Making the rich even richer? Interaction of structured reflection with prior knowledge in collaborative medical simulations

Constanze Richters, Matthias Stadler, Anika Radkowitsch
[constanze.richters][matthias.stadler]@psy.lmu.de; Radkowitsch@leibniz-ipn.de
Department of Psychology, LMU Munich
Felix Behrmann, Marc Weidenbusch, Martin R. Fischer
[felix.behrmann][marc.weidenbusch][martin.fischer]@med.uni-muenchen.de
Institute of Medical Education, University Hospital, LMU Munich
Ralf Schmidmaier, Medizinische Klinik und Poliklinik IV, University Hospital, LMU Munich,
ralf.schmidmaier@med.uni-muenchen.de
Frank Fischer, Department of Psychology, LMU Munich, frank.fischer@psy.lmu.de

Abstract: Collaborative diagnostic reasoning refers to knowledge and skills involved in complex individual and collaborative activities with which medical students and physicians often struggle. This study examined the effects of structured reflection and collaboration scripts on collaborative diagnostic reasoning in medical students depending on their prior knowledge level. 151 advanced medical students were asked to diagnose patients with the help of an agent-based radiologist. In the meantime, students received either collaboration scripts, reflection prompts, or both, or no instructional support. The results showed that structured reflection with and without collaboration scripts enhances the collaborative diagnostic skills only of learners with high prior knowledge. These findings cast doubts on the generally positive effects of structured reflection for learning with simulations put forward by prior research.

Introduction and theoretical background

Collaborative diagnostic reasoning plays an important role in everyday medical practice. Physicians must be able to determine malfunctions in human organisms to explain symptoms and implement potential remedies. In doing so, medical experts from different sub-disciplines such as internists and radiologists often work together in teams to identify a patient's problem (e.g., fever of unknown origin). Skills involved in collaborative diagnostic reasoning incorporate cognitively demanding individual diagnostic activities (e.g., generating evidence and hypotheses or drawing conclusions; Fischer et al., 2014) as well as collaborative activities (e.g., sharing of evidence and hypotheses; Radkowitsch et al., 2020). Previous literature shows that medical students and even practitioners struggle in both individual (e.g., Norman & Eva, 2010) and collaborative processes (Tschan et al., 2009). To increase diagnostic quality and reduce diagnostic error rates, it seems promising to understand and foster the skills included in collaborative diagnostic reasoning.

Simulation-based learning environments with embedded instructional support offer a way to foster collaborative diagnostic reasoning skills (e.g., Chernikova, Heitzmann, Fink, et al., 2020; Chernikova, Heitzmann, Stadler, et al., 2020). In the context of collaborative learning, Vogel et al. (2017) suggest that collaboration scripts that are used to help learners to engage in collaborative activities to improve learning outcomes can particularly facilitate learning when they are combined with content-specific scaffolds. In the medical field, structured reflection as content-specific support (i.e., reflecting on suspected diagnoses considering patient information) has been found effective for enhancing individual diagnostic reasoning (e.g., Mamede & Schmidt, 2017). Further, collaboration scripts have been found effective in fostering medical students' collaboration skills (Radkowitsch et al., 2021). Thus, it seems promising that structured reflection and collaboration scripts can facilitate medical students' collaborative diagnostic reasoning skills synergistically. However, recent empirical research raises the question of whether learners with different amounts of prior knowledge benefit similarly from instructional support such as structured reflection and collaboration scripts or whether its effectiveness varies for learners with low and high prior knowledge depending on the provided levels of guidance (Chernikova, Heitzmann, Fink, et al., 2020). Therefore, the present study aims to investigate whether structured reflection and collaboration scripts can enhance the learning of collaborative diagnostic reasoning skills in medicine considering learners' prior knowledge.

Individual and collaborative diagnostic reasoning

Individual diagnostic reasoning can be understood as a targeted collection and interpretation of case-specific information to reduce uncertainty regarding a final accurate diagnosis (Heitzmann et al., 2019). Individual diagnostic reasoning includes knowledge and skills involved in epistemic activities such as generating evidence (case-relevant information) and hypotheses, evaluating hypotheses against the background of evidence, and drawing conclusions (Fischer et al., 2014). In medicine, physicians generate differential diagnoses (hypotheses) based on findings and symptoms and weigh them in light of new evidence until they settle on a final suspected diagnosis. Physicians' success in making diagnoses depends largely on the amount and organization of medical content knowledge (e.g., Boshuizen & Schmidt, 1992). Content knowledge refers to conceptual knowledge including pathophysiological relations and to strategic knowledge (necessary to apply the medical conceptual knowledge; Stark et al., 2011). Content knowledge is theorized to be stored in so-called *illness scripts* (Schmidt & Rikers, 2007) that are schematic mental representations of certain diseases with typical symptoms and findings. With growing medical expertise, conceptual knowledge is being more and more encapsulated so that medical experts no longer retrieve that knowledge consciously (Schmidt & Rikers, 2007). Over time, encountering many clinical cases leads to quick and accurate, yet error-prone, diagnoses (Charlin et al., 2007).

During their *collaborative diagnostic reasoning*, diagnosticians share and elicit evidence and hypotheses (Radkowitsch et al., 2020) to construct and maintain a shared conception of a problem (Roschelle & Teasley, 1995). For example, internists gather new evidence not only by interviewing the patient themselves (e.g., anamnesis) but also through examinations they request from radiologists, e.g., conducting an x-ray of the patient's thorax for suspected pneumonia. Collaboration may therefore lead to higher diagnostic accuracy through knowledge pooling of the collaborators (OECD, 2017). However, previous studies have shown that physicians often struggle to share relevant patient information (Tschan et al., 2009). Whether information is shared adequately depends largely on meta-knowledge on information about the collaborators' knowledge bases, roles, and tasks (Engelmann & Hesse, 2011). For example, an internist needs to know what information about the patient is relevant to the radiologist to conduct an examination (e.g., for examinations with radiation exposure, whether a female patient is pregnant). Such meta-knowledge enables collaborators to anticipate and assess their counterpart's behavior and adapt to it.

Facilitating collaborative diagnostic reasoning

To develop complex skills, the application of learners' knowledge to realistic case scenarios is considered to be crucial (Kolodner, 1992). Simulations enable learners to apply their knowledge and try out some of their future professional activities in lower-stakes standardized settings that are characterized by less complexity than real-world situations (Siebeck et al., 2011). A recent approach to foster collaboration skills is agent-based collaboration, in which learners solve problems collaboratively with a computer agent (Graesser et al., 2018). This approach allows for standardizing the collaboration by, for example, pre-determining the knowledge and skill level of the collaboration partner. Simulation-based learning environments have become increasingly important for the development of diagnostic reasoning skills (e.g., Chernikova, Heitzmann, Stadler, et al., 2020) further, they are particularly effective when additional instructional support is embedded. Since unsupported problem-solving is likely to overwhelm learners, particularly during the early phases of skill development, (e.g., Belland et al., 2017), additional instructional support seems beneficial to foster collaborative diagnostic reasoning.

As instructional support, deliberate reflection phases are promising to advance individual diagnostic reasoning in higher education (Chernikova, Heitzmann, Fink, et al., 2020). Reflection in the broadest sense refers to the process of analyzing and making judgments about what has happened. Learners are aware of and control their learning by assessing what they know, what they need to know, and how they bridge that gap (Dewey, 1933). In the medical field, particularly, a structured reflection that is used early in the diagnostic process can be a powerful tool to increase diagnostic quality and reduce diagnostic errors (Mamede & Schmidt, 2017; cf. Braun et al., 2019). Structured reflection encourages learners to compare signs, symptoms, and findings of the presented patient case with existing illness scripts of certain suspected diagnoses (Mamede et al., 2014). Mamede and Schmidt (2017) argue that through guided reflection, learners can be encouraged to interrupt their more intuitive diagnostic reasoning to use more controlled reasoning processes to identify flaws. However, since reflection requires high self-regulation skills it may be more appropriate for later stages of skill development when learners already have more prior content knowledge and practical experience available (Chernikova, Heitzmann, Fink, et al., 2020; *Matthew effect*, Walberg & Tsai, 1983). External collaboration scripts as a form of socio-cognitive scaffolding can structure and enhance collaborative processes by supporting collaborative activities (Vogel et al., 2017; Radkowitsch et al., 2021) and helping learners build important functional script components internally (*internal collaboration scripts*; Fischer et al., 2013). Based on the optimal scripting level principle (Fischer et al., 2013), collaboration scripts may be more promising for learning if they avoid instructing collaboration at a level where learners already have the required sub-skills. Adapting collaboration scripts to the learners' needs seems therefore promising. *Adaptivity* refers to the "ability of a learning system to diagnose a range of learner variables, and to accommodate a learner's specific needs by making appropriate adjustments to the learner's experience with the goal of enhancing learning outcomes" (Plass & Pawar, 2020, p. 276). Previous studies have shown, that adaptive support through collaboration scripts is promising in terms of knowledge and skill acquisition (e.g., Karakostas & Demetriadis, 2011). Beyond adapting the collaboration script to learners' current performance, learners' prior meta-knowledge could form an adaption basis. Meta-knowledge about the collaboration partner is positively related to internal collaboration scripts that may make an external collaboration script even a hindrance to learning (*expertise reversal effect*; Kalyuga et al., 2003). Instead, due to a high level of guidance, collaboration scripts may be more beneficial for learners at earlier stages of skill development when sufficient meta-knowledge is still lacking.

Moreover, existing research indicates that additionally employed content-specific scaffolds such as structured reflection as suggested by Mamede et al. (2014) may improve the benefit of an external collaboration script for learners (Vogel et al., 2017). Content-specific scaffolding may pre-structure the learning material so that learners are better able to collaboratively examine it in the manner specified by the collaboration script. Thus, in the context of collaborative diagnostic reasoning, learners might also be better able to understand and implement elements of the collaboration script through early reflection on suspected diagnoses. However, assuming that the effectiveness of structured reflection and collaboration scripts varies depending on learners´ prior knowledge, it remains an open question what role learners' prior knowledge plays in such a synergistic effect.

Research questions

RQ1: *Can structured reflection and adaptive collaboration scripts advance medical students' collaborative diagnostic reasoning (evidence sharing, hypotheses sharing, diagnostic quality) in a simulation?* We hypothesize that structured reflection (H1) and adaptive collaboration scripts (H2) positively affect collaborative diagnostic reasoning. Further, we expect that structured reflection and adaptive collaboration scripts positively interact with respect to collaborative diagnostic reasoning (i.e., we expect a positive effect beyond the additive effects of the two forms of instructional support (H3).

RQ2: *What roles do learners' prior content- and meta-knowledge play in the effects of structured reflection and adaptive collaboration scripts on collaborative diagnostic reasoning (evidence sharing, hypotheses sharing, diagnostic quality)?* We hypothesize that learners with high prior content knowledge benefit more from structured reflection than learners with low prior content knowledge (H4). In contrast, learners with low prior meta-knowledge benefit more from adaptive collaboration scripts than learners with high prior meta-knowledge (H5). The interaction between structured reflection and adaptive collaboration scripts is moderated by learners' prior content and meta-knowledge (H6).

Methods

We conducted an experiment with a 2x2-factorial design with reflection phases (levels: present, not present) and collaboration scripts (levels: present, not present). We recruited medical students from their 5th academic year and above from a 6-year medical study program and randomly assigned them to the four groups. In total, 151 medical students (N_{female} = 109) participated in the study. They have been studying for an average of 5.31 years of medical school (*SD* = 0.76) and were on average 25 years old (*SD* = 2.5). Participants received 30€ as compensation for participation.

The simulation-based learning environment and the learners' task

The participants were asked to solve five fictional but realistic patient cases with the leading symptom of fever in collaboration with an agent-based radiologist from whom they requested an examination to gain further evidence for the patient cases and reduce uncertainty regarding suspected diagnoses. The patient cases were embedded in a simulation-based learning environment developed to facilitate and assess collaborative diagnostic reasoning skills (see Radkowitsch et al., 2020). Every patient case was structured as the following:

At first, participants received a health record (an electronic folder that contains information about the respective patient admission, medical history, laboratory results, etc.). After studying the health record, participants filled out a request form for the radiologic examination, which closely resembles the way how requests for radiologic examinations are addressed in clinical practice. To justify the examination, participants provided the radiologist with information from the health record and suspected diagnoses they selected from a long menu (i.e., a list containing 249 different diagnoses that were more or less relevant to the case at hand). Provided that participants sufficiently justified their requests the radiologist performed the examination and shared her medical evaluation. Participants could use up to ten request forms. Finally, participants closed the case by selecting their final diagnosis from the same long menu described above and by justifying this diagnosis in a free-text field.

Experimental conditions and procedure

While working on the patient cases, participants in the experimental conditions (structured reflection, adaptive collaboration scripts, or both) received instructional support according to their assigned condition.

Structured reflection. Following Mamede et al. (2014), we used questions designed to stimulate participants' reflection on current suspected diagnoses by linking symptoms and findings from the health record to suspected diagnoses to reduce uncertainty. Participants received these questions after they reviewed the patient's health record individually and before interacting with the radiologist. After learners reflected on up to five suspected diagnoses, they were asked to sort these diagnoses by probability.

Adaptive collaboration script. Participants received three types of prompts containing meta-knowledge (i.e., information about the radiologist's role, task, and responsibilities). The first prompt, presented at the beginning of the interaction with the agent-based radiologist, included general, case-independent details on the radiologist's task and information that is helpful for them to complete the task. The second and the third prompt provided case-specific information about how radiologists generate evidence for specific suspected diagnoses and about which information helps radiologists to judge the risk of a specific radiologic test. These prompts were presented whenever participants did not sufficiently justify their radiologic request.

Procedure. The data collection was divided into two parts: The first part consisted of an online survey where participants answered single-choice and key-feature questions on content knowledge as well as on meta-knowledge about the collaboration partner (radiologist). After completing the test, participants were directed to the simulation to start the actual experiment (second part). At first, all participants worked on one patient case without additional instructional support (pretest case). Afterward, participants were asked to solve three more patient cases with instructional support according to their group membership (learning cases). Finally, all participants worked on another patient case without additional support (posttest case).

Measures

To map different sub-skills of collaborative diagnostic reasoning skills, we captured *diagnostic quality* as referring to the outcome of diagnostic reasoning and the quality of *evidence sharing* and *hypotheses sharing* as referring to the collaborative diagnostic reasoning process. All variables were operationalized in collaboration with medical experts and calculated for each patient case by the educational researchers among the authors.

Evidence sharing was calculated by the proportion of the shared relevant evidence (information selected by participants and submitted to the radiologist) in all relevant shareable evidence that medical experts have been previously defined depending on the selected diagnoses (long menu). We calculated one score per request indicating the quality of evidence sharing during this request. Subsequently, we calculated the mean score from all requests resulting in a range from 0 points indicating no relevant evidence was shared to 1 point indicating all relevant evidence was shared (Cronbach's $\alpha = 0.82$).

Hypotheses sharing was calculated by the proportion of all shared relevant hypotheses that were previously defined by medical experts in all shared hypotheses (diagnoses selected from the long menu by participants, and submitted to the radiologist) resulting in a range from 0 points indicating no relevant hypotheses were shared to 1 point all shared hypotheses were relevant (Cronbach's $\alpha = 0.51$).

Diagnostic quality was assessed by the correctness of the final diagnosis and the quality of its justification based on a coding manual defined by medical experts. Provided participants submitted the correct final diagnosis, we calculated the proportion of relevant information mentioned in the free-text field in all relevant information that would have fully justified the final diagnosis according to the medical experts. In case of an incorrect final diagnosis, we assigned 0 points for the diagnostic quality. We obtained a range from 0 points indicating an incorrect final diagnosis to 1 point indicating a correct and properly justified final diagnosis (Cronbach's $\alpha = 0.70$).

Prior content knowledge was operationalized by 35 single-choice and 37 key-feature (M. R. Fischer et al., 2005) questions referring to internal medicine and radiology. For the final score, we calculated mean values across all questions resulting in a range from 0 points to 1 point indicating participants´ prior content knowledge level (Cronbach's $\alpha = 0.74$).

Prior meta-knowledge was measured by seven text-based patient cases with symptoms that required a radiologic examination as the next step of the diagnostic work-up. All cases contained information on the patient (from an emergency medical service note) as well as on the predetermined planned radiologic examination. Each case had eleven items that are related to information from which to choose which would be shared with the radiologist. We calculated mean values across all items per case and subsequently across all cases resulting in a range from 0 points to 1 point indicating participants´ prior meta-knowledge level (Cronbach's $\alpha = 0.70$).

Analyses

We computed a multiple linear regression model for each of the collaborative diagnostic reasoning sub-skills (evidence sharing, hypotheses sharing, diagnostic quality). In each model, we used the score from the posttest case as the dependent variable and the score from the pretest case as a covariate to test the advancement (transfer performance). We used prior content and meta-knowledge as moderators, as well as reflection and collaboration script as independent variables. According to our hypotheses, we included two-way interaction terms (reflection x collaboration script, prior content knowledge x reflection, prior meta-knowledge x collaboration script) and two three-way interaction terms (prior content knowledge x reflection x collaboration script, prior meta-knowledge x reflection x collaboration script) in each model. To control for all two-way interactions for the assumed three-way interactions, we included two more interaction terms (reflection x meta-knowledge, content knowledge x collaboration script). All continuous variables (covariates and dependent variables) have been z-standardized to simplify the interpretation of the results.

Results

To answer RQ 1, we tested at first whether structured reflection positively affects collaborative diagnostic reasoning (H1). Results of the linear regression models revealed that structured reflection neither significantly affect the quality of evidence sharing ($\beta = .06$, $p = .777$) nor the quality of hypotheses sharing ($\beta = .10$, $p = .674$) nor overall diagnostic quality ($\beta = .34$, $p = .104$). These results are thus not in support of H1.

Further, we tested whether adaptive collaboration scripts positively affect collaborative diagnostic reasoning (H2). The results showed that collaboration scripts neither significantly affect the quality of evidence sharing ($\beta = .13$, $p = .517$) nor the quality of hypotheses sharing ($\beta = -.03$, $p = .923$) nor overall diagnostic quality ($\beta = .05$, $p = .810$). Thus, these findings are not in line with H2.

Moreover, we tested whether structured reflection and adaptive collaboration scripts positively interact with respect to collaborative diagnostic reasoning (H3). The analyses revealed that the interaction of structured reflection and collaboration scripts neither significantly affect the quality of evidence sharing ($\beta = .19$, $p = .493$) nor the quality of hypotheses sharing ($\beta = -.04$, $p = .910$) nor overall diagnostic quality ($\beta = -.18$, $p = .555$). We thus did not find evidence for H3.

To answer RQ 2, we tested at first whether learners with high prior content knowledge benefit more from structured reflection than learners with low prior content knowledge (H4). In contrast to our hypothesis, the analyses revealed that the interaction between structured reflection and prior content knowledge did not significantly affect the quality of evidence sharing ($\beta = .06$, $p = .762$). However, there was a significant interaction effect on the quality of hypotheses sharing ($\beta = .56$, $p = .016$). Comparisons of the estimated marginal means revealed that learners with low content knowledge who reflected on their individual diagnostic process hardly differed across groups. However, learners with high content knowledge scored higher when they reflected on their diagnostic process ($M = 0.14$, $SE = 0.17$) compared to when they did not ($M = -0.26$, $SE = 0.17$). The analysis revealed no significant effect on the overall diagnostic quality ($\beta = -.06$, $p = .758$). Thus, these findings partially support H4.

To test whether learners with low prior meta-knowledge benefit more from adaptive collaboration scripts than learners with high prior meta-knowledge (H5) we took a closer look at the interaction between collaboration scripts and meta-knowledge. The analyses revealed that the interaction of collaboration scripts and meta-knowledge neither significantly affected the quality of evidence sharing ($\beta = -.25$, $p = .220$) nor the quality of hypotheses sharing ($\beta = .18$, $p = .466$) nor the overall diagnostic quality ($\beta = -.18$, $p = .555$). Thus, these results do not support H5.

Finally, we tested whether the interaction between structured reflection and adaptive collaboration scripts is moderated by learners' prior content and meta-knowledge (H6). The analyses revealed a significant interaction effect of structured reflection, collaboration scripts and prior content knowledge on the quality of evidence sharing ($\beta = .67$, $p = .020$) whereas the three-way interaction between structured reflection, collaboration scripts and prior meta-knowledge did not significantly affect the quality of evidence sharing ($\beta = .41$, $p = .147$). Comparisons of the estimated means revealed that learners with low prior content knowledge scored higher when they exclusively learned with collaboration scripts ($M = 0.29$, $SE = 0.19$) compared to when they only reflected on their diagnostic process ($M = -0.29$, $SE = 0.18$), additionally reflected on their diagnostic process ($M = -0.18$, $SE = 0.23$) or when they received no additional support ($M = -0.23$, $SE = 0.20$). In contrast, learners with high prior content knowledge scored better when they reflected on their individual diagnostic process and additionally received collaboration scripts ($M = 0.68$, $SE = 0.21$) compared to when they only reflected on their diagnostic process ($M = 0.09$, $SE = 0.19$), only received collaboration scripts ($M = -0.28$, $SE = 0.19$) or when they received no additional support ($M = -0.03$, $SE = 0.20$).

Moreover, neither the three-way interaction with prior content knowledge ($\beta = -.48$, $p = .162$) nor the three-way interaction with prior meta-knowledge had a significant effect on the quality of hypotheses sharing ($\beta = -.37$, p

= .275). The same applied to the overall diagnostic quality. The analyses neither revealed significant results for the interaction with prior content knowledge ($\beta = .19$, $p = .528$) nor for prior meta-knowledge ($\beta = -.26$, $p = .399$). Thus, these findings partially support H6.

Discussion

In this study, we examined to what extent structured reflection and collaboration scripts enhance the learning of collaborative diagnostic reasoning skills in medicine considering prior medical content and meta-knowledge. Overall, our analyses revealed no significant main effects per se of structured reflection unlike those of previous work (cf. Mamede et al., 2014). Further, we neither found significant main effects of collaboration scripts (cf. Radkowitsch et al., 2021) nor interaction effects with prior meta-knowledge that would indicate a previously assumed expertise reversal effect (Kalyuga et al., 2003). Moreover, we found no synergistic effect between structured reflection and collaboration scripts (cf. Vogel et al., 2017).

However, as hypothesized, structured reflection advanced collaborative diagnostic reasoning skills of learners with high prior content knowledge. More precisely, structured reflection advanced learners' performance in hypotheses sharing. Learners with high prior content knowledge were presumably better able to relate suspected diagnoses to information from the case when reflecting (Mamede et al., 2014), resulting in them sharing less irrelevant suspected hypotheses. This interaction effect can be considered to be an instance of the Matthew effect (Walberg & Tsai, 1983) and is in line with findings from a recent meta-analysis which suggests that reflection is rather appropriate for later stages of skill development respectively for learners with high prior knowledge (Chernikova, Heitzmann, Fink, et al., 2020).

Moreover, in contrast to our assumption, learners with low prior meta-knowledge about the collaborators' role, task, and responsibilities did not benefit more from collaboration scripts than learners with high prior meta-knowledge. This could be due to the adaptivity of the collaboration script. More precisely, two of the three prompts were only presented whenever learners' requests were rejected by the radiologists, and rejection rates could be negatively associated with learners' meta-knowledge. Learners whose requests were less often rejected might have more meta-knowledge and thus received the collaboration script less often than learners who were rejected more frequently and who might have less meta-knowledge. Thus, the script might be already adapted to learners' prior meta-knowledge and no longer interacts with it. However, the first prompt was displayed independently of rejection and, further, rejections may not exclusively depend on meta-knowledge but other factors such as an ineffective individual diagnostic process. Future studies that address conditions under which learners get rejected when working with the radiologist would be helpful here.

Further, concerning the performance in evidence sharing, learners with high prior content knowledge benefited from structured reflection referring to the individual process combined with collaboration scripts referring to the collaborative process whereas learners with low prior content knowledge benefited from exclusively learning with collaboration scripts. Following the previous assumption that learners with high prior content knowledge benefit more from structured reflection (Chernikova, Heitzmann, Fink, et al., 2020) it seems possible that only these learners were able to pre-structure the diagnostic process so that they could implement the elements of the collaboration script (Vogel et al., 2017) that aimed at enhancing the performance in evidence sharing. Through early successful reflection on suspected diagnoses by identifying case-relevant information as evidence and linking this evidence to suspected diagnoses, high content knowledge learners may have pre-structured the evidence for the interaction with the radiologist. Subsequently, the collaboration script might have helped them to select, among this evidence, relevant evidence for the radiologist. On the other hand, learners with low prior content knowledge appeared to be cognitively overwhelmed by the combination of both forms of instructional support (Eckhardt et al., 2013 and were, instead, able to improve their performance in evidence sharing when only learning with collaboration scripts that offer higher levels of guidance. This suggests that a sufficient level of prior knowledge is not only a prerequisite for successful reflection but also a synergistic effect of both forms of instructional support.

Limitations

In interpreting our results, it is important to consider that the sample size of our study was rather small. This could be a reason for the not reproducible main effect of the collaboration script. Moreover, the lack of effects of the collaboration script on hypotheses sharing and the overall diagnostic quality could be due to its design, which rather addressed improving the performance in evidence sharing than in hypotheses sharing.

Further, by measuring the quality of evidence sharing, we did not consider whether learners shared all the available information or whether they indeed selected information carefully. However, we assume that the lack of important information in medical practice is more critical than additionally shared irrelevant information (Tschan et al., 2009). In the future, we plan to include a precision dimension to control for this factor. By measuring the performance in hypotheses sharing, on the other hand, we did not consider how many relevant hypotheses learners shared in total. However, we assume in medical practice that it might be more important to share as few as possible irrelevant suspected diagnoses than as many as possible relevant suspected diagnoses. We plan to consider how sensitive learners were in sharing hypotheses in the future.

Finally, we want to emphasize that the pattern of findings might at least in part may be caused by the replacement of real collaboration with an agent-based collaboration. However, researchers from the field of the learning sciences and medical experts designed the simulation´s interface (request form) and structure authentically based on real clinical situations. An interesting future line of research would be investigating the transfer to diagnostic settings in which two humans work together.

Conclusion and implication

Structured reflection with and without collaboration scripts can advance high prior but not low prior knowledge learners´ collaborative diagnostic reasoning skills in medical simulations. Learners with high prior knowledge who reflect on their individual diagnostic reasoning process perform better in hypotheses sharing. Collaboration scripts can additionally help these learners to review their generated and evaluated evidence from the reflective process with respect to its relevance to the collaboration partner. Instead, learners with low prior knowledge benefit at least to some extent from exclusively learning with collaboration scripts. These findings support recent meta-analytic evidence that the effectiveness of instructional support with different levels of guidance respectively self-regulation requirements varies for learners with different amounts of prior knowledge (Chernikova, Heitzmann, Fink, et al., 2020). Moreover, our results cast doubts on the generally positive effects of structured reflection for learning through problem-solving put forward by prior research (e.g., Mamede & Schmidt, 2017).

Future research may mainly follow this line and address *micro-* or *macro-adaptivity* (Plass & Pawar, 2020) of reflection phases by for example exploring different reflection designs with varying degrees of structure on learners with various levels of prior knowledge. Higher levels of structure or guidance in externally stimulated reflection might be helpful for learners at lower stages of learning. Moreover, conditions under which performance-adapted collaboration scripts advance collaborative diagnostic reasoning could be investigated further as well by exploring reasons why learners with low and high prior knowledge fail in collaboration to identify additional collaborative aspects that could be supported by collaboration scripts.

Overall, the present study sets a first step in investigating conditions under which stimulated reflection has a positive impact on medical students' collaborative diagnostic reasoning in the context of simulations. An important conclusion is that structured reflection in collaborative simulations may be particularly suitable for learners with high prior knowledge.

References

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

Boshuizen, H. P. A., & Schmidt, H. G. (1992). On the role of biomedical knowledge in clinical reasoning by experts, intermediates and novices. *Cognitive Science*, *16*(2), 153–184.

Braun, L. T., Borrmann, K. F., Lottspeich, C., Heinrich, D. A., Kiesewetter, J., Fischer, M. R., & Schmidmaier, R. (2019). Scaffolding clinical reasoning of medical students with virtual patients: Effects on diagnostic accuracy, efficiency, and errors. *Diagnosis*, *6*(2), 137–149.

Charlin, B., Boshuizen, H. P. A., Custers, E. J., & Feltovich, P. J. (2007). Scripts and clinical reasoning: Clinical expertise. *Medical Education, 41*(12), 1178–1184.

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*(1), 157–196.

Chernikova, O., Heitzmann, N., Stadler, M., Holzberger, D., Seidel, T., & Fischer, F. (2020). Simulation-based learning in higher education: A meta-analysis. *Review of Educational Research*, *90*(4), 499–541.

Dewey, J. (1933). *How we think: A restatement of reflective thinking to the educative process.* D. C. Heath and company.

Eckhardt, M., Urhahne, D., Conrad, O., & Harms, U. (2013). How effective is instructional support for learning with computer simulations? *Instructional Science*, *41*(1), 105–124.

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*(6), 2078–2087.

Fischer, F., Kollar, I., Stegmann, K., & Wecker, C. (2013). Toward a script theory of guidance in computer-supported collaborative learning. *Educational Psychologist*, *48*(1), 56–66.

Fischer, F., Kollar, I., Ufer, S., & Eberle, J. (2014). Scientific reasoning and argumentation: Advancing an interdisciplinary research agenda in education. *Frontline Learning Research*, *2*(3), 28-45.
Fischer, M. R., Kopp, V., Holzer, M., Ruderich, F., Jünger, J. (2005) A modified electronic key feature examination for undergraduate medical students: validation threats and opportunities. *Medical Teacher*, *27*(5), 450-455.
Graesser, A., Fiore, S., Greiff, S., Andrews-Todd, J., Foltz, P., & Hesse, F. (2018). Advancing the science of collaborative problem solving. *Psychological Science in the Public Interest*, *19*(2), 59–92.
Heitzmann, N., Seidel, T., Hetmanek, A., Wecker, C., Fischer, M. R., Ufer, S., Schmidmaier, R., Neuhaus, B., Siebeck, M., Stürmer, K., Obersteiner, A., Reiss, K., Girwidz, R., Fischer, F., & Opitz, A. (2019). Facilitating diagnostic competences in simulations in higher education: A conceptual framework and a research agenda. *Frontline Learning Research*, *7*(4), 1–24.
Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist*, *38*(1), 23–31.
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.
Kolodner, J. L. (1992). An introduction to case-based reasoning. *Artificial Intelligence Review, 6*(1), 3–34.
Mamede, S., & Schmidt, H. G. (2017). Reflection in medical diagnosis: A literature review. *Health Professions Education*, *3*(1), 15–25.
Mamede, S., van Gog, T., Sampaio, A. M., de Faria, R. M. D., Maria, J. P., & Schmidt, H. G. (2014). How can students' diagnostic competence benefit most from practice with clinical cases? The effects of structured reflection on future diagnosis of the same and novel diseases: *Academic Medicine*, *89*(1), 121–127.
Norman, G. R., & Eva, K. W. (2010). Diagnostic error and clinical reasoning: Diagnostic error and reasoning. *Medical Education*, *44*(1), 94–100.
OECD (2017). *PISA 2015 Assessment and analytical framework: Science, reading, mathematic, financial literacy and collaborative problem solving.* Paris: France. OECD Publishing. https://doi.org/10.1787/9789264281820-en
Plass, J. L., & Pawar, S. (2020). Toward a taxonomy of adaptivity for learning. *Journal of Research on Technology in Education*, *52*(3), 275–300.
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. https://doi.org/10.3205/ZMA001344
Radkowitsch, A., Sailer, M., Schmidmaier, R., Fischer, M. R., & Fischer, F. (2021). Learning to diagnose collaboratively – Effects of adaptive collaboration scripts in agent-based medical simulations. *Learning and Instruction*, *75*, 101487.
Roschelle, J., & Teasley, S. D. (1995). The construction of shared knowledge in collaborative problem solving. In C. O'Malley (Ed.), *Computer supported collaborative learning* (pp. 69–97). Springer.
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.
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: Learning with simulation. *Medical Education*, *45*(10), 1025–1031.
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.
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.
Vogel, F., Wecker, C., Kollar, I., & Fischer, F. (2017). Socio-cognitive scaffolding with computer-supported collaboration scripts: A meta-analysis. *Educational Psychology Review*, *29*(3), 477–511.
Walberg, H. J., & Tsai, S.-L. (1983). "Matthew" effects in education. *American Educational Research Journal*, *20*(3), 359-373.

Acknowledgments

This research was funded by a grant from the Deutsche Forschungsgemeinschaft DFG (COSIMA, FOR 2385) for project 6 CoSiMed (FI 792/11-1).