



Collaborative Problem-Solving in Knowledge-Rich Domains: A Multi-Study Structural Equation Model

Laura Brandl¹ · Matthias Stadler^{1,2} · Constanze Richters¹ · Anika Radkowsitsch³ · Martin R. Fischer² · Ralf Schmidmaier⁴ · Frank Fischer¹

Received: 18 September 2023 / Accepted: 9 May 2024 / Published online: 24 June 2024
© The Author(s) 2024, corrected publication 2024

Abstract

Collaborative skills are crucial in knowledge-rich domains, such as medical diagnosing. The Collaborative Diagnostic Reasoning (CDR) model emphasizes the importance of high-quality collaborative diagnostic activities (CDAs; e.g., evidence elicitation and sharing), influenced by content and collaboration knowledge as well as more general social skills, to achieve accurate, justified, and efficient diagnostic outcomes (Radkowsitsch et al., 2022). However, it has not yet been empirically tested, and the relationships between individual characteristics, CDAs, and diagnostic outcomes remain largely unexplored. The aim of this study was to test the CDR model by analyzing data from three studies in a simulation-based environment and to better understand the construct and the processes involved ($N = 504$ intermediate medical students) using a structural equation model including indirect effects. We found various stable relationships between individual characteristics and CDAs, and between CDAs and diagnostic outcome, highlighting the multidimensional nature of CDR. While both content and collaboration knowledge were important for CDAs, none of the individual characteristics directly related to diagnostic outcome. The study suggests that CDAs are important factors in achieving successful diagnoses in collaborative contexts, particularly in simulation-based settings. CDAs are influenced by content and collaboration knowledge, highlighting the importance of understanding collaboration partners' knowledge. We propose revising the CDR model by assigning higher priority to collaboration knowledge compared with social skills, and dividing the CDAs into information elicitation and sharing, with sharing being more transactive. Training should focus on the development of CDAs to improve CDR skills.

Keywords Collaborative Problem-solving · Simulation-based Learning Environment · Diagnostic Activities · Diagnostic Reasoning · Medical Education

Introduction

Collaborative skills are highly relevant in many situations, ranging from computer-supported collaborative learning to collaborative problem-solving in professional practice (Fiore et al., 2018). While several broad collaborative problem-solving frameworks exist

(OECD, 2017), most of them are situated in knowledge-lean settings. However, one example of collaborative problem-solving of knowledge-rich domains is collaborative diagnostic reasoning (CDR; Radkowsch et al., 2022)—which aligns closely with medical practice—as this is a prototypical knowledge-rich domain requiring high collaboration skills in daily practice. In daily professional practice, physicians from different specialties often need to collaborate with different subdisciplines to solve complex problems, such as diagnosing, that is, determining the causes of a patient's problem. Moreover, research in medical education and computer-supported collaborative learning suggests that the acquisition of medical knowledge and skills is significantly enhanced by collaborative problem-solving (Hautz et al., 2015; Koschmann et al., 1992). For problem-solving and learning, it is crucial that all relevant information (e.g., evidence and hypotheses) is elicited from and shared with the collaboration partner (Schmidt & Mamede, 2015). However, CDR is not unique to the medical field but also relevant in other domains, such as teacher education (Heitzmann et al., 2019).

The CDR model has been the basis of empirical studies and describes how individual characteristics and the diagnostic process are related to the diagnostic outcome. However, it has not yet been empirically tested, and the relationships between individual characteristics, diagnostic process, and diagnostic outcome remain mostly unexplored (Fink et al., 2023). The aim of this study is to test the CDR model by analyzing data from three studies with similar samples and tasks investigating CDR in a simulation-based environment. By undertaking these conceptual replications, we aspire to better understand the construct and the processes involved. As prior research has shown, collaboration needs to be performed at a high quality to achieve accurate problem solutions respectively learning outcomes (Pickal et al., 2023).

Collaborative Diagnostic Reasoning (CDR) Model

Diagnosing can be understood as the process of solving complex diagnostic problems through “goal-oriented collection and interpretation of case-specific or problem-specific information to reduce uncertainty” in decision-making through performing diagnostic activities at a high quality (Heitzmann et al., 2019, p. 4). To solve diagnostic problems, that is, to identify the causes of an undesired state, it is increasingly important to collaborate with experts from different fields, as these problems become too complex to be solved individually (Abele, 2018; Fiore et al., 2018). Collaboration provides advantages such as the division of labor, access to diverse perspectives and expertise, and enhanced solution quality through collaborative sharing of knowledge and skills (Graesser et al., 2018).

The CDR model is a theoretical model focusing on the diagnostic process in collaborative settings within knowledge-rich domains (Radkowsch et al., 2022). The CDR model is based on scientific discovery as a dual-search model (SDDS; Klahr & Dunbar, 1988) and its further development by van Joolingen and Jong (1997). The SDDS model describes individual reasoning as the coordinated search through hypothetical evidence and hypotheses spaces and indicates that for successful reasoning it is important not only that high-quality cognitive activities within these spaces are performed but also that one is able to coordinate between them (Klahr & Dunbar, 1988). In the extended SDDS model (van Joolingen & Jong, 1997) focusing on learning in knowledge-rich domains, a learner hypothesis space was added including all the hypotheses one can search for without additional knowledge. Although Dunbar (1995) found that conceptual change occurs more often in groups than in individual work, emphasizing the importance of collaborative

processes in scientific thinking and knowledge construction, the SDDS model has hardly been systematically applied in computer-supported collaborative learning and collaborative problem-solving.

Thus, the CDR model builds upon these considerations and describes the relationship between individual characteristics, the diagnostic process, and the diagnostic outcome. As in the SDDS model we assume that CDR involves activities within an evidence and hypotheses space; however, unlike the SDDS in the CDR model, these spaces are understood as cognitive storages of information. Which aligns more to the extended dual search space model of scientific discovery learning (van Joolingen & Jong, 1997). In summary we assume that coordinating between evidence (data) and hypothesis (theory) is essential for successful diagnosing. Further, the CDR model is extended to not only individual but also collaborative cognitive activities and describes the interaction of epistemic activities (F. Fischer et al., 2014) and collaborative activities (Liu et al., 2016) to construct a shared problem representation (Rochelle & Teasley, 1995) and effectively collaborate. Thus, we define CDR as a set of skills for solving a complex problem collaboratively “by generating and evaluating evidence and hypotheses that can be shared with, elicited from, or negotiated among collaborators” (Radkowsitch et al., 2020, p. 2). The CDR model also makes assumptions about the factors necessary for successful CDR. First, we look at what successful CDR means, why people differ, and what the mediating processes are.

Diagnostic Outcome: Accuracy, Justification, and Efficiency

The primary outcome of diagnostic processes, such as CDR, is the accuracy of the given diagnosis, which indicates problem-solving performance or expertise (Boshuizen et al., 2020). However, competence in diagnostic reasoning, whether it is done individually or collaboratively, also includes justifying the diagnosis and reaching it effectively. This is why, in addition to diagnostic accuracy, diagnostic justification and diagnostic efficiency should also be considered as secondary outcomes of the diagnostic reasoning process (Chernikova et al., 2022; Daniel et al., 2019). Diagnostic justification makes the reasoning behind the decision transparent and understandable for others (Bauer et al., 2022). Good reasoning entails a justification including evidence, which supports the reasoning (Hitchcock, 2005). Diagnostic efficiency is related to how much time and effort is needed to reach the correct diagnosis; this is important for CDR, as diagnosticians in practice are usually under time pressure (Braun et al., 2017). Both diagnostic justification and diagnostic efficiency are thus indicators of a structured and high-quality reasoning process. So, while in many studies, the focus of assessments regarding diagnostic reasoning is on the accuracy of the given diagnosis (Daniel et al., 2019), the CDR model considers all three facets of the diagnostic outcome as relevant factors.

Individual Characteristics: Knowledge and Social Skills

Research has shown that content knowledge, social skills, and, in particular, collaboration knowledge are important prerequisites for, and outcomes of, computer-supported collaborative learning (Jeong et al., 2019; Vogel et al., 2017). CDR has integrated these dependencies into its model structure. Thus, the CDR model assumes that people engaging in CDR differ with respect to their content knowledge, collaboration knowledge, and domain general social skills.

Content knowledge refers to conceptual and strategic knowledge in a specific domain (Förtsch et al., 2018). Conceptual knowledge encompasses factual understanding of domain-specific concepts and their interrelations. Strategic knowledge entails contextualized knowledge regarding problem-solving during the diagnostic process (Stark et al., 2011). During expertise development, large amounts of content knowledge are acquired and restructured through experience with problem-solving procedures and routines (Boshuizen et al., 2020). Research has repeatedly shown that having high conceptual and strategic knowledge is associated with the diagnostic outcome (e.g., Kieseewetter et al., 2020; cf. Fink et al., 2023).

In addition to content knowledge, the CDR model assumes that collaborators need collaboration knowledge. A key aspect of collaboration knowledge (i.e., being aware of knowledge distribution in the group; Noroozi et al., 2013) is the pooling and processing of non-shared information, as research shows that a lack of collaboration knowledge has a negative impact on information sharing, which in turn has a negative impact on performance (Stasser & Titus, 1985).

Finally, general social skills influence the CDR process. These skills mainly influence the collaborative aspect of collaborative problem-solving and less the problem-solving aspect (Graesser et al., 2018). Social skills are considered particularly important when collaboration knowledge is low (F. Fischer et al., 2013). CDR assumes that in particular the abilities to share and negotiate ideas, to coordinate, and to take the perspective are relevant for the diagnostic process and the diagnostic outcome (Radkowsch et al., 2022; see also Liu et al., 2016, and Hesse et al., 2015).

Diagnostic Process: Collaborative Diagnostic Activities

The diagnostic process is thought to mediate the effect of the individual characteristics on the diagnostic outcome and is described in the CDR model using collaborative diagnostic activities (CDAs), such as evidence elicitation, evidence sharing, and hypotheses sharing (Heitzmann et al., 2019; Radkowsch et al., 2022). One of the main functions of CDAs is to construct a shared problem representation (Rochelle & Teasley, 1995) by sharing and eliciting relevant information, as information may not be equally distributed among all collaborators initially. To perform these CDAs at a high quality, each collaborator needs to identify information relevant to be shared with the collaboration partner as well as information they need from the collaboration partner (OECD, 2017).

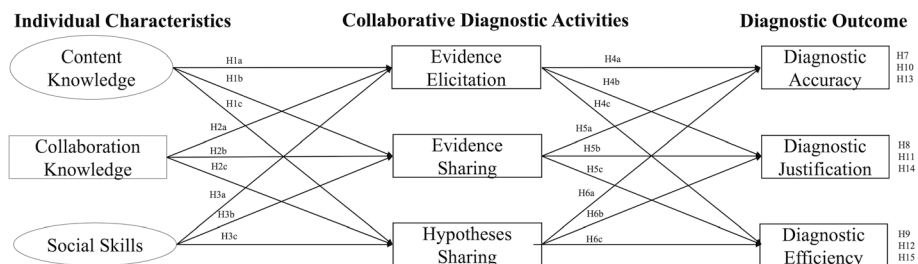
Evidence elicitation involves requesting information from a collaboration partner to access additional knowledge resources (Weinberger & Fischer, 2006). Evidence sharing and hypothesis sharing involve identifying the information needed by the collaborator to build a shared problem representation. This externalization of relevant information can be understood as the novelty aspect of transactivity (Vogel et al., 2023). However, challenges arise from a lack of relevant information due to deficient sharing, which can result from imprecise justification and insufficient clustering of information. In particular, research has shown that collaborators often lack essential information-sharing skills, such as identifying information relevant for others from available data, especially in the medical domain (Kieseewetter et al., 2017; Tschan et al., 2009).

It is crucial for the diagnostic outcome that all relevant evidence and hypotheses are elicited and shared for the specific collaborators (Tschan et al., 2009). However, diagnostic outcomes seem to be influenced more by the relevance and quality of the shared information than by their quantity (Kiesewetter et al., 2017; Tschan et al., 2009). In addition, recent research has shown that the diagnostic process is not only an embodiment of individual characteristics but also adds a unique contribution to diagnostic outcome (Fink et al., 2023). However, it remains difficult to assess and foster CDAs.

Collaboration in Knowledge-Rich Domains: Agent-Based Simulations

There are several challenges when it comes to modelling collaborative settings in knowledge-rich domains for both learning and research endeavors. First, many situations are not easily accessible, as they may be scarce (e.g., natural disasters) or too critical or overwhelming to be approached by novices (e.g., some medical procedures). In these cases, the use of simulation-based environments allows authentic situations approximating real-life diagnostic problems to be provided (Cook et al., 2013; Heitzmann et al., 2019). Further, the use of technology-enhanced simulations allows data from the ongoing CDR process to be collected in log files. This enables researchers to analyze process data without the need for additional assessments with dedicated tests. Analyzing process data instead of only product data (the assessment's outcome) permits insights into the problem-solving processes leading to the eventual outcome (e.g., Goldhammer et al., 2017). Second, when using human-to-human collaboration, the results of one individual are typically influenced by factors such as group composition or motivation of the collaboration partner (Radkowsitch et al., 2022). However, we understand CDR as an individual set of skills enabling collaboration, as indicated by the broader definition of collaborative problem-solving (OECD, 2017). Thus, the use of simulated agents as collaboration partners allows a standardized and controlled setting to be created that would otherwise be hard to establish in collaborations among humans (Rosen, 2015). There is initial research showing that performance in simulations using computerized agents is moderately related to collaborative skills in other operationalizations (Stadler & Herborn et al., 2020). Thus, computerized agents allow for enhanced control over the collaborative process without significantly diverging from human-to-human interaction (Graesser et al., 2018; Herborn et al., 2020). Third, in less controlled settings it is hard to ensure a specific process is taking place during collaborative problem-solving. For example, when using a human-to-human setting, it is possible that, even though we envision measuring or fostering a specific activity (i.e. hypotheses sharing), it is not performed by the student. Through using an agent-based simulated collaboration partner, we can ensure that all required processes are taking place while solving the problem (Rosen, 2015).

Summarizing, by fostering a consistent and controlled setting, simulated agents facilitate the accurate measurement and enhancement of collaborative problem-solving. Evidential support for the application of simulated agents spans a variety of contexts, including tutoring, collaborative learning, knowledge co-construction, and collaborative problem-solving itself, emphasizing their versatility and effectiveness in educational settings (Graesser et al., 2018; Rosen, 2015).



Circles represent latent variables, and rectangles represent manifest variables

Fig. 1 Visualization of hypothesized relationships between individual characteristics, collaborative diagnostic activities, and diagnostic outcome

Research Question and Current Study

In computer-supported collaborative learning there has been the distinction between approaches addressing collaboration to learn and approaches focusing on learning to collaborate. Our study is best understood as addressing the second approach, learning to collaborate. We want to better understand CDR to be able to facilitate collaborative problem-solving skills in learners. Thus, in this paper, we examine what it takes to be able to collaborate in professional practice of knowledge-rich domains, such as medical diagnosing.

When solving diagnostic problems, such as diagnosing a patient, it is often necessary to collaborate with experts from different fields (Radkowsch et al., 2022). In CDR, the diagnostic outcome depends on effectively eliciting and sharing relevant evidence and hypotheses among collaborators, who often lack information-sharing skills (Tschan et al., 2009). Thus, the CDR model emphasizes the importance of high-quality CDAs influenced by content and collaboration knowledge as well as social skills to achieve accurate, justified, and efficient diagnostic outcomes (Radkowsch et al., 2022).

This study reviews the relationships postulated in CDR model across three studies to test them empirically and investigate the extent to which the relationships in the CDR model are applicable across studies. By addressing this research question, the current study contributes to a better understanding of the underlying processes in collaborative problem-solving.

We derived a model (Fig. 1) from the postulated relationships made by the CDR model. We assume that the individual characteristics are positively related to the CDAs (Hypotheses 1–3), as well as that the CDAs are positively related to the diagnostic outcome (Hypotheses 4–6). Further, we expect that the relationship between the individual characteristics and the diagnostic outcome is partially mediated by the CDAs (Hypotheses 7–15).

We used data from three studies with similar samples and tasks investigating CDR in an agent-based simulation in the medical domain. The studies can therefore be considered conceptual replication studies. Furthermore, we decided to use an agent-based simulation of a typical collaboration setting in diagnostic reasoning, namely the interdisciplinary collaboration between an internist and a radiologist (Radkowsch et al., 2022).

Table 1 Sample description per study

Study	N	Gender	Age	Year of study
Study A	157	Male = 49 Female = 108	$M = 25.1$ $SD = 3.1$	$M = 5.3$ $SD = 0.9$
Study B	155	Male = 44 Female = 111	$M = 25.3$ $SD = 3.0$	$M = 5.4$ $SD = 0.8$
Study C	192	Male = 62 Female = 130	$M = 23.3$ $SD = 3.4$	$M = 3.2$ $SD = 0.4$

N = sample size, M = mean, SD = standard deviation

Methods

Sample

To test the hypotheses, three studies were analyzed.¹ Study A was carried out in a laboratory setting in 2019 and included medical students in their third to sixth years. Study B included medical students in their fifth to sixth years. Data collection for this study was online due to the pandemic situation in 2020 and 2021. In both studies, participation was voluntary, and participants were paid 10 per hour. Study C was embedded as an online session in the curriculum of the third year of medical school in 2022. Participation was mandatory, but permission to use the data for research purposes was given voluntarily. All participants took part in only one of the three studies. All three studies received ethical approval from LMU Munich (approval numbers 18-261, 18-262 & 22-0436). For a sample description of each study, see Table 1. We would like to emphasize that none of the students were specializing in internal medicine, ensuring that the study results reflect the competencies of regular medical students without specialized expertise.

Procedure

Each of the three studies was organized in the same way, with participants first completing a pretest that included a prior knowledge test, socio-demographic questions, and questions about individual motivational-affective characteristics (e.g., social skills, interest, and motivation). Participants then moved on to the CDR simulation and worked on the patient case. The patient case was the same for studies B and C, but was different for study A. The complexity and difficulty of the patient case did not vary systematically between the patient cases.

Simulation and Task

In the CDR simulation, which is also used as a learning environment, the task was to take over the role of an internist and to collaborate with an agent-based radiologist to obtain further information by performing radiological examinations to diagnose fictitious patient

¹ Please note that the data employed in this study have been used in previous publications (e.g., Brandl et al., 2021; Radkowsch, et al., 2021; Richters et al., 2022). However, the research question and the results reported in this study are completely unique to this study.

Table 2 Overview of the number of questions in the content knowledge test

Study	Conceptual knowledge in internal medicine	Conceptual knowledge in radiology	Strategic knowledge in internal medicine	Strategic knowledge in radiology
Study A	20	15	24 8 cases 3 questions per case	16 8 cases 2 questions per case
Study B	20	15	24 8 cases 3 questions per case	16 8 cases 2 questions per case
Study C	13	12	24 8 cases 3 questions per case	12 6 cases 2 questions per case

Table 3 Example items for each subscale for measuring social skills

Subscale	Item
Direct Measurement	I enjoy working with others.
Perspective taking	It is easy for me to put myself in the position of my collaboration partners.
Information sharing	When I collaborate with others, I purposefully share relevant information.
Negotiating	I can negotiate compromises when working with others.
Coordination	When I work with others, we have a clear common goal in mind.

cases with the chief symptom of fever. Medical experts from internal medicine, radiology, and general medicine constructed the patient cases. Each case was structured in the same way: by studying the medical record individually, then collaborating with an agent-based radiologist, and finally reporting the final diagnosis and its justification again individually. For a detailed description on the development and validation of the simulation, see Radkowsch and colleagues (2020).

Before working within the simulation, participants were presented with an instruction for the simulated scenario and informed what they were to do with it. Then, we instructed participants how to access further information in the medical record by clicking on hyperlinks, as well as how they could use the toolbar to make notes for the later in the process. Furthermore, we acquainted the students with how they could request further information through collaborating with a radiologist.

During the collaboration with an agent-based radiologist, participants were asked to fill out request forms to obtain further evidence from radiological examinations needed to diagnose the patient case. To effectively collaborate with radiologists, it is crucial for internists to clearly communicate the type of evidence required to reduce uncertainty (referred to as “evidence elicitation”) and share any relevant patient information such as signs, symptoms, and medical history (referred to as “evidence sharing”) as well as suspected diagnoses under investigation (referred to as “hypotheses sharing”) that may impact the radiologists’ diagnostic process. Only when participants shared evidence and hypotheses appropriately for their requested examination did they receive a description and evaluation of the radiologist’s radiologic findings. What was considered appropriate was determined by medical

Table 4 Means, standard deviations, and internal consistency for individual characteristics, collaborative diagnostic activities, and diagnostic outcome per study

Variable	Study A			Study B			Study C		
	M	SD	ω	M	SD	ω	M	SD	ω
Conceptual knowledge in internal medicine ^c	0.61	0.12	.37	0.65	0.13	.53	0.49	0.18	.49
Conceptual knowledge in radiology ^c	0.67	0.15	.55	0.68	0.16	.62	0.52	0.18	.39
Strategic knowledge in internal medicine ^c	0.58	0.12	.53	0.61	0.13	.54	0.48	0.14	.40
Strategic knowledge in radiology ^c	0.42	0.12	.38	0.47	0.14	.45	0.44	0.16	.44
Collaboration knowledge ^c	0.70	0.09	.83	0.72	0.09	.83	0.65	0.10	.82
Direct measurement ^d	4.45	0.63	.79	4.36	0.64	.74	4.59	0.67	.81
Perspective taking ^d	4.36	0.57	.70	4.35	0.57	.62	4.50	0.58	.62
Information sharing ^d	4.55	0.52	.59	4.49	0.55	.62	4.56	0.51	.66
Negotiating ^d	4.76	0.57	.46	4.71	0.57	.51	4.71	0.49	.28
Coordination ^d	4.51	0.58	.72	4.51	0.60	.71	4.58	0.59	.76
Evidence elicitation ^{a,c}	0.49	0.50	– ^b	0.70	0.46	– ^b	0.67	0.47	– ^b
Evidence sharing ^c	0.55	0.19	– ^b	0.60	0.22	– ^b	0.53	0.24	– ^b
Hypotheses sharing ^{a,c}	0.61	0.49	– ^b	0.62	0.49	– ^b	0.51	0.50	– ^b
Diagnostic accuracy ^{a,c}	0.82	0.38	– ^b	0.90	0.30	– ^b	0.92	0.27	– ^b
Diagnostic justification ^c	0.34	0.23	– ^b	0.43	0.22	– ^b	0.41	0.20	– ^b
Diagnostic efficiency ^c	0.07	0.02	– ^b	0.07	0.04	– ^b	0.05	0.02	– ^b

^aBinary indicator^bSingle measure item^cRanging from 0 to 1^dRanging from 1 to 6

experts for each case and examination in preparation of the cases. Therefore, this scenario involves more than a simple division of tasks, as the quality of one person's activity (i.e., description and evaluation of the radiologic findings) depends on the collaborative efforts (i.e., CDAs) of the other person (OECD, 2017)

Measures—Individual Characteristics

The individual characteristics were measured in the pretest. The internal consistencies of each measure per study are displayed in Table 4 in the Results section. We want to point out that the internal consistency of knowledge as a construct—determined by the inter-correlations among knowledge pieces—typically exhibits a moderate level. Importantly, recent research argues that a moderate level of internal consistency does not undermine the constructs' capacity to explain a significant amount of variance (Edelsbrunner, 2024; Stadler et al., 2021; Taber, 2018).

Content knowledge was separated into radiology and internal medicine knowledge, as these two disciplines play a major role in the diagnosis of the simulated patient cases. For each discipline, conceptual and strategic knowledge was assessed (Kiesewetter et al., 2020; Stark et al., 2011). The items in each construct were presented in a randomized way in each study. However, the items for study C were shortened due to the embedding of the data

collection in the curriculum. Therefore, items with a very high or low item difficulty in previous studies were excluded (Table 2).

Conceptual knowledge was measured using single-choice questions including five options adapted from a database of examination questions from the Medical Faculty of the LMU Munich, focusing on relevant and closely related diagnoses of the patient cases used in the simulation. A mean score of 0–1 was calculated, representing the percentage of correct answers and indicating the average conceptual knowledge of the participant per medical knowledge domain.

Strategic content knowledge was measured contextually using key features questions (M. R. Fischer et al., 2005). Short cases were introduced followed by two to three follow up questions (e.g., What is your most likely suspected diagnosis?, What is your next examination?, What treatment do you choose?). Each question had eight possible answers, from which the learners were asked to choose one. Again, a mean score of 0–1 was calculated, representing the percentage of correct responses, indicating the average strategic content knowledge of the participant per domain.

The measure of collaboration knowledge was consistent across the three studies and specific to the simulated task. Participants were asked to select all relevant information for seven different patient cases with the cardinal symptom of fever (internal medicine). The patient cases were presented in a randomized order and always included 12 pieces of information regarding the chief complaints, medical history, and physical examination of the patient cases. We then assessed whether each piece of information was shared correctly (i.e. whether relevant information was shared and irrelevant information was not shared) and assigned 1 point and divided it by the maximum of 12 points to standardized the range of measure to 0–1. Then we calculated a mean score for each case and then across all cases, resulting in a range of 0–1 indicating the participants' collaboration knowledge

The construct of social skills was consistent across the three data collections and was measured on the basis of self-report on a 6-point Likert scale ranging from total disagreement to total agreement. The construct was measured using 23 questions divided into five subscales; for example items, see Table 3. Five questions aimed to measure the overall construct, and the other four subscales were identified using the complex problem-solving frameworks of Liu et al. (2016) and Hesse et al. (2015): perspective taking (four questions), information sharing (five questions), negotiation (four questions), and coordination (five questions). For the final score, the mean of all subcategories was calculated, ranging from 1 to 6, representing general social skills.

Measures—Collaborative Diagnostic Activities (CDA)

We operationalize CDAs in the pretest case in terms of quality of evidence elicitation, evidence sharing, and hypotheses sharing. The internal consistencies of each measure per study are displayed in Table 4 in the Results section.

The quality of evidence elicitation was measured by assessing the appropriateness of the requested radiological examination for the indicated diagnosis. An expert solution was developed to indicate which radiological examinations were appropriate for each of the possible diagnoses. If participants requested an appropriate radiological examination for the indicated diagnoses, they received 1 point for that request attempt. Finally, a mean score across all request attempts (maximum of 3) was calculated and scored. The final mean score was transformed into a binary indicator, with 1 indicating that all requested radiological examinations were appropriated and 0 indicating that inappropriate radiological

examinations were also requested, due to the categorical nature of the original data and its skewed distribution, with a majority of responses concentrated in a single category.

The quality of evidence sharing was measured using a precision indicator. This was calculated as the proportion of shared relevant evidence out of all shared evidence. Relevant evidence is defined per case and per diagnosis and indicated by the expert solution. The precision indicator was first calculated per radiological request. We then calculated the mean score, summarizing all attempts in that patient case. This resulted in a range from 0 points, indicating that only irrelevant evidence was shared, to 1 point, indicating that only relevant evidence was shared.

The quality of hypotheses sharing was also measured using a precision indicator. For each patient case, the proportion of diagnoses relevant for the respective patient case to all shared diagnoses was calculated. Which diagnoses were considered relevant for a specific case was determined by an expert solution. As with evidence elicitation, this score was evaluated and converted into a binary variable, where 1 indicated that only relevant diagnoses were shared and 0 indicated that also irrelevant diagnoses were shared, due to the categorical nature of the original data and its skewed distribution, with a majority of responses concentrated in a single category.

Measures—Diagnostic Outcome

We operationalize diagnostic outcome in the pretest case in terms of diagnostic accuracy, diagnostic justification, and diagnostic efficiency.

For diagnostic accuracy, a main diagnosis was assigned to each patient case as expert solution. After working on the patient case and requesting the radiological examination, participants indicated their final diagnosis. To do this, they typed in the first three letters of their desired diagnosis and then received suggestions from a list of 249 possible diagnoses. Diagnostic accuracy was then calculated by coding the agreement between the final diagnosis given and the expert solution. Accurate diagnoses (e.g., hospital-acquired pneumonia) were coded as 1, correct but inaccurate diagnoses (e.g., pneumonia) were coded as 0.5, and incorrect diagnoses were coded as 0. A binary indicator was used for the final diagnostic accuracy score, with 0 indicating an incorrect diagnosis and 1 indicating an at least inaccurate diagnosis, due to the categorical nature of the original data and its skewed distribution, with a majority of responses concentrated in a single category.

A prerequisite for diagnostic justification and diagnostic efficiency is the provision of at least an inaccurate diagnosis. If a participant provided an incorrect diagnosis (coded as 0), diagnostic justification and diagnostic efficiency were immediately scored as 0.

After choosing a final diagnosis, participants were asked to justify their decision in an open text field. Diagnostic justification was then calculated as the proportion of relevant reported information out of all relevant information that would have fully justified the final accurate diagnosis. Again, medical experts agreed on an expert solution that included all relevant information to justify the correct diagnosis. The participants' solution was coded by two independent coders, each coding the full data, and differences in coding were discussed until the coders agreed. We obtained a range from 0 points, indicating a completely inadequate justification, to 1 point, indicating a completely adequately justified final diagnosis.

Diagnostic efficiency was defined as diagnostic accuracy (non-binary version) divided by the minutes required to solve the case.

Statistical Analyses

To answer the research question, a structural equation model (SEM) was estimated using MPlus Editor, version 8 (Muthén & Muthén, 2017). We decided to use a SEM, as it is a comprehensive statistical approach widely used in psychology and educational sciences for its ability to model complex relationships among observed and latent variables while accounting for measurement error (Hilbert & Stadler, 2017). SEM support the development and verification of theoretical models, enabling scholars to refine theories and interventions in psychology and education based on empirical evidence, as not only can one relationship be investigated but a system of regressions is also considered simultaneously (Nachtigall et al., 2003).

We included all links between the variables and applied a two-step approach, using mean-adjusted and variance-adjusted unweighted least squares (ULSMV, Savalei & Rhemtulla, 2013) as the estimator and THETA for parametrization, first examining the measurement model and then the structural model. The assessment of model fit was based on chi-square (χ^2), root mean square error of approximation (RMSEA), and comparative fit index (CFI). Model fit is generally indicated by small chi-squared values; RMSEA values of < 0.08 (acceptable) and < 0.06 (excellent), and CFI values ≥ 0.90 . We do not consider standardized root mean squared residual (SRMR), because, according to the definition used in MPlus, this index is not appropriate when the sample size is 200 or less, as natural variation in such small samples contributes to larger SRMR values (Asparouhov & Muthén, 2018). For hypotheses 1–6, we excluded path coefficients < 0.1 from our interpretation, as they are relatively small. In addition, at least two interpretable path coefficients, of which at least one is statistically significant, are required to find support for the hypothesis. For hypotheses 7–15, specific indirect effects (effect of an individual characteristic on diagnostic outcome through a specific CDA) and total indirect effects (mediation of the effect of an individual characteristic on diagnostic outcome through all mediators) were estimated.

We reported all measures in the study and outlined differences between the three samples. All data and analysis code have been made publicly available at the Open Science Framework (OSF) and can be accessed at <https://osf.io/u8t62>. Materials for this study are available by email through the corresponding author. This study's design and its analysis were not pre-registered.

Results

The descriptive statistics of each measure per study are displayed in Table 4. The intercorrelations between the measures per study can be found in Appendix Table 7.

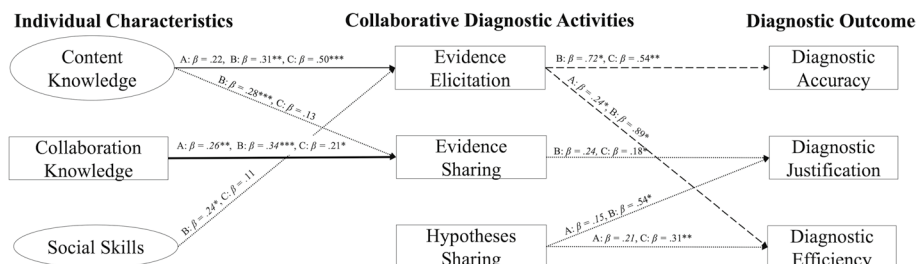
Overall Results of the SEM

All loadings were in the expected directions and statistically significant, except for conceptual knowledge in internal medicine in study C ($\lambda = 0.241$, $p = .120$), conceptual knowledge in radiology in study A ($\lambda = 0.398$, $p = .018$), and strategic knowledge in internal medicine ($\lambda = 0.387$, $p = .206$) and radiology ($\lambda = -0.166$, $p = .302$) in study B. Standardized factor loadings of the measurement model are shown in Appendix Table 8.

Table 5 Standardized paths coefficients (β) and standard errors (SE) for paths between individual characteristics, collaborative diagnostic activities, and diagnostic outcome per study

Hypothesis	Path between Independent variable	Dependent variable	Study A		Study B		Study C		Conclusion
			β	SE	β	SE	β	SE	
1a	Content knowledge	EE	.22	0.12	.31**	0.12	.50***	0.10	Supported
1b		ES	.08	0.08	.28***	0.08	.13	0.09	Supported
1c		HS	.12	0.12	-.14	0.12	.1	0.12	Not supported
2a	Collaboration knowledge	EE	-.03	0.10	.16	0.10	-.04	0.08	Not supported
2b		ES	.26***	0.08	.34***	0.07	.21*	0.08	Fully supported
2c		HS	.06	0.10	.09	0.10	.03	0.09	Not supported
3a	Social skills	EE	.01	0.11	.24*	0.11	.11	0.10	Supported
3b		ES	.00	0.08	.23***	0.08	-.03	0.08	Not supported
3c		HS	-.1	0.11	-.1	0.11	.11	0.11	Not supported
4a	EE	Diagnostic accuracy	-.25	0.16	.72*	0.37	.54**	0.19	Supported
4b		Diagnostic justification	.03	0.12	-.69	0.37	.25	0.16	Not supported
4c		Diagnostic efficiency	.24*	0.12	.89*	0.40	.08	0.18	Supported
5a	ES	Diagnostic accuracy	-.02	0.12	-.27	0.25	.20	0.20	Not supported
5b		Diagnostic justification	-.19	0.10	.24	0.21	.18*	0.09	Supported
5c		Diagnostic efficiency	.09	0.09	-.18	0.22	.05	0.07	Not supported
6a	HS	Diagnostic accuracy	.13	0.16	-.22	0.33	.13	0.17	Not supported
6b		Diagnostic justification	.15	0.12	.54*	0.27	-.18	0.12	Supported
6c		Diagnostic efficiency	.21	0.11	-.4	0.28	.31**	0.11	Supported

EE evidence elicitation, ES evidence sharing, HS hypotheses sharing. * $p < .05$, ** $p < .01$, and *** $p < .001$



Only coefficients > 0.10 are displayed. The full thick line represents a positive statistically significant relationship in three studies; the full line represents a positive relationship in three studies, of which two are statistically significant; the dashed line represents a positive statistically significant relationship in two studies; and the dotted line represents a positive statistically significant relationship in two studies, of which one is statistically significant. $p < .05$. $**p < .01$. $***p < .001$

Fig. 2 Evidence on supported relationships between individual characteristics, collaborative diagnostic activities, and diagnostic outcome

The SEM has a good fit for study A [$X^2(75) = 74.086$, $p = .508$, RMSEA = 0.00, CFI = 1.00], study B [$X^2(75) = 68.309$, $p = .695$, RMSEA = 0.000, CFI = 1.00], and study C [$X^2(75) = 93.816$, $p = .070$, RMSEA = 0.036, CFI = 1.00].

Paths between Individual Characteristics, CDAs, and Diagnostic Outcome

The standardized path coefficients and hypotheses tests for the theoretical model are reported in Table 5. An overview of the paths supported by the data is shown in Fig. 2.

Overall, the R^2 for the CDAs ranged from medium to high for evidence elicitation and evidence sharing, depending on the study, and were consistently low for hypotheses sharing across all three studies. Looking at diagnostic outcome, R^2 is consistently large for diagnostic accuracy and medium to large for diagnostic justification and diagnostic efficiency (Table 6).

Table 6 R^2 for collaborative diagnostic activities and diagnostic outcome per study

Variable	Study A		Study B		Study C	
	R^2	SE	R^2	SE	R^2	SE
EE ^a	.051	0.05	.166	0.09	.289	0.11
ES ^a	.077	0.04	.234***	0.07	.061	0.04
HS ^a	.023	0.03	.035	0.04	.027	0.03
Diagnostic accuracy ^a	.286	0.13	.343	0.20	.332	0.17
Diagnostic justification	.165	0.08	.246	0.19	.146	0.07
Diagnostic efficiency	.141	0.07	.422	0.22	.143	0.10

EE evidence elicitation, ES evidence sharing, HS hypotheses sharing.

* $p < .05$, ** $p < .01$, and *** $p < .001$

^aFor dichotomous criterions, MPlus computes a pseudo- R^2

The path from content knowledge to evidence elicitation was positive and > 0.1 in all three studies, as well as statistically significant in two of them; therefore, we consider Hypothesis 1a supported. The path from content knowledge to evidence sharing was positive and > 0.1 in two studies, as well as statistically significant in one of them; therefore, Hypothesis 1b is also supported. In contrast, the path from content knowledge to hypotheses sharing was indeed also positive in two studies, but as neither was statistically significant, we conclude that Hypothesis 1c was not supported. The path from collaboration knowledge to evidence elicitation was positive and > 0.1 in only one study, but also not statistically significant. Thus, we found that Hypothesis 2a was not supported. For the path from collaboration knowledge to evidence sharing, we found relevant positive and statistically significant coefficients in all three studies. Hypothesis 2b is therefore fully supported by the data. This is not the case for Hypothesis 2c, for which we found no coefficient > 0.1 for the path from collaboration knowledge to hypotheses sharing. For the path from social skills to evidence elicitation, we found positive coefficients > 0.1 in two out of three studies, of which one was also statistically significant. Thus, we consider Hypothesis 3a to be supported. For the path from social skills to evidence sharing, we again found one statistically significant positive coefficient, but in the other two studies it was < 0.1 . Therefore, we do not consider Hypothesis 3b to be supported by the data. The same applies to the path from social skills to hypotheses sharing, where the coefficient is < 0.1 in two studies. We therefore do not consider Hypothesis 3c to be supported.

The path from evidence elicitation to diagnostic accuracy was statistically significant and large in magnitude in two out of three studies. Hypothesis 4a is therefore supported. The path from evidence elicitation to diagnostic justification was only positive and > 0.1 in one study, which was also not statistically significant. Therefore, we find no support for Hypothesis 4b. In contrast, the path from evidence elicitation to diagnostic efficiency was positive and statistically significant in two out of three studies, with one large effect. Hypothesis 4c is therefore supported. The path from evidence sharing to diagnostic accuracy was only positive and reasonably large in one study. Therefore, we do not find support for Hypothesis 5a. The path from evidence sharing to diagnostic justification was positive and > 0.1 in two studies as well as statistically significant in one of them, so Hypothesis 5b is supported. In contrast, we did not find a positive coefficient > 0.1 for the path from evidence sharing to diagnostic efficiency. Therefore, Hypothesis 5c is not supported by the data. Although we found coefficients > 0.1 in two studies for the path from hypotheses sharing to diagnostic accuracy, we found no support for Hypothesis 6a, as none of these was statistically significant. This is different for Hypothesis 6b, as we found two positive paths from hypotheses sharing to diagnostic justification, one of which was statistically significant and large. Finally, we found two positive paths from evidence sharing to diagnostic efficiency in three studies, one of which was statistically significant. Hypothesis 6c is therefore supported.

Indirect Effects between Individual Characteristics, CDA, and Diagnostic Outcome

Indirect effects of CDAs on the effect of individual characteristics on the diagnostic outcome in CDR were estimated to test hypotheses 7–15. Although we found a mediating effect of all CDAs ($\beta = .31$, $p = .008$), and specifically for evidence elicitation ($\beta = .27$, $p = .021$) from content knowledge on diagnostic accuracy in study C, and some significant overall and direct effects for other relationships (Appendix Table 9), none of these were consistent across all of the studies. Thus, we conclude no consistent support for any of the Hypotheses 7–15.

Discussion

The aim of the current study was to investigate the extent to which the relationships specified in the CDR model (Radkowsch et al., 2022) are applicable across studies, to better understand the processes underlying CDR in knowledge-rich domains. Not only is this exploration crucial for the medical field or collaborative problem-solving in knowledge-rich domains, but it also offers valuable insights for computer-supported collaborative learning research. Despite CDR's specific focus, the principles and findings have relevant implications for understanding and enhancing collaborative processes in various educational and professional settings.

Specifically, we investigated how individual learner characteristics, the CDAs, and the diagnostic outcome are related. We therefore analyzed data from three independent studies, all from the same context, a simulation-based environment in the medical domain. Our study found positive relationships between content knowledge and the quality of evidence elicitation as well as the quality of evidence sharing, but not for the quality of hypotheses sharing. Furthermore, collaboration knowledge is positively related to the quality of evidence sharing, but not to the quality of evidence elicitation and the quality of hypotheses sharing. Social skills are only positively related to the quality of evidence elicitation. This underscores the multifaceted nature of collaborative problem-solving situations. Thus, effective CDR, a form of collaborative problem-solving, necessitates a nuanced understanding of the interplay between individual characteristics and CDAs.

The relevance of content knowledge for diagnostic competence is well established in research (Chernikova et al., 2020). To develop any diagnostic skills in knowledge-rich domains, learners need to acquire large amounts of knowledge and to restructure it through experience with problem-solving procedures and routines (Boshuizen et al., 2020). In the case of CDR this enables the diagnostician to come up with an initial suspected diagnosis, which is likely to be relevant information for the collaboration partner and to guide the further CDAs effectively. The finding that content knowledge only has a relation to the quality of evidence elicitation but none of the other CDAs can be explained by the fact that evidence elicitation is the least transactive CDA within the collaborative decision-making process. When eliciting evidence, the collaboration partner is used as an external knowledge resource (Weinberger & Fischer, 2006). So, despite being a collaborative activity, evidence elicitation is about what information from the collaboration partner is needed rather than what the collaboration partner needs. Thus, elicitation is less transactive than sharing, which is focused at what the collaboration partner needs.

Not only content knowledge but also collaboration knowledge is related to the quality of evidence sharing. This finding implies that collaboration knowledge may influence the CDR above and beyond individual content knowledge. It also supports the differentiation of knowledge types made in the CDR model (Radkowsch et al., 2022). Thus, it is important to learn not only the conceptual and strategic medical knowledge that is required for diagnosing but also knowledge about what information is relevant for specific collaboration partners when diagnosing collaboratively. This finding underpins the importance of being aware of the knowledge distribution among collaboration partners and the relevance of the transactive memory (Wegner, 1987). Thus, for collaborative problem-solving in knowledge-rich domains—as for computer-supported collaborative learning more generally—knowledge and information awareness is crucial (Engelmann & Hesse, 2010).

Thus, the relevance of collaboration knowledge in collaborative problem-solving is an important finding of our study, highlighting that it is critical in facilitating effective

collaborative processes and outcomes. The current findings emphasize the need for educational strategies that explicitly target the development of collaborative knowledge to ensure that learners have the knowledge and skills necessary to participate in productive collaborative problem-solving and computer-supported collaborative learning processes. In doing so, the CDR model emphasizes the need for learners to master collaborative skills and build shared problem representations to take full advantage of collaborative learning opportunities.

As CDR is conceptualized to be an interplay of cognitive and social skills (Hesse et al., 2015), we also assumed that social skills are related to CDAs. However, we only found evidence of the expected relationship between social skills and CDAs for the quality of evidence elicitation. One explanation could be that collaboration knowledge was relatively high in all three samples, outweighing the influences of general skills. This is consistent with the assumption of the CDR model that the influence of more general social skills is reduced with an increasing level of professional collaboration knowledge (Radkowsch et al., 2022). When collaboration knowledge is available to the diagnosticians, it becomes more important than social skills. This finding again underlines the importance of collaboration knowledge, which can be seen as a domain- and profession-specific development of social skills. However, another explanation could be that, when collaborating with an agent, the effect of social skills decreases, as the agent was not programmed to respond to social nuances. The design of the simulation would thus buffer against the effect of social skills. Although the study by Herborn et al. (2020) found no differences between human-to-human and human-to-agent collaboration, this does not necessarily invalidate the potential variability in outcomes associated with the social skills incorporated into the agent. For a thorough investigation into the impact of social skills, the agent would need variable social abilities, enabling the variation of the importance of basic social skills for successful collaboration.

Further, we need to conclude that there is no support for a relationship between the individual characteristics and hypotheses sharing, as we found no stable support for the relationship between any of the individual characteristics and the quality of hypotheses sharing. One possible explanation could be that the binary precision measure used to operationalize quality in hypotheses sharing is not sensitive enough or is not capturing the relevant aspect of quality in that activity. Another explanation could be that there is no direct relationship between the individual characteristics and hypotheses sharing, as this relationship is mediated by evidence sharing and thus influenced by the activated knowledge scripts (Schmidt & Rikers, 2007).

Looking at the relationships between CDAs and the diagnostic outcome, the current results highlight the need to distinguish between primary (diagnostic accuracy) and secondary (diagnostic justification and efficiency) outcomes of diagnostic reasoning (Daniel et al., 2019). Achieving diagnostic accuracy, a purely quantitative outcome measure, is less transactive than other aspects of the diagnostic outcome. This is also where we find the link to evidence elicitation, as we consider this to be the least transactive CDA within the collaborative decision-making process. However, the ability to justify and reach this decision efficiently is then highly dependent on evidence sharing and hypotheses sharing, activities that are more focused on transactivity within CDR (Weinberger & Fischer, 2006).

Although individual learner characteristics are found to have an effect on CDAs, and CDAs impact the diagnostic outcome, the effect is not mediated by CDAs across studies. Thus, we assume that, for effective collaborative problem-solving in knowledge-rich domains, such as CDR, it is not enough to have sufficient content and collaboration knowledge; it is also necessary to be able to engage in high quality CDAs to achieve a

high-quality diagnostic outcome. This is consistent with research on individual diagnostic reasoning, which shows that diagnostic activities have a unique contribution to the diagnostic outcome after controlling for content knowledge (Fink et al., 2023).

In summary, we explored evidence elicitation, evidence sharing, and hypotheses sharing as crucial CDAs. The findings revealed diverse associations of these CDAs with individual characteristics and facets of the diagnostic outcome, supporting the notion that the CDR-process involves a variety of different skills (instead of being one overarching skill). On the basis of these results, we propose categorizing CDAs into activities primarily focused on individual goals and needs (e.g., elicitation) and more transactive activities directly targeted at the collaborator (e.g., sharing). To enhance quality in CDAs, instructional support should be considered. For instance, providing learners with an adaptive collaboration script has been shown to improve evidence sharing quality and promote the internalization of collaboration scripts, fostering the development of collaboration knowledge (Radkowitz et al., 2021). Further, group awareness tools, such as shared concept maps, should be considered to compensate for deficits in one's collaboration knowledge (Engelmann & Hesse, 2010). However, what is required to engage in high-quality CDAs remains an open question. One starting point is domain-general cognitive skills. These could influence CDAs, particularly in the early stages of skill development (Hetmanek et al., 2018). Previous research showed that, in diagnostic reasoning, instructional support is more beneficial when being domain-specific than domain-general (Schons et al., 2022). Thus, there is still a need for further research on how such instructional support might look like.

Future Research

Although we used data from three studies, all of them were in the same domain; thus, it remains an open question whether these findings are applicable across domains. The CDR model claims that the described relationships are not limited to the medical domain, but rather are valid across domains for collaboratively solving complex problems in knowledge-rich domains. Future research should explore generalizability, for example, for teacher education, which is a distinct field that also requires diagnosing and complex problem-solving (Heitzmann et al., 2019).

Regardless of domain, the non-mediating relationship of CDAs between individual characteristics and diagnostic outcomes, as well as the found effects of the CDAs in the current study, suggests that an isolated analysis of CDAs does not fully represent the complex interactions and relationships among activities, individual characteristics, and diagnostic outcomes. Future studies might assess CDAs as a bundle of necessary activities, including a focus on their possible non-linear interactions. We propose to use process data analysis to account for the inherent complexity of the data, as different activities in different sequences can lead to the same outcome (Y. Chen et al., 2019). More exploratory analyses of fine-grained, theory-based sequence data are needed to provide insights into more general and more specific processes involved in successful solving complex problems collaboratively (Stadler et al., 2020).

As our results have shown, collaboration knowledge and thus awareness of the knowledge distribution among collaboration partners is highly relevant. While a recent meta-analysis showed a moderate effect of group awareness of students' performance in computer-supported collaborative learning (D. Chen et al., 2024), it has so far not been systematically investigated in collaborative problem-solving. Thus, more research on the influence collaboration knowledge in collaborative problem-solving is needed.

Further, additional factors associated with success in collaborative problem-solving—not yet incorporated into the model and thus not yet investigated systematically—include communication skills (OECD, 2017), the self-concept of problem-solving ability (Scalise et al., 2016), and positive activating emotions during problem-solving tasks (Camacho-Morles et al., 2019).

Limitations

There are, however, some limitations to be considered. One is that we have only considered CDAs and how they relate to individual characteristics and outcomes. However, the CDR model also introduces individual diagnostic activities, such as the generation of evidence and the drawing of conclusions. These occur before and after the CDAs and may therefore also have an impact on the described relationships. However, we decided to focus on the CDAs within the CDR process because they are particularly relevant for constructing a shared problem representation, being central to CDR. Future research might consider these individual diagnostic activities, as they could, for example, further explain the how content knowledge is related to the diagnostic outcome.

Another limitation of the current analyses is the operationalization of quality for the CDAs. We chose the appropriateness of radiological examination for the indicated diagnosis for quality of evidence elicitation and precision for quality of evidence sharing and hypotheses sharing. However, all of these only shed light on one perspective of each activity, while possibly obscuring others. For example, it may be that content knowledge is not related to the precision of hypotheses sharing, but this may be different when looking at other quality indicators, such as sensitivity or specificity. However, we decided to use the precision aspect of activities, as research shows that collaborators often fail to identify relevant information, and the amount of information is not related to performance (Tschan et al., 2009). Future research may explore a broader variety of quality indicators to be able to assess the quality of CDAs as comprehensively as possible. It should also be noted that in study B a suppression effect (Horst, 1941) between hypothesis sharing and evidence elicitation artificially inflated the observed effect size. This is to be expected with process data that can be highly correlated and needs to be considered when interpreting the effect sizes.

In addition, it should be noted that the omega values obtained for the conceptual and strategic knowledge measures were below the commonly accepted threshold of 0.7. While we chose to use omega values as a more appropriate measure of reliability in our context, given the complex and multifaceted nature of the knowledge constructs, these lower-than-expected values raise important questions about the quality of the data and the robustness of the findings. Thus, it is important to understand that knowledge constructs, by their very nature, may not always exhibit high levels of internal consistency due to the diverse and interrelated components they encompass (Edelsbrunner, 2024; Stadler et al., 2021; Taber, 2018). This complexity may be reflected in the moderate omega values observed, which, while seemingly counterintuitive, does not invalidate the potential of the constructs to account for substantial variance in related outcomes. However, findings related to these constructs should be interpreted with caution, and the results presented should be considered tentative. Future research should further explore the implications of using different reliability coefficients in assessing complex constructs within the learning sciences, potentially providing deeper insights into the nuanced nature of knowledge and its measurement.

Another limitation of this study is related to the agent-based collaboration, as a predictive validation of collaborative problem-solving for later human-to-human collaboration in comparable contexts has not yet been systematically conducted. Although the agent-based collaboration situation used has been validated in terms of perceived authenticity, it still does not fully correspond to a real collaboration situation (Rosen, 2015). This could be an explanation for the low influence of social skills, as the setting might not require the application of a broad set of social skills (Hesse et al., 2015; Radkowsitch et al., 2020). In a real-life collaboration, the effects of social skills might be more pronounced. However, research showed that the human-to-agent approach did not lead to different results in collaborative problem-solving than the human-to-human approach in the 2015 PISA study, and correlations with other measures of collaborative skills have been found (Herborn et al., 2020; Stadler, Herborn et al., 2020). Future studies should specifically test the relevance of social skills for CDR in a human-to-human setting to strengthen the generalizability of our findings.

Conclusion

In conclusion, the current study highlights the importance of individual characteristics and CDAs as independent predictors for achieving good diagnoses in collaborative contexts, at least in the simulation-based settings we used in the studies included in our analysis. Collaboration knowledge emerged as a critical factor, demonstrating its importance over early acquired, general social skills. Therefore, it is imperative to revise the CDR approach by giving higher priority to the proficiency of collaboration knowledge compared with social skills. Furthermore, we conclude that, in simulation-based CDR, content knowledge does not play such a crucial role in predicting diagnostic success compared with many other educational settings, most probably because of the endless opportunities for retrying and revising in simulation-based learning environments.

With respect to CDAs, we suggest refining the perspective on the quality of CDAs and consider revising the CDR model by summarizing CDAs as information elicitation and information sharing, with the former being less transactive, and thus, less demanding than the latter. Adequate performance in both types of CDA is presumed to result in a high-quality shared problem representation, resulting in good diagnostic outcome. Collaborative problem-solving skills are highly relevant in professional practice of knowledge-rich domains, highlighting the need to strengthen these skills in students engaged in CDR and to provide learning opportunities accordingly. Further, the ability to effectively collaborate and construct shared problem representations is important, not only in CDR but also in collaborative problem-solving and computer-supported collaborative learning more in general, highlighting the need for integrating such skills into curricula and instructional design.

By emphasizing these aspects, we can improve the diagnostic skills of individuals in collaborative settings. Through advancing our understanding of CDR, we are taking a key step forward in optimizing collaborative problem-solving and ultimately contributing to improved diagnostic outcomes in various professional domains beyond CDR in medical education. In particular, integrating collaboration knowledge and skills into computer-supported collaborative learning environments can enrich learning experiences and outcomes in various knowledge-rich domains.

Appendix

Table 7 Intercorrelations (Pearson), for the latent and manifest variables

Study A		1	2	3	4	5	6	7	8	9
Variable										
1	Content Knowledge	–								
2	Collaboration Knowledge	0.21**	–							
3	Social Skills	0.13	–0.03	–						
4	EE	0.12	–0.03	0.04	–					
5	ES	0.14	0.26***	0.01	0.04	–				
6	HS	0.06	0.05	–0.08	–0.12	–0.05	–			
7	Diagnostic Accuracy	0.23**	0.02	0.02	–0.09	0.00	0.13	–		
8	Diagnostic Justification	0.21*	0.05	0.15	0.4	–0.15	0.14	– ^a	–	
9	Diagnostic Efficiency	–0.11	0.10	–0.16	0.10	0.11	0.11	– ^a	–0.00	–
Study B		1	2	3	4	5	6	7	8	9
Variable										
1	Content Knowledge	–								
2	Collaboration Knowledge	0.16*	–							
3	Social Skills	–0.04	0.06	–						
4	EE	0.26**	0.14	0.16*	–					
5	ES	0.25**	0.34***	0.19*	0.46***	–				
6	HS	–0.03	0.09	–0.07	0.29***	–0.01	–			
7	Diagnostic Accuracy	0.05	0.12	–0.06	0.25**	0.11	0.07	–		
8	Diagnostic Justification	0.10	0.05	0.04	–0.11	–0.02	0.12	– ^a	–	
9	Diagnostic Efficiency	0.10	0.14	–0.17	0.22*	0.20*	0.07	– ^a	–0.26**	–
Study C		1	2	3	4	5	6	7	8	9
Variable										
1	Content Knowledge	–								
2	Collaboration Knowledge	0.09	–							
3	Social Skills	0.17*	0.16*	–						
4	EE	0.29***	–0.03	0.15*	–					
5	ES	0.09	0.19**	0.02	0.03	–				
6	HS	0.07	0.00	0.08	0.25***	–0.06	–			
7	Diagnostic Accuracy	0.05	–0.03	0.01	0.25**	0.09	0.14	–		
8	Diagnostic Justification	0.16	0.15	0.05	0.14	0.24**	–0.08	– ^a	–	
9	Diagnostic Efficiency	0.12	0.03	–0.01	0.16*	0.07	0.26***	– ^a	–0.08	–

EE evidence elicitation, ES evidence sharing, HS hypotheses sharing. * $p < .05$, ** $p < .01$, and *** $p < .001$

^aThese correlations cannot be calculated as if diagnostic accuracy were 0; diagnostic justification and efficiency are coded as NA

Table 8 Standardized factor loadings (λ) and standard errors (SE) in the measurement model

Factor	Observed variable	Study A		Study B		Study C	
		λ	SE	λ	SE	λ	SE
Content knowledge	Conceptual knowledge in internal medicine	0.709***	0.168	0.741***	0.166	0.241	0.155
	Conceptual knowledge in radiology	0.398*	0.168	0.626***	0.148	0.424**	0.136
	Strategic knowledge in internal medicine	0.525**	0.169	0.387	0.206	0.797***	0.118
	Strategic knowledge in radiology	0.755**	0.217	-0.166	0.302	0.651***	0.113
Social skills	Direct measurement	0.612***	0.05	0.625***	0.056	0.681***	0.065
	Perspective taking	0.595***	0.062	0.642***	0.06	0.639***	0.059
	Information sharing	0.671***	0.047	0.719***	0.056	0.627***	0.068
	Negotiating	0.611***	0.055	0.614***	0.051	0.511***	0.065
	Coordination	0.685***	0.05	0.65***	0.054	0.754***	0.052

SE standard errors. * $p < .05$, ** $p < .01$, and *** $p < .001$

Table 9 Mediation between individual characteristics and diagnostic outcomes by collaborative diagnostic activities

Mediation		Study A		Study B		Study C	
Name	Type	β	SE	β	SE	β	SE
Content knowledge—diagnostic accuracy	Total	.45***	0.111	.26	0.151	.13	0.134
	Total indirect	-.04	0.063	.18	0.145	.31**	0.118
	Via EE	-.06	0.05	.22	0.148	.27*	0.117
	Via ES	0	0.01	-.08	0.076	.03	0.033
	Via HS	.01	0.023	.03	0.052	.01	0.022
	Direct	.49***	0.122	.09	0.173	-.18	0.166
Content knowledge—diagnostic justification	Total	.26*	0.109	.16	0.101	.19	0.104
	Total indirect	.01	0.043	-.22	0.149	.13	0.086
	Via EE	.01	0.027	-.22	0.147	.12	0.085
	Via ES	-.01	0.021	.07	0.061	.03	0.02
	Via HS	.02	0.021	-.07	0.079	-.02	0.025
	Direct	.26*	0.119	.38*	0.182	.06	0.139
Content knowledge—diagnostic efficiency	Total	-.13	0.102	.09	0.11	.17	0.104
	Total indirect	.09	0.05	.28	0.17	.08	0.093
	Via EE	.05	0.039	.28	0.187	.04	0.089
	Via ES	.01	0.012	-.05	0.065	.01	0.011
	Via HS	.03	0.029	.06	0.062	.03	0.037
	Direct	-.21*	0.104	-.2	0.198	.09	0.143
Collaboration knowledge—diagnostic accuracy	Total	.04	0.119	.18	0.146	-.08	0.241
	Total Indirect	.01	0.047	0	0.096	.02	0.071
	Via EE	.01	0.025	.12	0.092	-.02	0.046
	Via ES	0	0.033	-.09	0.089	.04	0.045
	Via HS	.01	0.016	-.02	0.038	0	0.011
	Direct	.03	0.127	.18	0.152	-.1	0.241
Collaboration knowledge—diagnostic justification	Total	.05	0.095	.05	0.08	.16	0.086
	Total indirect	-.04	0.037	.02	0.09	.03	0.031
	Via EE	0	0.005	-.11	0.09	-.01	0.021
	Via ES	-.05	0.031	.08	0.076	.04	0.024
	Via HS	.01	0.017	.05	0.057	0	0.016
	Direct	.1	0.1	.03	0.122	.14	0.091
Collaboration knowledge—diagnostic efficiency	Total	.12	0.1	.13	0.11	.04	0.092
	Total indirect	.03	0.04	.05	0.1	.01	0.035
	Via EE	-.01	0.024	.14	0.113	0	0.01
	Via ES	.03	0.026	-.06	0.077	.01	0.016
	Via HS	.01	0.023	-.04	0.047	0	0.028
	Direct	.09	0.103	.09	0.141	.04	0.089
Social skills—diagnostic accuracy	Total	-.05	0.128	-.1	0.125	.05	0.142
	Total indirect	-.02	0.035	.14	0.112	.07	0.061
	Via EE	0	0.027	.18	0.123	.06	0.054
	Via ES	0	0.001	-.06	0.059	-.01	0.018
	Via HS	-.01	0.022	.02	0.041	.01	0.023
	Direct	-.03	0.131	-.23	0.164	-.02	0.141

Table 9 (continued)

Mediation		Study A		Study B		Study C	
Name	Type	β	SE	β	SE	β	SE
Social Skills—Diagnostic Justification	Total	.14	0.104	.03	0.1	-.03	0.089
	Total indirect	-.02	0.027	-.17	0.109	0	0.037
	Via EE	0	0.004	-.17	0.114	.03	0.03
	Via ES	0	0.016	.05	0.049	0	0.016
	Via HS	-.01	0.021	-.06	0.069	-.02	0.022
	Direct	.15	0.101	.2	0.125	-.04	0.086
Social Skills—Diagnostic Efficiency	Total	-.14	0.094	-.2*	0.082	-.06	0.096
	Total indirect	-.02	0.032	.22	0.13	.04	0.04
	Via EE	0	0.026	.22	0.143	.01	0.022
	Via ES	0	0.008	-.04	0.051	0	0.004
	Via HS	-.02	0.025	.04	0.057	.03	0.036
	Direct	-.12	0.097	-.42**	0.136	-.1	0.089

EE evidence elicitation, ES evidence sharing, HS hypotheses sharing. * $p < .05$, ** $p < .01$, and *** $p < .001$

Funding Open Access funding enabled and organized by Projekt DEAL. 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 (FI 792/11-1 & FI 792/11-2)

Declarations

Conflict of interest statement On behalf of all authors, the corresponding author states that there is no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

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. <https://doi.org/10.1007/s12186-017-9183-x>
- Asparouhov, T., & Muthén, B. (2018). *SRMR in Mplus*. <https://www.statmodel.com/download/SRMR2.pdf>
- Bauer, E., Sailer, M., Kiesewetter, J., Fischer, M. R., & Fischer, F. (2022). Diagnostic argumentation in teacher education: Making the case for justification, disconfirmation, and transparency. *Frontiers in Education*, 7, Article 977631. <https://doi.org/10.3389/educ.2022.977631>
- 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*, 29, 100310. <https://doi.org/10.1016/j.edurev.2020.100310>

- Brandl, L., Richters, C., Radkowsitch, A., Obersteiner, A., Fischer, M. R., Schmidmaier, R., Fischer, F., & Stadler, M. (2021). Simulation-based learning of complex skills: Predicting performance with theoretically derived process features. *Psychological Test and Assessment Modeling*, 63(4), 542–560. https://www.psychologie-aktuell.com/fileadmin/Redaktion/Journale/ptam-2021-4/PTAM__4-2021_6_kor.pdf
- Braun, L. T., Zottmann, J. M., Adolf, C., Lottspeich, C., Then, C., Wirth, S., Fischer, M. R., & Schmidmaier, R. (2017). Representation scaffolds improve diagnostic efficiency in medical students. *Medical Education*, 51(11), 1118–1126. <https://doi.org/10.1111/medu.13355>
- Camacho-Morles, J., Slemp, G. R., Oades, L. G., Morrish, L., & Scoular, C. (2019). The role of achievement emotions in the collaborative problem-solving performance of adolescents. *Learning and Individual Differences*, 70, 169–181. <https://doi.org/10.1016/j.lindif.2019.02.005>
- Chen, D., Zhang, Y., Luo, H., Zhu, Z., Ma, J., & Lin, Y. (2024). Effects of group awareness support in CSCL on students' learning performance: a three-level meta-analysis. *International Journal of Computer-Supported Collaborative Learning*, 19(1), 97–129. <https://doi.org/10.1007/s11412-024-09418-3>
- Chen, Y., Li, X., Liu, J., & Ying, Z. (2019). Statistical analysis of complex problem-solving process data: an event history analysis approach. *Frontiers in Psychology*, 10, Article 486. <https://doi.org/10.3389/fpsyg.2019.00486>
- 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. <https://doi.org/10.1007/s10648-019-09492-2>
- Chernikova, O., Heitzmann, N., Opitz, A., Seidel, T., & Fischer, F. (2022). A theoretical framework for fostering diagnostic competences with simulations in higher education. In F. Fischer & A. Opitz (Eds.), *Learning to Diagnose with Simulations*. Springer, Cham. https://doi.org/10.1007/978-3-030-89147-3_2
- Cook, D. A., Brydges, R., Zendejas, B., Hamstra, S. J., & Hatala, R. M. (2013). Technology-enhanced simulation to assess health professionals: a systematic review of validity evidence, research methods, and reporting quality. *Academic Medicine: Journal of the Association of American Medical Colleges*, 88(6), 872–883. <https://doi.org/10.1097/ACM.0b013e31828ffdcf>
- Daniel, M., Rencic, J., Durning, S. J., Holmboe, E. S., Santen, S. A., Lang, V., Ratcliffe, T., Gordon, D., Heist, B., Lubarsky, S., Estrada, C. A., Ballard, T., Artino, A. R., Da Sergio Silva, A., Cleary, T., Stojan, J., & Gruppen, L. D. (2019). Clinical reasoning assessment methods: a scoping review and practical guidance. *Academic Medicine: Journal of the Association of American Medical Colleges*, 94(6), 902–912. <https://doi.org/10.1097/ACM.0000000000002618>
- Dunbar, K. (1995). How scientists really reason: scientific reasoning in real-world laboratories. In R. J. Sternberg & J. E. Davidson (Eds.), *The nature of insight* (pp. 365–395). MIT Press.
- Edelsbrunner, P. A. (2024). *Does interference between intuitive conceptions and scientific concepts produce reliable inter-individual differences?* Science & Education. Advance online publication. <https://doi.org/10.1007/s11191-024-00500-8>
- 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. <https://doi.org/10.1007/s11412-010-9089-1>
- Fink, M. C., Heitzmann, N., Reitmeier, V., Siebeck, M., Fischer, F., & Fischer, M. R. (2023). Diagnosing virtual patients: the interplay between knowledge and diagnostic activities. *Advances in Health Sciences Education: Theory and Practice*, 1–20. <https://doi.org/10.1007/s10459-023-10211-4>
- Fiore, S. M., Graesser, A. C., & Greiff, S. (2018). Collaborative problem-solving education for the twenty-first-century workforce. *Nature Human Behaviour*, 2(6), 367–369. <https://doi.org/10.1038/s41562-018-0363-y>
- 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. <https://doi.org/10.1080/00461520.2012.748005>
- Fischer, F., Kollar, I., Ufer, S., Sodian, B., Hussmann, H., Pekrun, R., Neuhaus, B. J., Dörner, B., Pankofter, S., Fischer, M. R., Strijbos, J.-W., Heene, M., & Eberle, J. (2014). Scientific reasoning and argumentation: advancing an interdisciplinary research agenda in education. *Frontline Learning Research*, 2(3), 28–45. <https://doi.org/10.14786/flr.v2i2.96>
- 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. <https://doi.org/10.1080/01421590500078471>
- Förtsch, C., Sommerhoff, D., Fischer, F., Fischer, M. R., Girwidz, R., Obersteiner, A., Reiss, K., Stürmer, K., Siebeck, M., Schmidmaier, R., Seidel, T., Ufer, S., Wecker, C., & Neuhaus, B. J. (2018). Systematizing professional knowledge of medical doctors and teachers: development of an

- interdisciplinary framework in the context of diagnostic competences. *Education Sciences*, 8(4), 207. <https://doi.org/10.3390/educsci8040207>
- Goldhammer, F., Naumann, J., Rölke, H., Stelter, A., & Tóth, K. (2017). Relating product data to process data from computer-based competency assessment. In D. Leutner, J. Fleischer, J. Grünkorn, & E. Klieme (Eds.), *Methodology of educational measurement and assessment. competence assessment in education* (pp. 407–425). Springer International Publishing. https://doi.org/10.1007/978-3-319-50030-0_24
- Graesser, A. C., Fiore, S. M., Greiff, S., Andrews-Todd, J., Foltz, P. W., & Hesse, F. W. (2018). Advancing the science of collaborative problem solving. *Psychological Science in the Public Interest: A Journal of the American Psychological Society*, 19(2), 59–92. <https://doi.org/10.1177/1529100618808244>
- Hautz, W. E., Kämmer, J. E., Schaubert, S. K., Spies, C. D., & Gaissmaier, W. (2015). Diagnostic performance by medical students working individually or in teams. *JAMA*, 313(3), 303–304. <https://doi.org/10.1001/jama.2014.15770>
- Heitzmann, N., Seidel, T., Hetmanek, A., Wecker, C., Fischer, M. R., Ufer, S., Schmidmaier, R., Neuhaus, B. J., Siebeck, M., Stürmer, K., Obersteiner, A., Reiss, K., Girwidz, R., Fischer, F., & Opitz, A. (2019). Facilitating diagnostic competences in simulations in higher education: a framework and a research agenda. *Frontline Learning Research*, 1–24. <https://doi.org/10.14786/flr.v7i4.384>
- Herborn, K., Stadler, M., Mustafić, M., & Greiff, S. (2020). The assessment of collaborative problem solving in PISA 2015: can computer agents replace humans? *Computers in Human Behavior*, 104, 105624. <https://doi.org/10.1016/j.chb.2018.07.035>
- Hesse, F. W., 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). Springer.
- Hetmanek, A., Engelmann, K., Opitz, A., & Fischer, F. (2018). Beyond intelligence and domain knowledge. In F. Frank, C. Clark A., E. Katharina, O. Jonathan, F. Fischer, C. A. Chinn, K. Engelmann, & J. Osborne (Eds.), *Scientific reasoning and argumentation* (pp. 203–226). Routledge. <https://doi.org/10.4324/9780203731826-12>
- Hilbert, S., & Stadler, M. (2017). Structural equation models. In V. Zeigler-Hill & T. K. Shackelford (Eds.), *Encyclopedia of Personality and Individual differences* (pp. 1–9). Springer International Publishing. https://doi.org/10.1007/978-3-319-28099-8_1285-1
- Hitchcock, D. (2005). Good reasoning on the Toulmin model. *Argumentation*, 19(3), 373–391. <https://doi.org/10.1007/s10503-005-4422-y>
- Horst, P. (1941). The prediction of personnel adjustment. *Social Science Research and Council Bulletin* (48), 431–436.
- Jeong, H., Hmelo-Silver, C. E., & Jo, K. (2019). Ten years of computer-supported collaborative learning: a meta-analysis of CSCL in STEM education during 2005–2014. *Educational Research Review*, 28, 100284. <https://doi.org/10.1016/j.edurev.2019.100284>
- Kiesewetter, J., Fischer, F., & Fischer, M. R. (2017). Collaborative clinical reasoning—a systematic review of empirical studies. *The Journal of Continuing Education in the Health Professions*, 37(2), 123–128. <https://doi.org/10.1097/CEH.0000000000000158>
- Kiesewetter, J., Sailer, M., Jung, V. M., Schönberger, R., Bauer, E., Zottmann, J. M., Hege, I., Zimmermann, H., Fischer, F., & Fischer, M. R. (2020). Learning clinical reasoning: how virtual patient case format and prior knowledge interact. *BMC Medical Education*, 20(1), 73–83. <https://doi.org/10.1186/s12909-020-1987-y>
- Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science*, 12(1), 1–48. https://doi.org/10.1207/s15516709cog1201_1
- Koschmann, T. D., Feltoovich, P. J., Myers, A. C., & Barrows, H. S. (1992). Implications of CSCL for problem-based learning. *ACM SIGCUE Outlook*, 21(3), 32–35. <https://doi.org/10.1145/130893.130902>
- Liu, L., Hao, J., Davier, A. A. von, Kyllonen, P., & Zapata-Rivera, J.-D. (2016). A tough nut to crack: measuring collaborative problem solving. In J. Keengwe, Y. Rosen, S. Ferrara, & M. Mosharraf (Eds.), *Handbook of Research on Technology Tools for Real-World Skill Development* (pp. 344–359). IGI Global. <https://doi.org/10.4018/978-1-4666-9441-5.ch013>
- Muthén, L. K., & Muthén, B. O. (2017). *Mplus: Statistical Analysis with Latent Variables: User's Guide* (Version 8) [Computer software]. Authors.
- Nachtigall, C., Kroehne, U., Funke, F., & Steyer, R. (2003). (Why) should we use SEM? Pros and cons of structural equation modeling. *Methods of Psychological Research Online*, 8(2), 1–22.
- Noroozi, O., Biemans, H. J., 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>

- OECD. (2017). PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematic, Financial Literacy and Collaborative Problem Solving, revised edition. PISA, OECD Publishing. <https://doi.org/10.1787/9789264281820-en>
- Pickal, A. J., Engelmann, K., Chinn, C. A., Girwidz, R., Neuhaus, B. J., & Wecker, C. (2023). Fostering the collaborative diagnosis of cross-domain skills in video-based simulations. In *Proceedings of the International Conference on Computer-supported for Collaborative Learning, Proceedings of the 16th International Conference on Computer-Supported Collaborative Learning—CSCL 2023* (pp. 139–146). International Society of the Learning Sciences. <https://doi.org/10.22318/csl2023.638463>
- Radkowsitch, 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>
- Radkowsitch, A., Sailer, M., Fischer, M. R., Schmidmaier, R., & Fischer, F. (2022). Diagnosing collaboratively: A theoretical model and a simulation-based learning environment. In F. Fischer & A. Opitz (Eds.), *Learning to Diagnose with Simulations*. Springer, Cham. https://doi.org/10.1007/978-3-030-89147-3_10
- Radkowsitch, 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. <https://doi.org/10.1016/j.learninstruc.2021.101487>
- Richters, C., Stadler, M., Radkowsitch, A., Behrmann, F., Weidenbusch, M., Fischer, M. R., Schmidmaier, R., & Fischer, F. (2022). Making the rich even richer? Interaction of structured reflection with prior knowledge in collaborative medical simulations. In A. Weinberger, W. Chen, D. Hernández-Leo, & B. Che (Chair), *International Society of the Learning Sciences*. Hiroshima, Japan.
- Rochelle, J., & Teasley, S. (1995). The construction of shared knowledge in collaborative problem solving. In C. O'Malley (Ed.), *Computer-Supported Collaborative Learning* (pp. 66–97). 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>
- Savalei, V., & Rhemtulla, M. (2013). The performance of robust test statistics with categorical data. *British Journal of Mathematical and Statistical Psychology*, 66(2), 201–223. <https://doi.org/10.1111/j.2044-8317.2012.02049.x>
- Scalise, K., Mustafic, M., & Greiff, S. (2016). Dispositions for collaborative problem solving. In S. Kuger, E. Klieme, N. Jude, & D. Kaplan (Eds.), *Methodology of Educational Measurement and Assessment. Assessing Contexts of Learning* (pp. 283–299). Springer International Publishing. https://doi.org/10.1007/978-3-319-45357-6_11
- Schmidt, H. G., & Mamede, S. (2015). How to improve the teaching of clinical reasoning: a narrative review and a proposal. *Medical Education*, 49(10), 961–973. <https://doi.org/10.1111/medu.12775>
- 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. <https://doi.org/10.1111/j.1365-2923.2007.02915.x>
- Schons, C., Obersteiner, A., Reinhold, F., Fischer, F., & Reiss, K. (2022). *Developing a simulation to foster prospective mathematics teachers' diagnostic competencies: the effects of scaffolding*. Advance online publication. <https://doi.org/10.1007/s13138-022-00210-0>
- Stadler, M., Herborn, K., Mustafic, M., & Greiff, S. (2020). The assessment of collaborative problem solving in PISA 2015: an investigation of the validity of the PISA 2015 CPS tasks. *Computers & Education*, 157, 103964. <https://doi.org/10.1016/j.compedu.2020.103964>
- Stadler, M., Hofer, S., & Greiff, S. (2020). First among equals: log data indicates ability differences despite equal scores. *Computers in Human Behavior*, 111, 106442. <https://doi.org/10.1016/j.chb.2020.106442>
- Stadler, M., Sailer, M., & Fischer, F. (2021). Knowledge as a formative construct: a good alpha is not always better. *New Ideas in Psychology*, 60. <https://doi.org/10.1016/j.newideapsych.2020.100832>
- 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>
- 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), 1467–1478.
- Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, 48(6), 1273–1296. <https://doi.org/10.1007/s11165-016-9602-2>
- 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. <https://doi.org/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. <https://doi.org/10.1023/A:1002993406499>
- 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. <https://doi.org/10.1007/s10648-016-9361-7>
- Vogel, F., Weinberger, A., Hong, D., Wang, T., Glazewski, K., Hmelo-Silver, C. E., Uttamchandani, S., Mott, B., Lester, J., Oshima, J., Oshima, R., Yamashita, S., Lu, J., Brandl, L., Richters, C., Stadler, M., Fischer, F., Radkowsch, A., Schmidmaier, R., . . . Noroozi, O. (2023). Transactivity and knowledge co-construction in collaborative problem solving. In *Proceedings of the International Conference on Computer-Supported for Collaborative Learning, Proceedings of the 16th International Conference on Computer-Supported Collaborative Learning—CSCCL 2023* (pp. 337–346). International Society of the Learning Sciences. <https://doi.org/10.22318/csccl2023.646214>
- 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). Springer New York. https://doi.org/10.1007/978-1-4612-4634-3_9
- Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers & Education*, 46(1), 71–95. <https://doi.org/10.1016/j.compedu.2005.04.003>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Please note that the data employed in this study have been used in previous publications (e.g., Brandl et al., 2021; Radkowsch, et al., 2021; Richters et al., 2022). However, the research question and the results reported in this study are completely unique to this study. An initial version of this article is presented as a poster at ISLS 2024.

Authors and Affiliations

Laura Brandl¹  · Matthias Stadler^{1,2} · Constanze Richters¹ · Anika Radkowsch³ · Martin R. Fischer² · Ralf Schmidmaier⁴ · Frank Fischer¹

✉ Laura Brandl
L.Brandl@psy.lmu.de

¹ Department of Psychology, Ludwig-Maximilians-Universität München, Leopoldstr. 13, 80802 Munich, Germany

² Institute of Medical Education, LMU University Hospital, Ludwig-Maximilians-Universität München, Munich, Germany

³ IPN Leibniz Institute for Science and Mathematics Education, Department of Mathematics Education, Kiel, Germany

⁴ Medizinische Klinik und Poliklinik IV, LMU University Hospital, Ludwig-Maximilians-Universität München, Munich, Germany