



Pre-service teachers' evidence-based reasoning during pedagogical problem-solving: better together?

Andras Csanadi^{1,2}  · Ingo Kollar³ · Frank Fischer¹

Received: 25 January 2019 / Revised: 18 December 2019 / Accepted: 3 February 2020 /

Published online: 18 March 2020

© The Author(s) 2020

Abstract

This study investigates if collaboration and the level of heterogeneity between collaborating partners' problem-solving scripts can influence the extent to which pre-service teachers engage in evidence-based reasoning when analyzing and solving pedagogical problem cases. We operationalized evidence-based reasoning through its content and process dimensions: (a) to what extent pre-service teachers refer to scientific theories or evidence of learning and instruction (content level) and (b) to what extent they engage in epistemic processes of scientific reasoning (process level) when solving pedagogical problems. Seventy-six pre-service teachers analyzed and solved a problem about an underachieving student either individually or in dyads. Compared with individuals, dyads of pre-service teachers referred less to scientific content, but engaged more in hypothesizing and evidence evaluation and less in generating solutions. A greater dyadic heterogeneity indicated less engagement in generating solutions. Thus, collaboration may be a useful means for engaging preservice teachers in analyzing pedagogical problems in a more reflective and evidence-based manner, but pre-service teachers may still need additional scaffolding to do it based on scientific theories and evidence. Furthermore, heterogeneous groups regarding the collaborating partners' problem-solving scripts may require further instructional support to discuss potential solutions to the problem.

Keywords Evidence-based reasoning · Pre-service teachers · Collaborative reasoning · Group heterogeneity · Problem-solving scripts

✉ Andras Csanadi
csaoand@gmail.com

Ingo Kollar
ingo.kollar@phil.uni-augsburg.de

Frank Fischer
frank.fischer@psy.lmu.de

Extended author information available on the last page of the article

Introduction

Many voices, both from educational policy-making and from educational research, emphasize a need for evidence-based practice in teaching (EC 2013; Petty 2009). This means that teachers should be able to ground their decisions and the way they handle pedagogical problems on scientific evidence (Hetmanek et al. 2015; Spencer et al. 2012). For example, when noticing that a particular student seems to lack study motivation in a lesson, a teacher might draw on theory and research conducted in the context of self-determination theory to face this problem (e.g., she might decide to provide the student autonomy with respect to the kind of activity she wants to engage in to address her need for autonomy; Ryan and Deci 2000).

Developing skills related to evidence-based teaching is no doubt an ongoing task for teachers on their course towards professionalism. Yet, we argue that pre-service teacher education is of particular importance in this context: During university education, pre-service teachers can be systematically exposed to typical (often simulated) problems from classroom practice and be scaffolded to engage in evidence-based teaching and reasoning activities that eventually might become part of their teaching repertoire in later phases during their career. Nevertheless, empirical research shows that pre-service teachers often have difficulties with the application of scientific knowledge (theories and evidence) to problems from educational practice (Klein et al. 2015; Yeh and Santagata 2015). Furthermore, they also often do not succeed in reasoning about pedagogical problems they encounter in a systematic, hypothetico-deductive manner (see Kiemer and Kollar 2018): Very often, they have difficulties in identifying the problems, and even if they do so, they often are not able to explain and solve those problems in a systematic way (Sherin and Van Es 2009).

One way to help pre-service teachers develop skills for evidence-based teaching is to have them reflect on cases from educational practice (e.g., Piwovar et al. 2018; Fischer et al. 2013). From research on collaborative learning, we may speculate that this might be especially effective when it is done in collaboration with a peer (e.g., Baeten and Simons 2014; Chi and Wylie 2014). Yet, even though there is a lot of research on the benefits of collaborative learning in other domains (for an overview, see Springer et al. 1999), there is not much evidence on whether collaborative learning will help the development of evidence-based teaching skills. Further, it is unclear how exactly collaboration should be organized to be effective in this context. One issue here refers to the effects of different group compositions: While some research suggests that heterogeneous groups typically succeed over homogeneous group, other findings seem to suggest the opposite (Bowers et al. 2000; Rummel and Spada 2005; Wiley et al. 2013). Also, it is an open question with respect to what characteristics groups should be homo- or heterogeneous. One dimension that has hardly been investigated in the past is the homo-/heterogeneity of group members' procedural knowledge on how to approach a given task. In the context of this article, such knowledge will be termed "problem-solving scripts" (Kiemer and Kollar 2018).

This article therefore has two aims: First, we are interested in how groups of pre-service teachers (in our case dyads) compare with individuals when it comes to how they solve authentic pedagogical problem cases. We are interested in the effects of collaboration both on (a) the extent to which they apply scientific evidence and (b) the extent to which they approach such problems in a systematic, hypothetico-deductive manner. Second, when solving such cases collaboratively, we look at the effects of homo- resp. heterogeneity of the collaborating pre-service teachers' problem-solving scripts (i.e., the procedural knowledge on problem-solving they apply during problem-solving).

Evidence-based teaching as a venue for scientific reasoning

In many professions, the ability to solve domain-specific problems in accordance with scientific evidence has been labeled “evidence-based practice” (e.g., Spencer et al. 2012). When conceptualizing evidence-based practice in the teaching domain, and when conceptualizing the underlying cognitive processes, we propose to heuristically differentiate between two different, but related, kinds of reasoning processes: (a) the application of scientific knowledge to the problem (content level) and (b) the reasoning about the problem in a systematic, scientific manner (process level). At the *content level*, when pre-service teachers solve authentic problems from their future practice, they need to be able to apply theoretical and empirical knowledge from their domain (Gruber et al. 2000). In this sense, teachers are expected to refer to relevant psychological and educational theories on learning, motivation, instruction, etc. and related research findings when they are confronted with a pedagogical problem (e.g., how best to scaffold an underperforming student; Csanadi et al. 2015; Voss et al. 2011).

At the *process level*, evidence-based reasoning can be viewed as an inquiry process (Klahr and Dunbar 1988) that is characterized by an engagement in certain epistemic processes while analyzing and solving a problem (Fischer et al. 2014). Based on Fischer et al. (2014), we argue that when being confronted with problems from their actual practice, pre-service teachers should engage in a set of eight epistemic processes that are similar to those that scientists engage in when solving a research problem. They (1) need to identify the problem itself (problem identification), (2) ask questions or make statements that guide their further exploration on the problem (questioning), (3) set up candidate explanations for the problem (hypothesis generation), (4) take into account or generate further information necessary to understand or solve the problem (evidence generation), (5) evaluate the information in the context of their hypotheses (evidence evaluation), (6) plan interventions/solutions (constructing artefacts), (7) engage in discussions with others to re-evaluate their thoughts (communicating and scrutinizing), and (8) sum up their process to arrive at well-warranted conclusions on how to explain and/or solve the problem (drawing conclusions).

Nevertheless, as prior research has shown, pre-service teachers seem to have difficulties on both the content and process levels of evidence-based reasoning (Csanadi et al. 2015; Stark et al. 2009; Yeh and Santagata 2015). With respect to the *content level*, Lockhorst et al. (2010), for example, found that teachers indicated a lack of scientific knowledge to assess pupils' learning processes. Moreover, pre-service teachers seem to rarely refer to scientific sources (Csanadi et al. 2015), and they show difficulties with applying their knowledge to solve pedagogical problems (McNeill and Knight 2013). In rather rare cases in which they use scientific theories and models, they often do so in a superficial manner (Sampson and Blanchard 2012; Stark et al. 2009). With respect to the *process level*, van de Pol et al. (2011) showed that teachers often fail to collect enough evidence to properly make decisions on how to scaffold their pupils. Similarly, when pre-service teachers assess learning processes, for example, by watching videos of teacher-student interactions in the classroom, they seem to show relatively lower engagement in generating evidence-based hypotheses on their own compared with their performance after a methods course (Yeh and Santagata 2015).

Taken together, pre-service teachers show difficulties with respect to both dimensions of scientific reasoning, i.e., referring to scientific theories and evidence and performing epistemic processes, when it comes to solving pedagogical problems from their future professional practice. Thus, searching for ways to help pre-service teachers build up skills for evidence-based teaching is urgently needed.

Reasoning about authentic problem cases: Better together?

One potential way to foster pre-service teachers' evidence-based reasoning is to implement more opportunities for collaborative reflection on authentic problem cases in teacher education (e.g., Kolodner 2006). Through case-based reasoning (CBR), students may gradually develop new strategies to understand and solve new problems (e.g., Tawfik and Kolodner 2016). When applying these strategies to authentic problem cases, they can test and practice these newly acquired strategies in simulated real-world situations.

Very often, CBR is done in groups (Fischer et al. 2014; Kolodner et al. 2003). From a socio-cognitive perspective (Chi 2009; Fischer et al. 2013), collaborating partners may mutually stimulate or "scaffold" each other by articulating and building on the knowledge they individually bring to the discussion (Asterhan and Schwarz 2009; Chi and Wylie 2014; Teasley 1995). The potentials of collaboration for pre-service teachers have particularly been put forward by proponents of co-teaching (e.g., Rytivaara and Kershner 2012). This research argues that the engagement in collaborative discourse and the exchange of different perspectives and scientific knowledge may contribute to teachers' professional growth on becoming more reflective and evidence-based practitioners (Baeten and Simons 2014; Birrell and Bullough 2005; Korthagen and Kessels 1999). In one study, King (2006) surveyed 161 pre-service teachers with respect to their views regarding the potentials and problems of "paired placement," i.e., of sending pre-service teachers during initial teacher education to schools in pairs rather than individually. Among others, more than 80% of the pre-service teachers preferred being in a paired placement over being sent to school individually, and more than 70% of them stated that they felt that they learned a lot from attending their peer's lessons. A more recent study by Simons and Baeten (2016) further found that also teacher mentors attribute a high potential to co-teaching to foster pre-service teachers' professional growth.

Unfortunately though, research on how exactly pre-service teachers collaborate while solving authentic problem cases seems to be very scarce. Yet, research from other areas and with different subjects revealed a wealth of evidence demonstrating strong potentials of collaborative vs. individual learning on knowledge acquisition and problem-solving (for an overview, see Springer et al. 1999). Teasley (1995), for example, observed children working with a computer simulation either alone or in dyads. They had the opportunity to use different control keys to steer a "spaceship" while their main task was to find out the function of a "mystery key." This task gave the students the opportunity to develop hypotheses about the function of the mystery key as well as to test their assumptions. As a result, dyads engaged significantly more in hypothesizing than individuals did and produced proportionally more interpretive vs. descriptive talk than individuals. In another study (Okada and Simon 1997), male science major undergraduates participated in an online simulation environment and were asked to conduct molecular genetics experiments either as individuals or as pairs. Results showed that pairs of students engaged more frequently in hypothesizing and in evidence evaluation (planned experiments to test the initial hypothesis) than individuals. As the authors note, one reason for the heightened engagement in explanatory behavior that the dyads showed may be the need to communicate in a more explicit manner (Okada and Simon 1997). Such discussions can also be part of a coordination attempt for meaning-making and knowledge co-construction (Roschelle and Teasley 1995; Weinberger et al. 2007) which has been described as a relevant aspect of scientific reasoning as well (Fischer et al. 2014).

Despite these studies showing positive effects of collaboration on different epistemic processes, other studies demonstrate that collaboration does not always have positive effects.

In co-teaching, for example, planning the lesson with a partner requires coordination between the teachers (Jang 2008; Nokes et al. 2008). This might lead to difficulties, especially in case of disagreements between teaching partners (Bullough et al. 2002), as collaboration might become time-consuming (Jang 2008) and contribute to a higher perceived workload (Nokes et al. 2008). These findings resonate with evidence from further research on collaborative learning that has been conducted in other contexts. For example, Strijbos et al. (2004) as well as Weinberger et al. (2010) have shown that unclarified roles, i.e., who should do what and when during collaboration, may slow down or hinder the development of mutual understanding and constructive communication processes (Rummel and Spada 2005). Sociopsychological research further shows that groups often rely excessively on discussing what everybody in the group already knows anyway (“shared knowledge”), while knowledge and ideas that only single group members possess (“unshared knowledge”) are often discussed to a much lesser extent (e.g., Wittenbaum and Stasser 1996).

Given these equivocal results on the effects of collaboration on problem-solving, it seems that further variables might have an impact on whether groups benefit from collaboration or not. In their review on the effects of co-teaching, Baeten and Simons (2014), for example, mention that the successfulness of collaboration between pre-service teachers may be threatened when there is a lack of compatibility between teaching partners (e.g., different conceptions on teaching). This implies that group composition (i.e., who collaborates with whom) may be a factor that may impact the successfulness of collaboration. In the following, we argue that in particular pre-service teachers' homo- vs. heterogeneity of their procedural knowledge on how to tackle problematic classroom situations (i.e., their problem-solving scripts) might be crucial in this regard.

Group composition as a factor influencing the effectiveness of collaborative learning: homo-/heterogeneity of collaborating teachers' problem-solving scripts

The question whether group members should be similar (i.e., homogeneous) or dissimilar (i.e., heterogeneous) to each other (e.g., Bowers et al. 2000) in order to promote student engagement in high-level knowledge construction processes has already been addressed in pre-service teacher education. For example, studies in the context of pre-service teacher professionalization emphasize the importance of heterogeneous “paired placements” (Nokes et al. 2008) in order to foster a more reflective integration between scientific knowledge and practice (e.g., Korthagen and Kessels 1999). Further arguments favoring heterogeneous over homogeneous group composition can be derived from the broader research on collaborative learning. As Paulus (2000) described, in heterogeneous groups, reasoners may stimulate (Paulus 2000) each other by bringing new perspectives to a problem-solving task. A meta-analysis by Bowers et al. (2000) concludes that while homogeneous teams tend to perform better on low-complexity tasks that are well-defined (e.g., solving a puzzle), heterogeneous teams perform better on rather complex problem-solving tasks (e.g., business games). Similar findings have been reported by Canham et al. (2012): They asked university students to solve statistical probability problems either in homogeneous or in heterogeneous dyadic groups. Homo-/heterogeneity of the dyads was varied as the type of training that group members received before collaboration. In homogeneous dyads, both members received the same type of training, while in the heterogeneous condition, the two members of each dyad received different types

of training. After their training, dyads solved both problems that they practiced during their training (familiar problem) and novel ones (transfer problems). Results showed that although homogeneous dyads performed better on solving familiar problems, heterogeneous dyads performed better on transfer problems.

Yet, even though evidence for a superiority of heterogeneous over homogeneous dyads especially with regard to solving more complex problems seems to be strong, working in heterogeneous groups may also come with costs. For example, heterogeneous problem-solving partners may need more time and effort to reach a common understanding of the given problem and coordinate their strategies on how to solve that problem (Bullough et al. 2002; Rummel and Spada 2005). Taken together, findings indicate that in comparison with homogeneous pairings, heterogeneous learning partners may although need more time to coordinate their ideas with each other, they may be more reflective and evidence-based when solving complex problems.

Another open question is with respect to what criterion groups should or should not be heterogeneous. While previous studies focused on e.g., homo-/heterogeneity regarding expertise (Wiley and Jolly 2003; Wu et al. 2015) or demographic background (Curşeu and Pluut 2013), the present article is particularly interested in homo-/heterogeneity in terms of the partners' "problem-solving scripts." We define problem-solving scripts as individuals' knowledge and expectations on how to reason about problems from professional practice and assume that these scripts guide pre-service teachers' understanding and behavior in problem-solving situations (Fischer et al. 2013; Schank 1999). Evidence for the power of scripts during professional problem-solving comes especially from the medical field. For example, Charlin et al. (2007) showed that physicians possess clinical reasoning scripts, e.g., "illness scripts" that guide their medical professionals' diagnostic behavior. As an example, a physician meeting a pale patient may notice a problem, engage in a systematic evidence generation procedure (e.g., listening to the symptoms, conducting further investigations), and generate and revise plausible hypotheses (diagnoses) (Charlin et al. 2007). Similar problem-solving processes might also occur among pre-service teachers when analyzing authentic problem cases: For example, in the description of a hypothetical classroom situation, one pre-service teacher might detect that one student underperforms in a given task. This pre-service teacher may then generate hypotheses, i.e., explanations of the problem (e.g., that the student's overall ability is too low or that the teacher's explanations were too complicated); evaluate the evidence (e.g., by looking at the students' grades, she might decide that the student's ability may not play a role); and generate solutions (e.g., that the teacher in the scenario should improve her explanations so that students can comprehend them better). Yet, another pre-service teacher may hold a somewhat different problem-solving script. For example, she might start with mentally going through past experiences and then quickly jump to conclusions without conducting a further evaluation of the available information. If these two hypothetical pre-service teachers would work together when reflecting about an authentic problem case from teaching practice, their problem-solving scripts would have to be conceptualized as rather heterogeneous. Whether this heterogeneity would rather stimulate or hinder the pair from effective collaboration is one open question that we aim to answer in the empirical study we describe in the following.

Research questions

In our empirical study, we were interested in whether and how pre-service teachers engage in scientific reasoning while reflecting about authentic problem cases from their future professional practice. More specifically, we asked:

- *RQ1*: Do dyads of pre-service teachers differ from individual pre-service teachers in the extent to which they (a) refer to scientific theories and evidence and (b) in their engagement in different epistemic processes of scientific reasoning when they reflect on authentic problem cases from teaching practice?

Based on earlier studies from co-teaching (Baeten and Simons 2014) as well as from psychologically oriented research on collaborative learning (Teasley 1995), we hypothesize that dyadic reasoners might be more reflective, i.e., engage in epistemic processes of scientific reasoning (Okada and Simon 1997) to a larger extent than pre-service teachers who engage in individual reasoning. Yet, due to possible coordination problems in groups (Strijbos et al. 2004) or biases in information sharing (Wittenbaum and Stasser 1996), we do not assume this difference to be particularly large.

- *RQ2*: Within dyads, does the degree of heterogeneity regarding group members' problem-solving scripts affect the extent to which dyads of pre-service teachers (a) refer to scientific theories and evidence and (b) engage in different epistemic processes of scientific reasoning during reflecting about authentic problem cases?

First, in accordance with findings that show that heterogeneous partners may mutually stimulate each other in complex problem-solving tasks (Bowers et al. 2000; Wiley et al. 2013), we assume that the level of dyadic heterogeneity with respect to its members' problem-solving scripts would stimulate (Bowers et al. 2000; Paulus 2000) a reflective dialogue. In other words, the more heterogeneous pre-service teachers are, the more they should engage in epistemic processes of scientific reasoning. Yet, as high levels of heterogeneity might also increase coordination demands (Bullough et al. 2002; Rummel and Spada 2005), we do not expect this effect to be very large.

Method

Participants and design

Seventy-six teacher education students (59 female, $M_{\text{Age}} = 21.22$, $SD = 3.98$) from a German university participated in the study, studying on their first ($N = 41$), second ($N = 26$), third ($N = 4$), fourth ($N = 2$), and fifth ($N = 2$) semesters ($N = 1$ missing). They received course credit for their participation. To answer research question 1, we varied the independent variable "reasoning setting" (individual setting vs. dyadic setting): Each participant was randomly assigned to either an individual (16 students, 13 female, $M_{\text{Age}} = 22.31$; $SD = 6.73$) or a dyadic (60 students, i.e., 30 dyads, 46 females, $M_{\text{Age}} = 20.93$; $SD = 2.85$) condition. To answer research question 2, we calculated an index to measure each dyad's degree of heterogeneity based on a test in which

students were to individually describe how they approach authentic problems from pedagogical practice (see below). The resulting score on the heterogeneity index was then used as a predictor for the use of scientific theories and evidence as well as for the dyads' engagement in the different epistemic processes of scientific reasoning. Thus, homo- vs. heterogeneity was not experimentally manipulated, but rather computed post hoc.

Procedure

The procedure of the study consisted of four steps. Regardless of the condition (dyadic vs. individual), in the first three steps, every student participated individually. In the last (problem-solving) step, students participated either as dyads or individually depending on the condition they were assigned to. First, students filled in a computer-based questionnaire on demographic variables. After that, they were given a computer-based card-sorting task to measure their problem-solving scripts (see below). Then, they were given five minutes to read five printed out presentation slides that included scientific content information originating from their introductory psychology class and short descriptions of theories and concepts: on short- and long-term memory (Atkinson and Shiffrin 1968), the level of encoding (Craik and Lockhart 1972), and a classification of learning strategies (Wild 2006). After that, participants were presented an authentic problem from their future professional practice, which was the following: "You are a teacher in a school. One of your students receives low grades in comparison to others. The student looks motivated and it seems she understands the content. You know from the parents that she studies diligently at home. You as a teacher, please find possible reasons and a solution to the problem."

For dealing with the problem, participants had 10 minutes. During these ten minutes, students in the individual condition were asked to think aloud while solving the problem, and dyads were asked to orally discuss the problem. All think-aloud data and collaborative discussions were audio-recorded and transcribed for further analysis (see below). Data from this problem-solving process were used to measure students' use of scientific theories and evidence as well as their engagement in the epistemic processes of scientific reasoning (see below). Finally, students were thanked for their participation and debriefed. The whole session took about one hour for each individual or dyad.

Independent variables

Reasoning setting As mentioned, reasoning setting was varied by randomly assigning participants to the problem-solving phase either as individuals or dyads. All dyads were randomly established.

Degree of homo-/heterogeneity of problem-solving scripts within dyads (dyadic heterogeneity) The degree of homo-/heterogeneity with regard to the students' problem-solving scripts in the dyadic reasoning setting was not manipulated experimentally, but rather assessed post hoc on the basis of the results of the card-sorting task that preceded the problem-solving phase and during which students participated individually. First, they were presented the practice-related problem case described above. Then, they were asked to use a set of prefabricated cards available on a MS PowerPoint slide to indicate what (epistemic) processes they would perform while solving the presented problem. The eight epistemic processes from

Fischer et al. (2014) and five additionally selected “distractor” processes (e.g., “giving feedback,” “improvising”) were written on the cards that were presented to the participants. Besides that, five blank cards were provided to give participants the opportunity to add further processes if they wanted to. From the resulting process sequences, participants’ problem-solving scripts were coded in the following way: First, we summed those epistemic process cards representing processes from the Fischer et al. (2014) model that were selected by both dyadic members. This number represented their shared *knowledge component index* (SKCI) on scientific reasoning. Then, we calculated a *disagreement on position index* (DPI) between dyadic members by calculating how many out of the epistemic processes they agreed on (SKCI) would need to be switched in position so as both members’ selection shows the same sequence of the shared epistemic processes on scientific reasoning. We also calculated a *pooled knowledge component index* (PKCI) on scientific inquiry by summing the number of epistemic processes of the Fischer et al. (2014) model that at least one dyadic member had selected. Finally, a *homogeneity index* was calculated as $(SKCI - DPI)/PKCI$ to account for agreements and, at the same time, controlling for disagreements between the two members of a group. Larger values on this index thus indicated more dyadic homogeneity, while lower values indicated more dyadic heterogeneity.

Dependent variables

Dependent variables were collected during the problem-solving phase that had students (dyads or individuals) solve the aforementioned authentic problem from their future professional practice either by thinking aloud (in the individual reasoning setting) or by discussing it (in the dyadic reasoning setting; see above).

All verbal data were transcribed. After that, we segmented the data into syntactical proposition-sized units (Chi 1997). Ten percent of the data were independently segmented by two researchers after a training on segmentation. Reliability was calculated as the proportion of agreement according to Strijbos et al. (2006). The estimation of the reliability of segmentation through one indicator (e.g., Cohen’s κ) may lead to imprecise measurements due to the different segmentation strategies of different segmenters (Strijbos et al., 2006; Strijbos and Stahl 2007). Thus, in order to ensure segmentation reliability, we calculated reliability for both segmenters’ perspectives and set the criteria to 80% of agreement (Strijbos et al., 2006). In 85.09% out of the total number of segment boundary indicated by segmenter 1, segmenter 2 also agreed that there is a segment boundary. At the same time, in 79.73% out of the total number of segments indicated by segmenter 2, segmenter 1 also indicated a segment boundary. These values showed a good reliability of the segmentation scheme. In the next step, one of the segmenters segmented the remaining data.

The content level: reference to scientific theories and evidence We developed a coding scheme to capture for each segment whether or not participants used scientific theories and/or evidence or not. Segments were coded as “use of scientific theories and evidence” if the speaker referred to scientific theories, concepts, or methods. Specifically, based on the frequently occurring topics, we created five content categories: “learning strategy,” “anxiety,” “motivation,” “other scientific evidence,” and “no scientific evidence.” The code “learning strategy” was applied when participants mentioned ways of how the students in the case description may or should process the learning material (“elaborated strategy”; “recollection

difficulties”). We coded a segment in the “anxiety” category when participants referred to test anxiety or emotional pressure that the student in the case description may experience. The code “motivation” was used when participants referred to motivational constructs such as “self-concept” or “intrinsic/extrinsic motivation” in their reflection. The code “other scientific evidence” was used when participants referred to scientific evidence from other theories and lines of research that was not included in the preparation slides (such as “self-fulfilling prophecy”; “mobbing”; “mind map.” If the above codes did not apply, the segment was coded as “non-scientific evidence.” For each transcript (individual or dyadic), we merged the first four categories to calculate participants’ overall engagement in use of scientific theories and evidence. Then, we divided this value by the total number of propositions. The reason for this was that preliminary analysis had shown that the two reasoning settings (dyads vs. individuals) were not equal regarding the total amount of talk, and consequently, working with raw sum scores would have biased our analyses. Five percent of the segments were coded by two independent coders with an agreement of Cohen’s $\kappa = .82$.

The process level: engagement in epistemic processes A second coding scheme (see Table 1) was developed to capture the eight epistemic processes of scientific reasoning proposed by Fischer et al. (2014). We also added a code “non-epistemic process” that was applied when the participants engaged in an activity that did not fall into one of the eight categories proposed by Fischer et al. (2014). Since it proved impossible to reliably differentiate between the two activities “evidence generation” and “evidence evaluation,” we merged these two categories into one: evidence evaluation. After the coding scheme had been developed and trained, two independent coders coded 10% of the data for the identification of epistemic

Table 1 Coding scheme to capture epistemic processes of scientific reasoning

Code for epistemic process	Brief description of code	Example of code
Problem identification (PI)	An initial attempt to build an understanding of the problem.	“So it is about a student, // who has low grades”
Questioning (Q)	A question or statement orienting further inquiry.	“Ok, so what is the reason for that?”
Hypothesis generation (HG)	Any explanation of the problem case.	“So if the reason is her learning method”
Evidence generation (EG) (later merged with EE)	Referring to case information; to scientific evidence; to anecdotal evidence; to lack of information; or planning further information collection.	“She studies diligently at home” “There are different learning strategies...” “I know someone who has exam nerves” “We also do not know her age.” [to find out] “how much time she needs to do her homework.”
Evidence evaluation (EE) (later merged with EG)	Evaluation of evidence (to support/falsify HG or GS).	“and then you can even exclude the problem of exam nerves”
Generating solutions (GS)	Planning an intervention, how to solve the problem.	“You should discourage her from using surface strategies”
Communicating and scrutinizing (CS)	Planning to engage others in the inquiry process.	“You can also talk to the parents”
Drawing conclusions (DC)	Concluding the outcomes of the earlier steps of inquiry.	“For me these would be the most important points to understand at all what her problem is.”
Non-epistemic (NE)	Propositions that cannot be coded under the other codes.	“Ok, have you read it through?”

processes. Inter-rater reliability was sufficient (Cohen's $\kappa = .68$). After coding each segment, the numbers of segments that fell in the same category were summed for each epistemic process: problem identification (PI), questioning (Q), hypothesis generation (HG), generating solutions (GS), evidence evaluation (EE), drawing conclusions (DC), communicating and scrutinizing (CS), and non-epistemic processes (NE). For the same reasons reported above (working with nonbiased proportional scores), the resulting sum score for each process was divided by the total amount of talk. Finally, as the data screening (see below) revealed that three epistemic processes (problem identification, questioning, and drawing conclusions) had relatively large non-coded ratios, we dummy coded the corresponding variables in each protocol by assigning 0 when there was no proposition coded under the given epistemic process and assigning 1 when there was at least one proposition coded under that epistemic process.

Statistical analyses

To answer the question whether dyads differ from individuals on their engagement in their use of scientific theories and evidence and in their engagement in epistemic processes (RQ1), ANOVAs and a MANOVA were conducted.

To answer the question if dyadic heterogeneity has an effect on pre-service teachers' use of scientific theories and evidence and their engagement in epistemic processes (RQ2), we analyzed only data from the dyadic reasoning setting and conducted linear regressions with "dyadic heterogeneity" as predictor and "use of scientific theories and evidence" as well as "engagement in SR activities" as criterion variables.

For all analyses, the unit of analysis was the segments of each transcript, no matter whether it came from an individual or a dyad. I.e., segments of transcripts from individual think-aloud transcripts were compared with segments of transcripts from dyadic discussions.

For the analyses, the alpha level was set to $p < .05$. For the regressions to answer the effect of dyadic heterogeneity on the engagement in epistemic processes (RQ2b), we set the alpha to a more stringent $p < .01$ criterion in order to control for the increased type I error rate due to multiple simultaneously conducted analyses (e.g., Chen et al. 2017). Although in such cases of multiple analyses, a form of correction (e.g., Bonferroni correction) is often recommended to account for the inflated alpha, other studies point out that such measures can be criticized as being too conservative (as they decrease the probability of type I error, but simultaneously increase the probability of type II error; Sinclair et al. 2013). For example, Streiner and Norman (2011) note that alpha corrections may be less appropriate in case of (a) studies with an exploratory focus or (b) hypotheses that are stated a priori. Since both apply to the present study, we decided to use an alpha of .01 instead of applying Bonferroni corrections.

Results

Preliminary analyses

An initial preliminary analysis was conducted to compare whether the two groups of reasoning setting were equal on the total amount of talk. Results showed that dyads talked significantly more ($M = 138.50$, $SD = 40.16$) than individuals ($M = 88.19$, $SD = 27.74$) did ($F_{1,46} = 19.93$,

$p < .001$, partial $\eta^2 = .31$). Therefore, instead of using raw scores, we computed proportions of each code by dividing raw frequencies by the overall number of segments and used those proportions in the subsequent analyses.

With respect to the *reference to scientific theories and evidence* (RQ1a), dyads showed a positively skewed ($z = 3.87$) as well as a leptokurtic ($z = 4.80$) deviation from normality. Data screening indicated one outlier case ($z = 3.50$), which also accounted for the skewness in the data. Thus, we removed this participant from further analysis. Furthermore, Levene's test indicated unequal variances between individuals and dyads ($p < .05$). To answer RQ1a, we therefore used Welch's F -tests due to its higher robustness for the violation of homogeneity of variances in the case of unequal sample sizes (Field 2009, pp. 379–380; Grissom 2000).

With respect to the *engagement in epistemic processes* (RQ1b and RQ2b), the frequency distributions of three epistemic processes suggested very high frequencies of missing values for the variables *drawing conclusions* (67.39%), *questioning* (58.70%), and *problem identification* (39.13%). Consequently, we decided to exclude these variables from further parametric analyses. Instead, we dummy coded them and conducted chi-square tests to test the odds of engaging vs. not engaging in each process as a function of the reasoning context (collaborative vs. individual). Furthermore, as Levene's test suggested unequal variances for solution generation, Welch's F -test was calculated as a follow-up test for solution generation. Considering the assumption checks, we decided to conduct MANOVA with a robust statistic, such as Pillai's trace (Meyers et al. 2006, p. 378). Finally, for the dummy-coded *problem identification*, *questioning*, and *drawing conclusion*, nonparametric (logistic) regressions were conducted.

RQ1: Effects of reasoning setting on scientific reasoning

To test the effect of dyadic vs. individual reasoning on scientific content use (RQ1a), Welch's F -test was conducted with reasoning setting as the independent variable and reference to scientific theories and content as the dependent variable in the analysis. Reasoning setting showed a significant effect on the *use of scientific theories and evidence* (see Table 2 for details). Namely, individuals referred proportionally more to scientific theories and evidence than dyads did.

To answer the question whether dyads differed from individuals on their engagement in epistemic processes (RQ1b), we conducted a MANOVA with the four epistemic processes

Table 2 Effects of reasoning setting on use of scientific theories and evidence as well as on the epistemic processes hypothesis generation, evidence evaluation, solution generation, communicating and scrutinizing, and non-epistemic propositions (ANOVAs)

Dependent variable	Reasoning setting				F	p	Partial η^2
	Individuals		Dyads				
	M	SD	M	SD			
Use of scientific theories and evidence	.44	.20	.29	.14	6.99	.014	.17
Hypothesis generation	.17	.11	.24	.09	6.06	.018	.12
Evidence evaluation	.26	.13	.33	.11	4.28	.045	.09
Solution generation	.45	.24	.29	.13	6.56	.019	.17
Communicating and scrutinizing	.04	.04	.05	.03	0.40	.532	.01
Non-epistemic propositions	.02	.03	.06	.04	10.48	.002	.19

eligible for parametric analysis and non-epistemic propositions as dependent variables and reasoning setting (individual vs. dyadic) as independent variable. Reasoning setting had a significant strong multivariate effect on the engagement in epistemic processes (Pillai's trace = .40, $F_{5,40} = 5.26$, $p < .001$, partial $\eta^2 = .40$).

Follow-up ANOVAs as well as Welch's F -test for solution generation revealed significant effects of reasoning setting on the engagement in *hypothesis generation*, in *evidence evaluation*, in *solution generation*, and in the production of *non-epistemic propositions* (see Table 2 for the results). As seen in Table 2, dyads engaged significantly more in *hypothesis generation* and in *evidence evaluation* than individuals did. Also, they produced more *non-epistemic propositions* than individuals. On the other hand, individuals engaged more in *solution generation* than dyads did.

Finally, a nonparametric chi-square test indicated a significant relationship between reasoning setting and engagement in *drawing conclusions*: the odds of engaging in drawing conclusions were 5.35 times higher for dyads than for individuals. No further significant effect was found.

Tables 2 and 3 present all the results reported above:

RQ2: Effects of dyadic heterogeneity on scientific reasoning

To answer the question if dyadic heterogeneity has an effect on the extent to which pre-service teachers *refer to scientific theories and evidence* (RQ2a), we conducted a linear regression analysis with dyadic heterogeneity as the predictor and reference to scientific content as the criterion variable. Dyadic heterogeneity did not predict the *reference to scientific theories and evidence* (see Table 4 for details).

To answer the question whether dyadic heterogeneity had an effect on the *engagement in epistemic processes* (RQ2b), we computed separate linear regression analyses that included dyadic heterogeneity as a predictor, and solution generation, hypothesis generation, evidence evaluation, and communicating and scrutinizing separately as criterion variables (see Table 4 for the results). The regression models revealed that dyadic heterogeneity had a marginal significant effect on *solution generation*: The more heterogeneous dyads were, the less they engaged in generating solutions. Moreover, dyadic heterogeneity negatively, yet non-significantly, predicted the following epistemic processes: *hypothesis generation*, *evidence evaluation*, and *communicating and scrutinizing*. Finally, the outcomes of logistic regressions (see Table 5) suggested no effect of dyadic heterogeneity on *problem identification*, on *questioning* or on *drawing conclusions*.

Table 3 Effects of reasoning setting on the epistemic processes problem identification, questioning, and drawing conclusions (chi-square tests)

Dependent variable	Reasoning setting		χ^2	p	Odd's ratio
	Individuals Occurrence (%)	Dyads Occurrence (%)			
Problem identification	56.25	63.33	0.22	.639	1.34
Questioning	37.50	43.33	0.15	.702	1.27
Drawing conclusions	12.50	43.33	4.51	.034	5.35

Table 4 Effects of dyadic homo-/heterogeneity on use of scientific theories and evidence as well as on the epistemic processes hypothesis generation, evidence evaluation, solution generation, communicating and scrutinizing, and non-epistemic propositions (linear regressions)

Dependent variable	<i>B</i>	95% CI _B	β	<i>p</i>	<i>R</i> ²	adj. <i>R</i> ²
Use of scientific theories and evidence	-.17	-.65; .31	-.14	.470	.02	-.02
Hypothesis generation	-.17	-.46; .11	-.23	.217	.05	.02
Evidence evaluation	-.25	-.60; .09	-.27	.144	.08	.04
Solution generation	.49	.11; .86	.45	.013	.20	.17
Communicating and scrutinizing	-.04	-.15; .07	-.14	.463	.02	-.02

General discussion

The first aim of our study was to answer the question whether collaboration case-based reasoning is a useful means to engage pre-service teachers in evidence-based reasoning. Furthermore, we were interested in the effects of homo-/heterogeneous problem-solving scripts of collaborating partners on our participants' evidence-based reasoning. With respect to evidence-based reasoning, we differentiated between (a) a *content* level (i.e., the extent to which participants referred to scientific theories and evidence during reasoning) and (b) a *process* level (i.e., the extent to which students engaged in the epistemic activities proposed by Fischer et al. 2014, during reasoning).

Regarding the *content* aspect of evidence-based reasoning, we found that dyads referred significantly less often to scientific theory and evidence than individuals. Based on research on the information pooling paradigm (Wittenbaum and Stasser 1996), one possible explanation for this finding might be that dyads—despite the fact that two students should typically possess more knowledge than an individual student—might have focused their discussion on scientific knowledge they both possessed, and have been hesitant to discuss unshared knowledge (e.g., knowledge about a theory that only one learner possessed) with each other. Possible reasons for this can be manifold and include aspects such as social anxiety or social loafing, and also production blocking (i.e., listening to a partner's explanation for a problem may undermine one's own thinking about the problem; see Paulus 2000). Research from pre-service teacher education further reports that pre-service teachers often are not well-prepared for collaborating with each other (Baeten and Simons 2014; Shin et al. 2016), which might further have increased coordination demands which in turn might have consumed time that dyads could not use for discussing what scientific knowledge to apply to the case (Strijbos et al. 2004).

In contrast to the results at the content level, regarding the *process* aspect of evidence-based reasoning, we found that dyads of pre-service teachers engaged more in hypothesis generation (explaining the problem) and evidence evaluation than individuals. Moreover, dyads were also more likely to draw final conclusions and to sum up their reasoning processes. These findings are consistent with prior research on scientific reasoning (e.g., Okada and Simon 1997; Teasley

Table 5 Effects of dyadic homo-/heterogeneity on the epistemic processes problem identification, questioning, and drawing conclusions (logistic regressions)

Dependent variable	χ^2	<i>p</i>	Exp(<i>B</i>)	CI _{Exp(<i>B</i>)}	Nagelkerke's <i>R</i> ²
Problem identification	2.90	.088	0.00	0.00; 3.50	.13
Questioning	2.11	.146	0.01	0.00; 7.04	.09
Drawing conclusions	2.16	.141	124.29	0.15; 101,090.79	.09

1995) and with the many pleas for co-teaching (Baeten and Simons 2014; Nokes et al. 2008): Those studies indicate that when students solve problems together, they often want to share their understanding and thus articulate their point of view towards each other. Also, they often “scaffold” each other for such articulation by asking their reasoning partner for clarification and further elaboration on the expressed ideas. Similarly, in the context of co-teaching, researchers conclude that collaborative discussion of problem cases and teaching practices may facilitate the exchange of individual knowledge and the development of more positive attitudes towards a more reflective and evidence-based practice (Baeten and Simons 2014; Birrell and Bullough 2005; Nokes et al. 2008).

As our results show though, this reflective discussion might however come at the cost of engaging less in generating solutions to the problem. In other words, our findings indicate that there might be a trade-off between more explanatory and more solution-oriented processes during collaborative problem-solving: It appears that the lower engagement in solution generation of dyads can be explained by their enhanced engagement in hypothesis generation and evidence evaluation. One explanation for this result might be that the epistemic processes of giving explanations, referring to evidence, and summing up the discussion reflect the epistemic need to coordinate between reasoning partners to develop a better (shared) understanding of the problem (Rummel and Spada 2005). As Okada and Simon (1997) argue, in a collaborative situation, reasoners often need to be more explicit in communicating their views to make their partner understand and/or accept their ideas. As a result, an extensive engagement in these explanatory processes might then prepare the choice of an adequate solution, which then would not need to be extensively debated anymore within the group.

Based on earlier findings (Bowers et al. 2000), we further assumed that collaborative problem-solving might benefit from the heterogeneity of the problem-solving scripts of teachers who collaborated in dyads. In contrast to this assumption, our results showed that the *content* aspect of evidence-based reasoning was not affected by group composition in that respect, i.e., homo- and heterogeneous dyads did not differ in the extent to which they referred to scientific theories and evidence. Again, that heterogeneous dyads did not exploit the potentials of their dissimilarity may either have to do with factors such as social or cognitive inhibition (Paulus 2000), or with the increased coordination demands that of course affect heterogeneous dyads more than homogeneous dyads, as the latter have less reason to extensively discuss how to approach the task.

As for the *process* aspect of scientific reasoning, we however did find (albeit small) effects of heterogeneity: The more collaborative partners differed from each other regarding their problem-solving strategy, the less they discussed possible solutions to alleviate the problem. Furthermore, although dyadic heterogeneity did not have a significant impact on other epistemic processes, in tendency, the more heterogeneous the dyads are, the more they seem to engage in “explanatory behavior.” This tendency is in accordance with previous findings that dyadic heterogeneity might be beneficial in the case of complex problem-solving tasks (Bowers et al. 2000; Wiley et al. 2013), and they might explain why more heterogeneous dyads focused less on the solutions as they spent more “effort” in developing explanations to the problem. This can be regarded as evidence that heterogeneity of the problem-solving scripts of pre-service teachers can be expected to result in more reflective (explanatory and evaluative) practices. It might be speculated then if co-teaching and other collaborative approaches to pre-service teacher education would be especially effective when pre-service teachers would be grouped together based on the heterogeneity of their problem-solving scripts.

Limitations and conclusions

Even though our study revealed interesting results on the comparison of individual vs. dyadic problem-solving, as well as regarding the effects of dyadic heterogeneity, our findings are limited in several ways:

First, since the presented empirical study implemented only one problem case for the problem-solving task, the interpretation of the results might not be generalizable to pre-service teachers' reasoning on other topics. Second, our sample was recruited from just one university and thus may have had specific characteristics that may differentiate it from samples that might have been drawn from other pre-service teacher education programs. Third, in contrast to the approach taken by Canham et al. (2012), we did not experimentally manipulate the homo-/heterogeneity of our dyads. Instead, we used the natural variation of the problem-solving scripts and computed a post hoc heterogeneity index. While we would argue that our approach increased external validity of our study, it might well be that it possibly led to an underestimation of the effects of dyads' heterogeneity. It thus would be very interesting to run a subsequent study in which homo-/heterogeneity is experimentally manipulated (and possibly the degree of heterogeneity is increased compared with our study) to see whether effects of heterogeneity would increase as a result of this. Fourth, since the present research represents a lab study without a corresponding field context, ecological validity may be limited on that end. Therefore, future studies could consider addressing more authentic, and thus ecologically more valid, problem-solving scenarios (e.g., in a classroom), and also use extended samples coming from several universities. Also, it would be interesting to investigate if our results also generalize to in-service teachers, as co-teaching and similar approaches are also often used with this population. Finally, another limitation of the present study is the lack of a prior scientific content knowledge measure. As a result, we only know the extent to which our dyads were homo-/heterogeneous with respect to their problem-solving scripts, but not with respect to their prior content knowledge. Subsequent studies should thus also measure prior scientific content knowledge to disentangle the possible differential effects of within-dyad homo-/heterogeneity regarding both the content and the process dimension of scientific reasoning.

After discussing our results, the question remains under what circumstances and with respect to what goals collaboration can be a useful context to promote pre-service teachers' evidence-based reasoning when solving pedagogical problems. Our findings paint quite a differentiated picture in this respect. They suggest that having pre-service teachers collaborate during reasoning about pedagogical problems is a good decision when the goal is to foster an analytical focus during reasoning. To reach this goal, our results further suggest to form groups that are heterogeneous with respect to their problem-solving scripts. However, if the goal is to promote students' references to scientific theories and evidence, our results seem to imply that having pre-service teachers collaborate is perhaps not such a good idea, at least if groups are not further scaffolded on how to apply scientific theories and evidence on case information. Approaches that may be effective in this respect have, for example, been described by Wenglein et al. (2016) and Raes et al. (2012). For future research, an important task is to accumulate evidence on how to combine different scaffolds to effectively support pre-service teachers in both the application of scientific knowledge and a systematic engagement in scientific reasoning processes.

Acknowledgments Open Access funding provided by Projekt DEAL. We would like to thank Christian Ghanem for his support and for his collaborative participation in the coding training and reliability testing phases.

Funding information This work was supported by the Elite Network of Bavaria (project number: K-GS-2012-209).

Compliance with ethical standards

Disclaimer The design of the study, the findings, and the conclusions do not reflect the views of the funding agency.

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

- Asterhan, C. S. C., & Schwarz, B. B. (2009). Argumentation and explanation in conceptual change: indications from protocol analyses of peer-to-peer dialog. *Cognitive Science*, 33(3), 374–400. <https://doi.org/10.1111/j.1551-6709.2009.01017.x>.
- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: a proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation*. New York: Academic Press.
- Baeten, M., & Simons, M. (2014). Student teachers' team teaching: models, effects, and conditions for implementation. *Teaching and Teacher Education*, 41, 92–110. <https://doi.org/10.1016/j.tate.2014.03.010>.
- Birrell, J. R., & Bullough, R. V. (2005). Teaching with a peer: a follow-up study of the 1st year of teaching. *Action in Teacher Education*, 27(1), 72–81. <https://doi.org/10.1080/01626620.2005.10463375>.
- Bowers, C. A., Pharmer, J. A., & Salas, E. (2000). When member homogeneity is needed in work teams: a meta-analysis. *Small Group Research*, 31(3), 305–327. <https://doi.org/10.1177/104649640003100303>.
- Bullough, R. V., Young, J., Erickson, L., Birrell, J. R., Clark, D. C., Egan, M. W., Berrie, C. F., Hales, V., & Smith, G. (2002). Rethinking field experience: partnership teaching versus single-placement teaching. *Journal of Teacher Education*, 53(1), 68–80. <https://doi.org/10.1177/0022487102053001007>.
- Canham, M. S., Wiley, J., & Mayer, R. E. (2012). When diversity in training improves dyadic problem solving. *Applied Cognitive Psychology*, 26(3), 421–430. <https://doi.org/10.1002/acp.1844>.
- Charlin, B., Boshuizen, H. P. A., Custers, E. J., & Feltovich, P. J. (2007). Scripts and clinical reasoning: clinical expertise. *Medical Education*, 41(12), 1178–1184. <https://doi.org/10.1111/j.1365-2923.2007.02924.x>.
- Chen, S.-Y., Feng, Z., & Yi, X. (2017). A general introduction to adjustment for multiple comparisons. *Journal of Thoracic Disease*, 9(6), 1725–1729.
- Chi, M. T. H. (1997). Quantifying qualitative analyses of verbal data: a practical guide. *Journal of the Learning Sciences*, 6(3), 271–315. https://doi.org/10.1207/s15327809jls0603_1.
- Chi, M. T. H. (2009). Active-constructive-interactive: a conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, 1(1), 73–105. <https://doi.org/10.1111/j.1756-8765.2008.01005.x>.
- Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: linking cognitive engagement to active learning outcomes. *Educational Psychologist*, 49(4), 219–243. <https://doi.org/10.1080/00461520.2014.965823>.
- Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: a framework for memory research. *Journal of Verbal Learning & Verbal Behavior*, 11(6), 671–684.
- Csanadi, A., Kollar, I., & Fischer, F. (2015). *Internal scripts and social context as antecedents of teacher students' scientific reasoning*. Paper presented at the 16th Biennial Conference of the European Association for Research on Learning and Instruction (EARLI), Limassol, Cyprus.
- Curşeu, P. L., & Pluut, H. (2013). Student groups as learning entities: the effect of group diversity and teamwork quality on groups' cognitive complexity. *Studies in Higher Education*, 38(1), 87–103. <https://doi.org/10.1080/03075079.2011.565122>.
- European Commission [EC]. (2013). *Supporting teacher competence development for better learning outcomes*. Retrieved on the 13.09.2019 from https://ec.europa.eu/assets/eac/education/experts-groups/2011-2013/teacher/teachercomp_en.pdf.
- Field, A. (2009). *Discovering statistics using SPSS* (2nd ed.). London: Sage.

- 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., Dorner, B., Pankofer, S., Fischer, M., Srijbos, 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.
- Grissom, R. J. (2000). Heterogeneity of variance in clinical data. *Journal of Consulting and Clinical Psychology, 68*(1), 155–165. <https://doi.org/10.1037/0022-006X.68.1.155>.
- Gruber, H., Mandl, H., & Renkl, A. (2000). *Was lernen wir in Schule und Hochschule: Träges Wissen?* [What do learn in school and higher education: inert knowledge?]. In H. Mandl & J. Gerstenmeier (Eds.), *Die Kluft zwischen Wissen und Handeln: empirische und theoretische Lösungsansätze* (pp. 139–156). Göttingen: Hogrefe.
- Hetmanek, A., Wecker, C., Kiesewetter, J., Trempler, K., Fischer, M. R., Gräsel, C., & Fischer, F. (2015). *Wozu nutzen Lehrkräfte welche Ressourcen? Eine Interviewstudie zur Schnittstelle zwischen bildungswissenschaftlicher Forschung und professionellem Handeln im Bildungsbereich.* [What resources do teachers use for what? An interview study on the link between educational research and professional problem-solving in education]. *Unterrichtswissenschaft, 43*(3), 194–210.
- Jang, S.-J. (2008). Innovations in science teacher education: effects of integrating technology and team-teaching strategies. *Computers & Education, 51*(2), 646–659. <https://doi.org/10.1016/j.compedu.2007.07.001>.
- Kiemer, K. & Kollar, I. (2018). Evidence-based reasoning of pre-service teachers: a script perspective. In J. Kay, and Luckin, R. (Eds.), *Rethinking learning in the digital age: making the learning sciences count*, 13th International Conference of the Learning Sciences (ICLS) 2018, Volume 2. London: International Society of the Learning Sciences.
- King, S. (2006). Promoting paired placements in initial teacher education. *International Research in Geographical & Environmental Education, 15*(4), 370–386.
- Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science, 12*(1), 1–48. https://doi.org/10.1207/s15516709cog1201_1.
- Klein, M., Wagner, K., Klopp, E., & Stark, R. (2015). *Förderung anwendbaren bildungswissenschaftlichen Wissens bei Lehramtsstudierenden anhand fehlerbasierten kollaborativen Lernens: Eine Studie zur Replikation und Stabilität bisheriger Befunde und zur Erweiterung der Lernumgebung.* [Scaffolding application-oriented educational knowledge of pre-service teachers via failure-based collaborative learning: a study on the replication and stability of previous evidence and on an extension of a learning environment]. *Unterrichtswissenschaft, 43*(3), 225–244.
- Kolodner, J. L. (2006). Case-based reasoning. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 225–242). Cambridge: Cambridge University Press.
- Kolodner, J. L., Camp, P. J., Crismond, D., Fasse, B., Gray, J., Holbrook, J., Puntambekar, S., & Ryan, M. (2003). Problem-based learning meets case-based reasoning in the middle-school science classroom: putting Learning by Design™ into practice. *The Journal of the Learning Sciences, 12*(4), 495–547. https://doi.org/10.1207/S15327809JLS1204_2.
- Korthagen, F. A., & Kessels, J. P. (1999). Linking theory and practice: changing the pedagogy of teacher education. *Educational Researcher, 28*(4), 4–17.
- Lockhorst, D., Wubbels, T., & van Oers, B. (2010). Educational dialogues and the fostering of pupils' independence: the practices of two teachers. *Journal of Curriculum Studies, 42*(1), 99–121. <https://doi.org/10.1080/00220270903079237>.
- McNeill, K. L., & Knight, A. M. (2013). Teachers' pedagogical content knowledge of scientific argumentation: the impact of professional development on K–12 teachers. *Science Education, 97*(6), 936–972. <https://doi.org/10.1002/sce.21081>.
- Meyers, L. S., Gamst, G., & Guarino, A. J. (2006). *Applied multivariate research: design and interpretation*. London: Sage.
- Nokes, J. D., Bullough Jr., R. V., Egan, W. M., Birrell, J. R., & Hansen, M. (2008). The paired-placement of student teachers: an alternative to traditional placements in secondary schools. *Teaching and Teacher Education, 24*(8), 2168–2177. <https://doi.org/10.1016/j.tate.2008.05.001>.
- Okada, T., & Simon, H. A. (1997). Collaborative discovery in a scientific domain. *Cognitive Science, 21*(2), 109–146. [https://doi.org/10.1016/S0364-0213\(99\)80020-2](https://doi.org/10.1016/S0364-0213(99)80020-2).
- Paulus, P. (2000). Groups, teams, and creativity: the creative potential of idea-generating groups. *Applied Psychology, 49*(2), 237–262. <https://doi.org/10.1111/1464-0597.00013>.
- Petty, G. (2009). *Evidence-based teaching*. Cheltenham: Nelson Thornes.
- Piwowar, V., Barth, V. L., Ophardt, D., & Thiel, F. (2018). Evidence-based scripted videos on handling student misbehavior: the development and evaluation of video cases for teacher education. *Professional Development in Education, 44*(3), 369–384.

- Raes, A., Schellens, T., De Wever, B., Kollar, I., Wecker, C., Fischer, F., Tissenbaum, M., Slotta, J., Peters, V. L., & Songer, N. (2012). Scripting science inquiry learning in CSSL classrooms. In van Aalst, J., Thompson, K., Jacobson, M. J., & Reimann, P. (Eds.), *The Future of Learning: Proceedings of the 10th International Conference of the Learning Sciences (ICLS 2012) – Vol 2, Short Papers, Symposia, and Abstracts* p 118–125. Sydney: International Society of the Learning Sciences.
- Roschelle, J., & Teasley, S. D. (1995). The construction of shared knowledge in collaborative problem solving. In C. O'Malley (Ed.), *Computer supported collaborative learning* (pp. 69–97). Berlin: Springer. https://doi.org/10.1007/978-3-642-85098-1_5.
- Rummel, N., & Spada, H. (2005). Learning to collaborate: an instructional approach to promoting collaborative problem solving in computer-mediated settings. *Journal of the Learning Sciences*, 14(2), 201–241. https://doi.org/10.1207/s15327809jls1402_2.
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1), 68–78.
- Rytivaara, A., & Kershner, R. (2012). Co-teaching as a context for teachers' professional learning and joint knowledge construction. *Teaching and Teacher Education*, 28(7), 999–1008. <https://doi.org/10.1016/j.tate.2012.05.00>.
- Sampson, V., & Blanchard, M. R. (2012). Science teachers and scientific argumentation: trends in views and practice. *Journal of Research in Science Teaching*, 49(9), 1122–1148. <https://doi.org/10.1002/tea.21037>.
- Schank, R. C. (1999). *Dynamic memory revisited*. New York: Cambridge University Press. <https://doi.org/10.1017/CBO9780511527920>.
- Sherin, M. G., & Van Es, E. A. (2009). Effects of video club participation on teachers' professional vision. *Journal of Teacher Education*, 60(1), 20–37.
- Shin, M., Lee, H., & McKenna, J. W. (2016). Special education and general education preservice teachers' co-teaching experiences: a comparative synthesis of qualitative research. *International Journal of Inclusive Education*, 20(1), 91–107. <https://doi.org/10.1080/13603116.2015.1074732>.
- Sinclair, J., Taylor, P. J., & Hobbs, S. J. (2013). Alpha level adjustments for multiple dependent variable analyses and their applicability – a review. *International Journal of Sports Science and Engineering*, 7(1), 17–20.
- Simons, M., & Baeten, M. (2016). Student teachers' team teaching during field experiences: An evaluation by their mentors. *Mentoring & Tutoring: Partnership in Learning*, 24(5), 415–440.
- Spencer, T. D., Detrich, R., & Slocum, T. A. (2012). Evidence-based practice: a framework for making effective decisions. *Education and Treatment of Children*, 35(2), 127–151. <https://doi.org/10.1353/etc.2012.0013>.
- Springer, L., Stanne, M. E., & Donovan, S. S. (1999). Effects of small group learning on undergraduates in science, mathematics, engineering, and technology: a meta-analysis. *Review of Educational Research*, 69(1), 21–51. <https://doi.org/10.3102/00346543069001021>.
- Stark, R., Puhl, T., & Krause, U.-M. (2009). Improving scientific argumentation skills by a problem-based learning environment: effects of an elaboration tool and relevance of student characteristics. *Evaluation & Research in Education*, 22(1), 51–68. <https://doi.org/10.1080/09500790903082362>.
- Streiner, D. L., & Norman, G. R. (2011). Correction for multiple testing: is there a resolution? *Chest*, 140(1), 16–18.
- Strijbos, J. W., & Stahl, G. (2007). Methodological issues in developing a multi-dimensional coding procedure for small-group chat communication. *Learning and Instruction*, 17(4), 394–404.
- Strijbos, J.-W., Martens, R. L., Jochems, W. M. G., & Broers, N. J. (2004). The effect of functional roles on group efficiency: using multilevel modeling and content analysis to investigate computer-supported collaboration in small groups. *Small Group Research*, 35(2), 195–229. <https://doi.org/10.1177/1046496403260843>.
- Strijbos, J.-W., Martens, R. L., Prins, F. J., & Jochems, W. M. J. (2006). Content analysis: What are they talking about? *Computers & Education*, 46(1), 29–48.
- Tawfik, A. A., & Kolodner, J. L. (2016). Systematizing scaffolding for problem-based learning: a view from case-based reasoning. *Interdisciplinary Journal of Problem-Based Learning*, 10(1), 6.
- Teasley, S. D. (1995). The role of talk in children's peer collaborations. *Developmental Psychology*, 31(2), 207–220. <https://doi.org/10.1037/0012-1649.31.2.207>.
- van de Pol, J., Volman, M., & Beishuizen, J. (2011). Patterns of contingent teaching in teacher–student interaction. *Learning and Instruction*, 21(1), 46–57. <https://doi.org/10.1016/j.learninstruc.2009.10.004>.
- Voss, T., Kunter, M., & Baumert, J. (2011). Assessing teacher candidates' general pedagogical/psychological knowledge: test construction and validation. *Journal of Educational Psychology*, 103(4), 952–969. <https://doi.org/10.1037/a0025125>.
- Weinberger, A., Stegmann, K., & Fischer, F. (2007). Knowledge convergence in collaborative learning: Concepts and assessment. *Learning and Instruction*, 17(4), 416–426.
- Weinberger, A., Stegmann, K., & Fischer, F. (2010). Learning to argue online: Scripted groups surpass individuals (unscripted groups do not). *Computers in Human Behavior*, 26(4), 506–515.
- Wenglein, S., Bauer, J., Heininger, S., Prenzel, M. (2016, August). *Pre-service teachers' evidence-based argumentation competence: can a training of heuristics improve argumentative quality?* Paper presented

- at the Joint Meeting of the Special Interest Groups 20 and 26 of the European Association for Research on Learning and Instruction (EARLI). Ghent, Belgium.
- Wild, K.-P. (2006). Lernstrategien und Lernstile. In D. H. Rost (Ed.), *Handwörterbuch Pädagogische Psychologie* (pp. 427–432). Weinheim: Beltz.
- Wiley, J., & Jolly, C. (2003). When two heads are better than one expert. In R. Alterman & D. Kirsch (2017). *Proceedings of the 25th Annual Cognitive Science Society* (pp. 1242–1246) New Jersey: Lawrence Erlbaum Associates. <https://doi.org/10.4324/9781315799360>.
- Wiley, J., Goldenberg, O., Jarosz, A. F., Wiedmann, M., & Rummel, N. (2013). Diversity, collaboration, and learning by invention. In M. Knauff (Ed.), *Proceedings of the 35th Annual Meeting of the Cognitive Science Society* (pp. 3765–3770). Austin, TX: Cognitive Science Society.
- Wittenbaum, G. M., & Stasser, G. (1996). Management of information in small groups. In J. L. Nye & A. M. Brower (Eds.), *What's social about social cognition? Research on socially shared cognition in small groups* (pp. 3–28). Thousand Oaks, CA: Sage.
- Wu, D., Liao, Z., & Dai, J. (2015). Knowledge heterogeneity and team knowledge sharing as moderated by internal social capital. *Social Behavior and Personality: An International Journal*, 43(3), 423–436. <https://doi.org/10.2224/sbp.2015.43.3.423>.
- Yeh, C., & Santagata, R. (2015). Preservice teachers' learning to generate evidence-based hypotheses about the impact of mathematics teaching on learning. *Journal of Teacher Education*, 66(1), 21–34. <https://doi.org/10.1177/0022487114549470>.
- Zottmann, J. M., Stegmann, K., Strijbos, J. W., Vogel, F., Wecker, C., & Fischer, F. (2013). Computer-supported collaborative learning with digital video cases in teacher education: The impact of teaching experience on knowledge convergence. *Computers in Human Behavior*, 29(5), 2100–2108.

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

Andras Csanadi. Ludwig-Maximilian University of Munich, Munich, Germany, and University of Bundeswehr Munich, Neubiberg, Germany. E-mail: csaoand@gmail.com, andras.csanadi@unibw.de

Current research themes:

Computer-supported collaborative learning. Measurement of reasoning processes. Problem-solving scripts. Quantifying qualitative data

Most relevant publications in the field of Psychology of Education:

Csanadi, A., Eagan, B., Kollar, I., Shaffer, D.W., & Fischer, F. (2018). When coding-and-counting is not enough: using epistemic network analysis (ENA) to analyze verbal data in CSCL research. *International Journal of Computer-Supported Collaborative Learning*, 13(4). 419-438.

Ingo Kollar. University of Augsburg, Augsburg, Germany. E-mail: ingo.kollar@phil.uni-augsburg.de

Current research themes:

Digital teaching competences. Evidence-based teaching. Scientific reasoning. Scripted learning. Self- and co-regulated learning

Most relevant publications in the field of Psychology of Education:

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

Ghanem, C., Kollar, I., Fischer, F., Lawson, T. & Pankofer, S. (2018). How do social work novices and experts solve professional problems? A micro-analysis of epistemic activities and knowledge utilization. *European Journal of Social Work*, 21(1), 3-19.

- Kollar, I., Ufer, S., Reichersdorfer, E., Vogel, F., Fischer, F. & Reiss, K. (2014). Effects of heuristic worked examples and collaboration scripts on the acquisition of mathematical argumentation skills of teacher students with different levels of prior knowledge. *Learning and Instruction, 32*, 22-36.
- Kollar, I., Fischer, F. & Slotta, J. D. (2007). Internal and external scripts in computer-supported collaborative inquiry learning. *Learning and Instruction, 17*(6), 708-721.
- Schwaighofer, M., Vogel, F., Kollar, I., Ufer, S., Strohmaier, A., Terwedow, I., Ottinger, S., Reiss, K., & Fischer, F. (2017). How to combine collaboration scripts and heuristic worked examples to foster mathematical argumentation – when working memory matters. *International Journal of Computer-Supported Collaborative Learning, 12*(3), 281-305. doi:10.1007/s11412-017-9260-z
- Vogel, F., Kollar, I., Ufer, S., Reichersdorfer, E., Reiss, K. & Fischer, F. (2016). Developing argumentation skills in mathematics through computer-supported collaborative learning: the role of transactivity. *Instructional Science, 44*(5), 477-500.
- Vogel, F., Wecker, C., Kollar, I., & Fischer, F. (2017). Socio-cognitive scaffolding with collaboration scripts: a meta-analysis. *Educational Psychology Review, 29*(3), 477-511.
- Wang, X., Kollar, I. & Stegmann, K. (2017). Adaptable scripting to foster regulation processes and skills in computer-supported collaborative learning. *International Journal for Computer-Supported Collaborative Learning, 12*(2), 153-172.

Frank Fischer. Ludwig-Maximilian University of Munich, Munich, Germany. E-mail: frank.fischer@psy.lmu.de

Current research themes:

collaborative learning, evidence-based practice, fostering diagnostic competence, learning with digital media, scientific reasoning and argumentation

Most relevant publications in the field of Psychology of Education:

- Berndt, M., Strijbos, J.-W., & Fischer, F. (2018). Effects of written peer-feedback content and sender's competence on perceptions, performance, and mindful cognitive processing. *European Journal of Psychology of Education, 33*(1), 31-49. doi:10.1007/s10212-017-0343-z
- Bolzer, M., Strijbos, J. W., & Fischer, F. (2015). Inferring mindful cognitive-processing of peer-feedback via eye-tracking: the role of feedback-characteristics, fixation-durations and transitions. *Journal of Computer Assisted Learning, 31*(5), 422-434. doi:10.1111/jcal.12091
- Eberle, J., Stegmann, K., & Fischer, F. (2014). Legitimate peripheral participation in communities of practice. Participation support structures for newcomers in faculty student councils. *Journal of the Learning Sciences, 23*(2), 216-244. doi:10.1080/10508406.2014.883978
- Engelmann, K., Neuhaus, B. J., & Fischer, F. (2016). Fostering scientific reasoning in education - meta-analytic evidence from intervention studies. *Educational Research and Evaluation*. doi:10.1080/13803611.2016.1240089
- Fischer, F., Kollar, I., Stegmann, K., & Wecker, C. (2013). Toward a script theory of guidance in computer-supported collaborative learning. *Educational Psychologist, 48*(1), 56-66.
- Fischer, F., & Järvelä, S. (2014). Methodological advances in research on learning and instruction and in the learning sciences. *Frontline Learning Research, 2*(4), 1-6.
- Fischer, F., Kollar, I., Ufer, S., Sodian, B., Hussmann, H., Pekrun, R., et al. (2014). Scientific reasoning and argumentation: advancing an interdisciplinary research agenda in education. *Frontline Learning Research, 2*(3), 28-45.
- Förtsch, C., Sommerhoff, D., Fischer, F., Fischer, M. R., Girwidz, R., Obersteiner, ... & 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*, 207. doi:10.3390/educsci8040207
- Ghanem, C., Kollar, I., Fischer, F., Lawson, T. R., & Pankofer, S. (2016). How do social work novices and experts solve professional problems? A micro-analysis of epistemic activities and the use of evidence. *European Journal of Social Work, 1*-17. doi: 10.1080/13691457.2016.1255931
- Heitzmann, N., Fischer, F. & Fischer, M. R. (2017). Worked examples with errors: when self-explanation prompts hinder learning of teachers' diagnostic competences on problem-based learning. *Instructional Science, 46*(2), 245-271. doi:10.1007/s11251-017-9432-2

- Opitz, A., Heene, M., & Fischer, F. (2017). Measuring scientific reasoning - a review of test instruments. *Educational Research and Evaluation*, 23(3-4), 78-101. doi:10.1080/13803611.2017.1338586
- Schwaighofer, M., Vogel, F., Kollar, I., Ufer, S., Strohmaier, A., Terwedow, I., Ottinger, S., Reiss, K., & Fischer, F. (2017). How to combine collaboration scripts and heuristic worked examples to foster mathematical argumentation – when working memory matters. *International Journal of Computer-Supported Collaborative Learning*, 12(3), 281-305. doi:10.1007/s11412-017-9260-z
- Schwaighofer, M., Fischer, F., & Bühner, M. (2015). Does working memory training transfer? A meta-analysis including training conditions as moderators. *Educational Psychologist*, 50(2), 138–166. doi:10.1080/00461520.2015.1036274
- Sommerhoff, D., Szameitat, A., Vogel, F., Chernikova, O., Loderer, K., & Fischer, F. (2018). What do we teach when we teach the learning sciences? A document analysis of 75 graduate programs. *Journal of the Learning Sciences*, 27(2), 319-351. doi:10.1080/10508406.2018.1440353
- Schwaighofer, M., Bühner, M., & Fischer, F. (2016). Executive functions as moderators of the worked example effect: when shifting is more important than working memory capacity. *Journal of Educational Psychology*, 108(7), 982–1000. doi:10.1037/edu000011
- Vogel, F., Wecker, C., Kollar, I., & Fischer, F. (2017). Socio-cognitive scaffolding with collaboration scripts: a meta-analysis. *Educational Psychology Review*. 29(3), 477-511.
- Wecker, C., & Fischer, F. (2014). Where is the evidence? A meta-analysis on the role of argumentation for the acquisition of domain-specific knowledge in computer-supported collaborative learning. *Computers & Education*, 75, 218-228. doi:10.1016/j.compedu.2014.02.016

Affiliations

Andras Csanadi^{1,2} · Ingo Kollar³ · Frank Fischer¹

¹ Ludwig-Maximilian University of Munich, Munich, Germany

² University of Bundeswehr Munich, Neubiberg, Germany

³ University of Augsburg, Augsburg, Germany