer monitoring during the fading of instructional scripts. *Learning and Instruction, 21*(6), 746-756. doi:10.1016/j.learninstruc.2011.05.001

Wegner, D. M. (1987). Transactive memory: A contemporary analysis of the group mind. In B. Mullen & G. R. Goethals (Eds.), *Theories of group behavior* (pp. 185-208). New York, NY: Springer.

Weinberger, A., Kollar, I., Dimitriadis, Y., Mäkitalo-Siegl, K., & Fischer, F. (2009). Computer-supported collaboration scripts: Perspectives from educational psychology and computer science. In N. Balacheff, S. Ludvigsen, T. De Jong, A. Lazonder, S.-A. Barnes, & L. Montandon (Eds.), *Technology-enhanced learning - Principles and Products* (pp. 155-173). Dordrecht, Netherlands: Springer Netherlands.

Wildgans-Lang, A., Scheuerer, S., Obersteiner, A., Fischer, F., & Reiss, K. (2020). Analyzing prospective mathematics teachers' diagnostic processes in a simulated environment. *ZDM*, 1-14. doi:10.1007/s11858-020-01139-9

Wimmers, P. F., Splinter, T. A., Hancock, G. R., & Schmidt, H. G. (2007). Clinical competence: General ability or case-specific? *Advances in Health Sciences Education, 12*(3), 299-314. doi:10.1007/s10459-006-9002-x

Wise, A. F., & Schwarz, B. B. (2017). Visions of CSCL: Eight provocations for the future of the field. *International Journal of Computer-Supported Collaborative Learning, 12*(4), 423–467. doi:10.1007/s11412-017-9267-5

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry, 17*(2), 89-100. doi:10.1111/j.1469-7610.1976.tb00381.x

Woods, N. N., Howey, E. H., Brooks, L. R., & Norman, G. R. (2006). Speed kills? Speed, accuracy, encapsulations and causal understanding. *Medical Education, 40*(10), 973-979. doi:10.1111/j.1365-2929.2006.02556.x

Ziv, A., Wolpe, P. R., Small, S. D., & Glick, S. (2003). Simulation-based medical education: An ethical imperative. *Academic Medicine: Journal of the Association of American Medical Colleges, 78*(8), 783-788. doi:10.1097/00001888-200308000-00006

Zottmann, J., Dieckmann, P., Taraszow, T., Rall, M., & Fischer, F. (2018). Just watching is not enough: Fostering simulation-based learning with collaboration scripts. *GMS Journal for Medical Education, 35*(3), 1–18. doi:10.3205/zma001181

8 Appendices

Anika Radkowitsch

A Patient Cases Study 1 and 3

For Study 1 and 3, fictitious patient cases were developed. The following Table A1 describes an overview of which patient cases were used in Study 1 and 3. Afterwards, the health record of one patient case is presented exemplarily. All patient cases as used in Study 1 and 3 are provided via the electronic supplementary material (see Appendix M).

Table A1

Overview about all patient cases used in Study 1 and 3

Patient	Study 1	Study 3	Diagnosis
Marianne Freundorf	Case 2	Pretest	Acute pancreatitis
Sabine Winkler[a]	-	Intervention case 1	Community acquired pneumonia
Anton Fomin[a]	-	Intervention case 2	Tuberculosis
Mark Binder	Case 3	Intervention case 3	Pneumocystis jirovecii pneumonia (PJP)
Oliver Forster[a]	-	Intervention Case 4	Osteomyelitis
Herma Goettlich	Case 1	Posttest 1	Aspiration pneumonia
Maria Schenker	Case 5	Posttest 2	Hospital acquired pneumonia
Ute Wenninger	Case 4	-	Sigmoid diverticulitis

Note. [a] These patient cases were not used in Study 1. Their difficulty was evaluated in a second validation study that was not part of the presented thesis prior of their use in Study 3.

Example case: Herma Goettlich

Vorstellung der Patientin

Sie arbeiten seit einigen Monaten in einem mittelgroßen Kreiskrankenhaus und sind derzeit auf einer allgemeininternistischen Station eingesetzt. Heute betreuen Sie zusätzlich die Notaufnahme. Am späten Montagvormittag wird die 78-jährige Herma Göttlich vom Notarzt gebracht, der besorgte Ehemann begleitet sie.
Frau Göttlich leidet unter starker Atemnot, so dass der Ehemann einen Großteil Ihrer Fragen beantwortet. Sie haben Blut abgenommen und "eilig" ins Labor geschickt, Frau Göttlich so weit möglich anamnestiziert und untersucht. Als Sie damit fertig sind, ist auch ein Teil des Labors schon fertig und Sie können sich mit der Akte überlegen, was die nächsten diagnostischen Schritte sein sollen.

Rettungsdienstprotokoll

78-jährige Patientin mit Fieber seit heute Morgen und rasch progredienter Luftnot. PO2 initial 92, unter 2 l O2 Besserung der Symptomatik, vorerst auf Intubation verzichtet. EKG unauffällig. Troponin Schnelltest negativ.

Anamnese

Herr Göttlich berichtet, dass seine Frau in letzter Zeit einiges durchmachen musste. Sie habe einen Schlaganfall gehabt, habe insgesamt ca. 8 kg an Gewicht abgenommen. Es ging ihr bis heute den Umständen entsprechend gut - sie könne zwar nur schlecht laufen, sei jetzt überwiegend im Rollstuhl mobilisiert, habe sich aber wohl gefühlt und sei guter Dinge gewesen. Heute Morgen sei er davon aufgewacht, dass sie zunehmend unruhig geworden sei, vermehrt angefangen habe zu schwitzen und schlecht Luft bekommen habe. Gestern Mittag schien noch alles in Ordnung gewesen zu sein. Er selbst sei das erste Mal seit langer Zeit bei Freunden zu Besuch gewesen, weil seine Tochter, die den Tag über zu Besuch war, ausnahmsweise die Pflege der Patientin übernehmen und mit ihr zu Abend essen konnte.

Vorerkrankungen

- Z.n. Mediainfarkt mit Dysphagie und Hemiparese vor 6 Wochen,
- Osteoporose,
- Beginnendes dementielles Syndrom,
- Z.n. Tonsillektomie 1962,
- Z.n. tiefe Venenthrombose rechts 2005

Medikamente

ASS 100, Ramipril, Simvastatin, Calcium/D3

Genussmittelanamnese

Nikotin ca. 10 py, ex vor 10 Jahren; Alkohol selten.

Sozialanamnese

Rentnerin, früher Metzgereifachverkäuferin.

Allergien

Keine bekannten Allergien, auch nicht gegen Medikamente oder Kontrastmittel.

Körperliche Untersuchung

78-jährige Patientin in reduziertem AZ und gutem EZ (1,75 m, 72 kg, BMI 23,5 kg/m2).

Vitalparameter

- RR 105/60 mmHg,
- P 102/min reg.,
- Temp. 37,9 °C,
- Atemfrequenz 27/min,
- pO2 96 % unter 2 l O2.

Lymphknoten

nicht vergrößert, nicht druckdolent. Schilddrüse unauffällig.

Kardiovaskuläres System

Keine Zyanose. Herztöne rein, regelmäßig, tachykard, keine Extratöne oder pathologische Herzgeräusche. Keine Jugularvenenstauung.
Mäßige Unterschenkelödeme beidseits rechts > links 2cm Seitendifferenz, Periphere Pulse seitengleich tastbar. Schleimhäute unauffällig.

Respiratorisches System

Symmetrische Thoraxexkursion, keine Einziehungen, normale Thoraxform.
Kein Stimmfremitus, kein Stridor. Gleichstand der Zwerchfelle, bilateral 4 cm atemverschieblich, kein H.a. Pleuraerguss. Lunge ubiquitär belüftet mit grobblasigen RGs v.a. rechts, Husten mit übelriechendem Auswurf, kein Pleurareiben.

Abdomen

Bauchdecke weich, kein Druckschmerz, keine Resistenzen, keine Abwehrspannung, Darmgeräusche regelrecht in allen Quadranten.
Nieren nicht klopfschmerzhaft, Milz nicht vergrößert tastbar, Leber 11 cm in der rechten MCL, Oberfläche glatt. Keine Hernien. Keine sichtbaren Operationsnarben.

Haut

Unauffälliger Hautbefund. Extremitäten warm, keine Varikosis. Keine Nagelveränderungen.

Bewegungsapparat

Normale Beweglichkeit aller Gelenke. Keine Gelenkschmerzen, -schwellungen oder -deformitäten. Wirbelsäule nicht klopfschmerzhaft. Meyer-, Homans-, Payr-Zeichen negativ.

Orientierende neurologische Untersuchung

Freundlich zugewandt, agitiert, in allen Qualitäten orientiert, kein Hinweis auf formale Denkstörungen oder Suizidalität. Pupillenlichtreaktion direkt und indirekt prompt und seitengleich. Bekannte Fazialis- und Hemiparese rechts. Darüber hinaus keine Paresen, kein Sensibilitätsdefizit, keine pathologischen Reflexe, kein Absinken in den Halteversuchen. Kein Meningismus. Vibrationsempfinden 8/8 an allen vier Extremitäten.

Labor

Sie haben folgende Laborwerte erhalten:

Blutbild

	Befund	Einheit	Referenz Frauen	Einheit
Erythrozyten	3,8 x 106	/µl	3,5 - 5 x 106	/µl
Hämoglobin (Hb)	13,6	g/dl	12 - 15	g/dl
MCH	28	pg	27 - 34	pg
MCV	84	fl	81 - 100	fl
MCHC	33	g/dl	32 - 36	g/dl
Hämatokrit (Hkt)	38	%	33-43	%
Leukozyten	13,6 x 10^3	/µl	4 - 11 x 103	/µl
Thrombozyten	182.000	/µl	150.000 - 400.000	/µl
Retikulozyten	1	%	0,5 - 2	%
Differentialblutbild				
Neutrophile Granulozyten	78	%	45 - 78	%
Stabkernige	4	%	0 - 4	%
Segmentkernige	74	%	45 - 74	%
Eosinophile Granulozyten	1	%	0 - 7	%
Basophile Granulozyten	1	%	0 - 2	%
Lymphozyten	16	%	16 - 45	%
Monozyten	4	%	04 - 10	%

Gerinnung

	Befund	Einheit	Referenz Frauen	Einheit
Quick	100	%	70 - 120	%
INR	1		1	
PTT	38	sec.	28 - 40	sec.

Serum

	Befund	Einheit	Referenz Frauen	Einheit
Serum				
Natrium	142	mmol/l	136 - 148	mmol/l
Kalium	4,7	mmol/l	3,6 - 5,2	mmol/l
Calcium (gesamt)	2,3	mmol/l	2,1 - 2,6	mmol/l
Kreatinin	0,9	mg/dl	< 0,9	mg/dl
eGFR	>60	ml/min/1,73 m2 KOF	> 60	ml/min/1,73 m2 KOF
Harnstoff	21	mg/dl	10 - 50	mg/dl
Alkalische Phosphatase	45	U/I	40 - 190	U/I
Bilirubin (gesamt)	1	mg/dl	< 1,1	mg/dl
Bilirubin (direkt)	0,6	mg/dl	< 0,6	mg/dl
CHE	4,6	kU/I	2,5 - 7,4	kU/I
GOT (ASAT)	13	U/I	< 15	U/I
GPT (ALAT)	8	U/I	< 17	U/I
ɣ-GT	14	U/I	< 18	U/I
α-Amylase	22	U/I	10 - 53	U/I
Lipase	89	U/I	< 190	U/I
Blutzucker	89	mg/dl	55 - 100	mg/dl
HbA1c	5,4	%	4 - 6	%
CK	34	U/I	< 80	U/I
CK-MB	4	U/I	< 10	U/I
CRP	53	mg/l	< 6	mg/l
Ferritin	83	µg/l	15 - 250	µg/l
TSH basal	1,8	µU/ml	0,2 - 3,1	µU/ml
Blutsenkungsgeschwindigkeit	10/23	mm	6-10 / 5-20	mm

Urin-Stix

	Befund	Referenz Frauen
pH	5	5 - 7
Eiweiß	-	-
Bilirubin	-	-
Urobilinogen	-	-
Nitrit	-	-
Glucose	-	-
Aceton	-	-
Blut	-	-

B Measures and Sample Solutions Study 1

B.1 Diagnostic Accuracy

Material is available upon request due to ongoing studies in the project at the time of printing.

B.2 Intrinsic cognitive load

In Study 1, intrinsic cognitive load was assessed using a scale with one item adapted from Opfermann (2008). The 5-point Likert scale ranged from 1 (very easy) to 5 (very difficult).

- Wie leicht oder schwer finden Sie im Moment das Thema „Anforderung einer radiologischen Untersuchung beim Leitsymptom Fieber“

Reference

Opfermann, M. (2008). *There's more to it than instructional design: The role of individual learner characteristics for hypermedia learning*. Berlin: Logos.

B.3 Authenticity

In Study 1, I assessed authenticity by using a 5-point Likert scale with three items adapted from Schubert, Friedmann, & Regenbrecht (2001). The scale ranged from 1 (does not apply) to 5 (does apply). Participants rated authenticity with respect to (1) the overall simulation and (2) the collaboration.

Authenticity with respect to the overall simulation

- Ich schätze die Simulation als authentisch ein.
- Die Simulation hat wie eine echte berufliche Anforderungssituation gewirkt.
- Das Erleben in der Simulation glich dem Erleben in einer realen beruflichen Anforderungssituation.

Authenticity with respect to the collaboration

- Ich schätze die simulierte Zusammenarbeit mit dem Radiologen als authentisch ein.
- Die simulierte Zusammenarbeit mit dem Radiologen hat wie eine echte berufliche Anforderungssituation gewirkt.
- Das Erleben in der simulierten Zusammenarbeit mit dem Radiologen glich dem Erleben in einer realen beruflichen Anforderungssituation.

Reference

Schubert, T., Friedmann, F., & Regenbrecht, H. (2001). The Experience of Presence: Factor Analytic Insights. *Presence: Teleoperators & Virtual Environments, 10*, 266-281.

C Coding Scheme Study 2

C.1 Coding for inclusion

Coding of title and abstract

Independent variable (IV)

Variable: Is the paper about CSCL scripts as defined below?
Name: script

Code	Category
0	No CSCL script or the CSCL script is not varied within a study (e.g. only described in the theory section)
1	CSCL script

Definition of a computer-supported collaboration script:

A collaboration script structures collaborative learning activities by scaffolds (that are named “script” or scaffolds with another name) which are instructional treatments that sequence and distribute at least two different activities among the collaborative learning partners by

(1a) asking or inducing to fulfil specific activities, by explicitly prompting to fulfil the activities (e.g. Rummel et al., 2009) *and/or*

giving a limited amount of activities that can be executed (e.g. by labelling messages; offering different chat-boxes with different headings, Stegmann et al., 2007) *and/or*

inducing distributed activities by asking for an activity that could only be executed in a collaborative way (e.g. discussing, clarifying open questions…)

and/or

(1b) distributing at least two roles* between collaborative learners that explicitly ask for the activities that are connected to the specific roles

and

(2) At least one of the prompted/expected activities must be a contribution that is addressing the opposing learning partner(s) (e.g., showing or explaining something to the learning partner; answering the learning partner’s questions; discussing with the learning partner; etc.). This can be implemented by either explicitly asking to address the learning partner within the script prompt, or by inducing to address the learning partner in the way the learning environment is designed (e.g. discussion forum with script prompts that scaffold how to write messages), or by another implicit demand to address the learning partner (e.g. “clarifying open questions”).

and

(3) Further to be a computer-supported collaboration script the collaborative learning that is supported by the script must be at least partially conducted using computational technique (e.g. learning environment at the computer; receiving script prompts via computer; etc.).

*Note: Collaborative activities that are structured by exclusively distributing roles among learners by distributing knowledge resources (e.g. one learning partner reads an informational text whereas another does not) is not defined as collaboration script.

Data Level

Variable: Is quantitative data provided?
Name: data_level

Code	Category	Explanation
0	No data provided	No study is reported at all (e.g., review, framework…)
1	Qualitative data only	Qualitative data (e.g., content analysis) is provided
2	Quantitative data	Quantitative data (i.e., control versus experimental group) provided

Study Design

Variable: Is an (quasi-) experimental study reported comparing a scripted condition with an unscripted condition?
Name: study_design

Code	Category	Explanation
0	No (quasi-) experimental study reported.	There is either no study reported at all, or it is not a (quasi-)experimental study (correlational only).
1	(quasi-) experimental study reported	At least one treatment group is compared to a control condition. Quasi-experimental allocation is possible.

Language

Variable: Is the article available in English?
Name: Language

Code	Category
0	Not available in English
1	Available in English

Peer reviewed

Variable: Is the article published in a peer reviewed Journal?
Name: peer_review

Code	Category
0	No peer reviewed article
1	Peer reviewed article

Inclusion in fulltext coding

Variable: Should the article be included in the final sample?
Name: inclusion_1

Code	Category	Explanation
0	No	If the remaining variables are coded 0 (resp. 1 for data_level) Exception: data (authors will be contacted before exclusion)
1	Yes	If the remaining variables are coded 1 (resp. 2 for data_level) Exception: data (authors will be contacted before exclusion)

Coding of full text

Post-Test

Variable: When is the outcome variable assessed?
Name: post-test

Code	Category	Explanation
0	Only process data reported	Data is defined as process data if it is assessed during the treatment (i.e., the CSCL script)
1	Post-test data	Post-test data is assessed after the treatment phase (control and experimental group had the same instruction for the post-test)

Dependent Variable

Variable: Is the dependent variable of interest (i.e., domain-specific knowledge, collaboration skill, motivation)?
Name: DV

Code	Category	Explanation
0	DV is not of interest	Other outcome as domain-specific knowledge, collaboration skill or motivation is reported
1	DV is of interest	At least one of domain-specific knowledge, collaboration skill, or motivation is reported

Sufficiency of Data

Variable: Is enough data reported to calculate an effect size?
Name: data

Code	Category	Explanation
0	Not enough data reported	
1	Enough data reported	Mean & SD & N for both groups, or N and F-, t-, Chi^2-, r-statistic reported

Final inclusion of the article

Variable: Should the article be included in the final sample? **Name**: final_inclusion		
Code	**Category**	**Explanation**
0	No	If at least one of the inclusion criteria has the code 0 (resp. 1 for data_level) Exception: data (authors will be contacted before exclusion)
1	Yes	If all inclusion criteria have the code 1 (resp. 2 for data_level) Exception: data (authors will be contacted before exclusion)

C.2 Coding of moderators

Script Activities

Collaboration scripts target different types of collaborative processes and by that foster the social regulation of learning activities. We distinguish between coordination, negotiation, information sharing. Identify all activities prompted/expected by the collaboration script in the experimental condition. Please write down all activities identified in the column *Script_activities* and give the respective type of collaborative activity in brackets (e.g. exchange questions with peers (info), answer questions of your peers (info))

Caution/Hints:

- Sometimes, learners receive support in form of hints (e.g., hints for constructing a good question/argument) that either pop up automatically or can be requested by the learner. These are no activity prompts and should not be coded here. Only, if learners are explicitly requested to do something (e.g., sentence openers), the prompt/hint is regarded as activity.
- Sometimes, it is not clear how often receive specific activity prompts (e.g., sentence openers), or the number of activity prompt differs between learners (e.g., if the prompt appears dependent on the behavior of the group). In this case, the prompt is counted once.
- If a script consists of different types of prompts, then each prompt is counted separately (e.g., in Adamson et al., 2014 Study 4, there are two different agent prompts (agree-disagree and explanation prompts), count each once).
- If an activity is introduced as a higher-level goal, it is not coded as activity prompt here (e.g., “in order to learn the content, students were asked to read the text → reading is coded, learning is not coded)

- If an activity is implicitly mentioned without mentioning the actual verb, the activity is coded anyhow (e.g. from Adamson et al., 2014: "task could not be completed without knowledge from each of the student experts" → implies exchange of knowledge).

Coordination

Activities that are related to the management of group processes, including the negotiation of learning/task goals and group strategies, strategic control of emotional and motivational processes as well as responsibility taking (see Järvelä & Hadwin, 2013; Hesse et al., 2015). Coordination is also if the script makes discussion of learning strategies unnecessary. Examples are:

- discuss learning goal
- discuss learning strategies
- discuss how to proceed
- distribution of roles
- distribution of activity into learning phases / sequencing
- **Not**: Write argument sequence

Variable: Collaborative activities
Name: coord

Code	Category
0	Not prompted
1	Undertaken by the script
2	Prompted

Negotiation

Activities that are related to argumentation and negotiation. These activities aim to resolute conflict, achieving compromise, and to advance knowledge building (e.g., Hesse et al., 2015). Negotiation contains an exchange of arguments. Examples are:

- resolution of conflict
- compromising
- the discussion of content related aspects (not the discussion of collaboration itself),
- formulation of an argument sequence. (e.g.. clarify misunderstandings, discuss,…).
- To agree on something, find a consensus
- To criticize something
- **Not**: to evaluate the general quality of something
- **Not**: construction of single arguments (this is information sharing)

Variable: Collaborative activities **Name**: nego	
Code	**Category**
0	Not prompted
1	Prompted

Information sharing

Activities that are related to the interchange of knowledge or ideas and the construction of shared knowledge, including recognizing differences in the knowledge base between oneself and the learning partner (transactive memory) and to initiate and to request information sharing processes. Information sharing scripts may also prompt to share information in a way that it is suitable for a specific addressee (Hesse et al., 2015). Examples are:

- exchange evidence,
- ask questions
- exchange information (theoretical content, own ideas or solutions of a problem),
- constructing arguments
- **Not**: argument sequences

Variable: Collaborative activities **Name**: info	
Code	**Category**
0	Not prompted
1	Prompted

C.3 Coding of dependent variable

Domain specific knowledge

Domain-specific knowledge refers to content knowledge and skills that are directly related to the targeted domain. For example, in Stegmann et al. (2007), the attribution theory would be the domain-specific knowledge. Domain-specific knowledge is often tested by using knowledge tests (multiple choice or open questions). The tests are either declarative or application oriented knowledge tests.

Coordination skills

Collaboration skills refer to knowledge about and skills related to collaborative processes that are directly prompted, induced and/or structured by the collaboration script (if the skill that is measured is not prompted by the CSCL script and it is not domain-specific knowledge, it is not perceived as relevant). For example, in Stegmann et al. (2007), the collaboration script fosters argumentation skills by directly prompting the argumentative activities “argument”, “counter-argument” and “synthesis”. As different collaborative tasks require different collaboration skills, here it is differentiated between three different types of collaboration skills (Coordination skills, Negotiation skills, Information sharing skills).

Motivation

Motivation comprises all factors that are important for the initiation and maintenance of actions (e.g., Heckhausen, 1974). Motivation is typically assessed using self-report questionnaires.

Variable: Type of the dependent variable
Name: DV_type

Code	Category	Description:
11	Declarative knowledge test	All tests that ask the learner to **recall** or **recognize** the content learnt during the learning phase. Example: a test that asks learners to recall key features of the attribution theory of Weiner.
12	Application test	All tests that ask the learner to **apply** the content learnt during the learning phase. Example: a test that asks the learners to apply the key features of the attribution theory of Weiner to a case scenario.
13	Mixed test	Both types of tasks are tested in the test
21	Coordination skill	Knowledge about and skills related to the management of group processes, including the negotiation of learning/task goals and group strategies, strategic control of emotional and motivational processes as well as responsibility taking
22	Negotiation skill	Knowledge about and skills related to argumentation and negotiation. This includes the skills to resolute conflict and to make compromises, to discuss content related aspects, and to formulate an argument sequence.
23	Information sharing skill	Knowledge about and skills related to the interchange of knowledge and the construction of shared knowledge. This includes recognizing differences in the knowledge base between oneself and the learning partner and to initiate and to request information sharing processes.
24	Mixed	Please leave a comment describing which skills are mixed
3	Motivation	See above

References

Adamson, D., Dyke, G., Jang, H., & Rosé, C. P. (2014). Towards an agile approach to adapting dynamic collaboration support to student needs. *International Journal of Artificial Intelligence in Education, 24*(1), 92–124. https://doi.org/10.1007/s40593-013-0012-6.

Heckhausen, H. (1974). *Leistung und Chancengleichheit [Achievement and equal opportunities]*. Göttingen, Germany: Hogrefe

Hesse, F., Care, E., Buder, J., Sassenberg, K., & Griffin, P. (2015). A framework for teachable collaborative problem solving skills. In P. Griffin & E. Care (Eds.), *Assessment and teaching of 21st century skills* (pp. 37–56). Dordrecht: Springer.

Liu, L., Hao, J., von Davier, A. A., Kyllonen, P., & Zapata-Rivera, D. (2015). A tough nut to crack: Measuring collaborative problem solving. In Y. Rosen, S. Ferrara, & M. Mosharraf (Eds.), *Handbook of research on computational tools for real-world skill development* (pp. 344–359). Hershey, PA: IGI Global.

Järvelä, S., & Hadwin, A. (2013). New Frontiers: Regulating Learning in CSCL. Educational Psychologist, 48, 25–39. doi:10.1080/00461520.2012.748006

Rummel, N., Spada, H., & Hauser, S. (2009). Learning to collaborate while being scripted or by observing a model. *International Journal of Computer-Supported Collaborative Learning, 4*, 69–92. https://doi.org/10.1007/s11412-008-9054-4.

Stegmann, K., Weinberger, A., & Fischer, F. (2007). Facilitating argumentative knowledge construction with computer-supported collaboration scripts. *International Journal of Computer-Supported Collaborative Learning, 2*(4), 421–447. doi: 10.1007/s11412-007-9028-y

D Measures and Sample Solutions Study 3

D.1 Evidence elicitation

Material is available upon request due to ongoing studies in the project at the time of printing.

D.2 Evidence sharing

Material is available upon request due to ongoing studies in the project at the time of printing.

D.3 Basic Psychological Needs

In Study 3, all participants rated their basic psychological need satisfaction using a 7-point Likert scale adapted from Sailer, Hense, Mayr, and Mandl (2017). The scale ranged from 1 (I do not agree) to 7 (I totally agree) and consists of the three subscales perceived competence, perceived autonomy, and perceived social relatedness.

Perceived competence

- Ich bin zufrieden mit meiner Leistung bei der Tätigkeit.
- Ich habe mich während der Tätigkeit kompetent gefühlt.
- Ich habe mich während der Tätigkeit fähig und effektiv gefühlt.
- Ich hatte während der Tätigkeit Erfolgserlebnisse.

Perceived autonomy

- Ich konnte selbst entscheiden, welche Handlungen ich bei der Tätigkeit ausführe.
- Ich habe selbst entscheiden können, was ich während der Tätigkeit mache.
- Ich konnte bei dieser Tätigkeit selbst Entscheidungen treffen.

Perceived social relatedness

- Ich habe mich als Teil eines Teams gefühlt.
- Ich habe mich während der Tätigkeit sozial eingebunden gefühlt.
- Ich habe mich während der Tätigkeit emotional mit anderen verbunden gefühlt.

Reference

Sailer, M., Hense, J. U., Mayr, S. K., & Mandl, H. (2017). How gamification motivates: An experimental study of the effects of specific game design elements on psychological need satisfaction. *Computers in Human Behavior, 69*, 371-380.

E Collaboration Script

Material is available upon request due to ongoing studies in the project at the time of printing.

F Experimenter Guidelines

F.1 Study 1

Material is available upon request due to ongoing studies in the project at the time of printing.

F.1 Study 3

Material is available upon request due to ongoing studies in the project at the time of printing.

G Ethical Approval by Ethics Committee

G.1 Study 1

Ethical clearance was declared by the ethics committee of LMU Klinikum, Ludwig-Maximilians-Universität, prior to data collection.

G.2 Study 3

Ethical clearance was declared by the ethics committee of LMU Klinikum, Ludwig-Maximilians-Universität, prior to data collection.

H Information for Participants and Declaration of Consent

H.1 Study 1

Material is available upon request due to ongoing studies in the project at the time of printing.

H.2 Study 3

Material is available upon request due to ongoing studies in the project at the time of printing.