Scientific Reasoning and Problem Solving in a Practical Domain: Are Two Heads Better Than One?

Andras Csanadi, Ludwig Maximilian University of Munich, andras.csanadi@psy.lmu.de
Ingo Kollar, University of Augsburg, ingo.kollar@phil.uni-augsburg.de
Frank Fischer, Ludwig Maximilian University of Munich, frank.fischer@psy.lmu.de

Abstract: To meet high-level problem solving standards, practitioners should optimally solve problems in an evidence-based manner: by using scientific knowledge while engaging in scientific reasoning processes. It can be argued that professionals do this better in groups than individually. Also, heterogeneous groups might have more potential to engage in higher levels of scientific reasoning than homogeneous groups or individuals. The present study investigated these questions in the context of teacher education. 76 teacher students solved a problem case from their future practice either individually or in dyads that were homo- or heterogeneous with respect to their members' problem solving scripts. The results show that although dyads have an advantage on generating hypotheses to explain the problem; individuals engage more in generating solutions. Moreover, especially heterogeneous dyads seem to differ from individuals' approaches. Future studies could investigate further how to use groups' potential in generating hypotheses to facilitate them in generating solutions.

Keywords: scientific reasoning, collaborative problem solving, group composition.

Introduction

Practitioners need to solve and reason on problems in their work on an everyday basis. It can be argued that in many domains, the kinds of reasoning processes that underlie competent solutions can be conceptualized as scientific reasoning. To make appropriate practical decisions, practitioners need to refer to relevant scientific knowledge from their domain and engage in reasoning and argumentation processes that are similar to the processes scientists in the particular domain would undertake (Fischer et al., 2014). However, utilizing scientific knowledge in a science-like way for solving complex problems is frequently a problem for practitioners (Gruber, Mandl & Renkl, 2000).

One possible way to support practitioners during problem solving is by allocating them into groups, i.e. to realize a collaborative problem solving setting (Rummel & Spada, 2005). A general notion is that groups tend to outperform single individuals on inquiry tasks (Lazonder, 2005). Collaborative partners, for example, can engage more in explaining and interpreting a phenomenon (Teasley, 1995), can facilitate better the regulation of the process of inquiry (Lazonder, 2005) and coordinate their ideas in a way that leads to a more abstract or deeper understanding of the problem (Schwartz, 1995). However, group settings do not guarantee better performance (Weinberger, Stegmann & Fischer, 2010). One way to explain these unequivocal findings is by considering group composition as a moderator of the effectiveness of collaborative learning. Group members' understanding on how they should reason during problem solving is guided by their knowledge or problem solving scripts (Fischer, Kollar, Stegmann & Wecker, 2013). The more diverging group members' views are on what processes and knowledge to build the own problem solving process on, the more creative they might be while solving the problem. However, introducing new perspectives might also increase the need for coordination in order to establish a joint problem space (Roschelle & Teasley, 1995).

The main aim of this study is to investigate whether scientific reasoning of future practitioners who solve an authentic problem from their practice differs if they work (1) individually or as a group; (2) as a homogeneous or as a heterogeneous group. These questions are investigated in the profession of teaching.

Solving practical problems as scientific reasoning

Based on the partially diverse conceptualizations of scientific reasoning (Fischer et al., 2014), we differentiate two aspects of scientific reasoning: the process and the content level (Zimmerman, 2000). At the *process* level, scientific reasoning can be viewed as an inquiry process (Klahr & Dunbar, 1988) that is characterized by an engagement in certain epistemic activities while solving a problem (Fischer et al., 2014). As the framework of Fischer et al. (2014) suggests, practitioners might engage in the same epistemic activities as scientists do for the sake of solving a practical problem. First, they need to identify the problem itself (Problem Identification); develop directions that target their further exploration on the problem (Questioning); consider potential explanations of

the problem (Hypothesis generation); take into account or generate further information necessary to understand or solve the problem (Evidence Generation); evaluate the information in the context of their hypotheses (Evidence Evaluation); generate solutions (Constructing artefacts); engage in discussions with others to re-evaluate their thoughts (Communicating and scrutinizing) and sum up their process to come to well-warranted conclusions on how to solve the problem (Drawing conclusions).

At the *content* level, we can look to what extent reasoners build on relevant theoretical concepts and empirical scientific findings of a domain during their problem-solving process. For example, teachers and teacher students should be able to apply teaching methods, theories of learning and relevant research findings from the Learning Sciences in order to solve a given problem (Voss, Kunter & Baumert, 2011), such as dealing with an underperforming student.

Collaborative problem solving and scientific reasoning

People reasoning together bring different perspectives on how to solve a given problem and have the opportunity to share their ideas (Hesse, Care, Buder, Sassenberg & Griffin, 2015). Groups have the potential of better and more innovative reasoning processes as reasoning partners may stimulate each others' cognitive processes (Paulus, 2000). Collaborative reasoning gives the opportunity to critically evaluate each other's as well as one's own ideas (Asterhan & Schwarz, 2009). Teasley (1995) demonstrated that dyads' typically produced more interpretive talk and hypothesis generation than individuals; in contrast, individuals engaged in more descriptive talk. Similarly, Okada and Simon (1997) found that reasoning with a partner resulted in higher engagement in explanatory processes (eg. hypothesis generation) compared to individual reasoning. One reason for such heightened engagement in explanatory behavior may be the need to communicate in a more explicit manner (Okada & Simon, 1997). Another is to think about the limitations of each learners' ideas, via challenging and defending one's own ideas (Asterhan & Schwarz, 2009). More recent studies (Métrailler, Reijnen, Kneser & Opwis, 2008) also confirmed these findings showing that dyads hypothesize, challenge and justify ideas more often than individuals do. Such dialectic conversational turns occurring in a collaborative scenario might be beneficial for understanding (Schwartz, 1995), performing (Teasley, 1995) and learning (Asterhan & Schwarz, 2009).

On the other hand, groups typically underperform to their potential, a phenomenon identified as productivity/process loss (Weinberger, Stegmann & Fischer, 2010). Several factors have been identified accounting for process losses in collaborative vs. pooled performance settings (Paulus, 2000). Some authors (Weinberger et al., 2010) see process loss a group coordination problem. Also, several cognitive and social factors such as production blocking, cognitive load, downward social comparison, social loafing or evaluation apprehension might account for moderating collaborative performance (Paulus, 2000).

Collaborative problem solving in homogeneous vs heterogeneous groups

One aspect that might explain the ambiguous findings on whether groups outperform individuals or not, is group composition. For example, heterogeneous groups are expected to bring wider perspectives to the interaction than homogeneous groups (Paulus, 2000). A meta-analysis by Bowers, Pharmer and Salas (2000) concludes that although there are mixed results, heterogeneous teams perform better on complex problem solving tasks. More recent empirical research is in accordance with their findings. Studies (Canham, Wiley & Mayer, 2012; Wiley & Jolly, 2003) indicate that while homogeneous dyads perform better on routine-like tasks, heterogeneous dyads have an advantage when applying knowledge to solve novel problems. While heterogeneous group setting might pose higher coordination demands on its members, these demands can also lead to more elaborated reasoning (Teasley, 1995), for the benefit of problem solving. However, beyond the scattered findings (Canham et al., 2012), there is yet much light to shed on how reasoning processes are affected by group composition.

If group composition can indeed account for the ambiguous performance of groups, the question is what attributes of group members should we consider? This study is specifically concerned with individuals' knowledge about problem-solving in the domain in which they are situated. Such knowledge that directs collaborating partners' understanding and behavior in a situation can be described as "scripts" (Fischer, et al., 2013; Schank, 1999). While one learner may understand problem solving as the engagement in multiple epistemic activities in a given sequence, another learner's problem-solving script may include only some activities and even in a very different order. Reasoning partners with such diverse knowledge structures or problem solving scripts might stimulate each other, but may also face coordinational constraints and invest more effort to develop a common ground (Roschelle & Teasley, 1995). So far, however, there is a lack of empirical research investigating the effect of group composition with respect to the members' problem solving scripts.

Research questions

Utilizing scientific knowledge in a science-like way for solving complex problems is frequently a problem for future practitioners (Gruber, Mandl & Renkl, 2000). Collaborative problem solving is a highlighted cross-domain competence (e.g., Hesse et al., 2015; Rummel & Spada, 2005). While studies (e.g., Okada and Simon, 1997) indicate the advantage of collaboration on the engagement in scientific reasoning processes, groups frequently face process losses (Weinberger, Stegmann & Fischer, 2010). Moreover, research illuminating the role of homo- vs. heterogeneous collaboration on scientific reasoning processes is scarce.

These considerations led us to ask two research questions to be addressed in this paper:

RQ1: To what extent do individuals and dyads of future practitioners differ in engaging in scientific reasoning processes and utilizing scientific knowledge during problem solving?

RQ2: Does dyadic composition (homo- vs. heterogeneity) with respect to problem solving scripts affect future practitioners' scientific reasoning processes and scientific knowledge utilization during problem solving?

Methods

Participants and design

76 teacher education students (59 female, 17 male, $M_{age} = 21.22$, $SD = 3.98$) from a German university participated in the study. Each participant was randomly assigned to either an individual (16 students, $M_{age} = 22.31$; $SD = 6.73$) or a dyadic (60 students, i.e. 30 dyads, $M_{age} = 20.93$; $SD = 2.85$) condition. Within the dyadic condition, 22 students (11 dyads) reasoned in homogeneous groups ($M_{age} = 20.32$; $SD = 1.84$), i.e. collaborated with a learning partner with a similar problem solving script, while 38 students (19 dyads) reasoned in heterogeneous groups ($M_{age} = 21.29$; $SD = 3.28$), i.e. collaborated with a partner with a dissimilar problem solving script (for the operationalization of homo-/heterogeneity, see below).

Procedure

The study consisted of four steps. In the first three steps students participated individually; in the fourth step, they participated according to their condition: either individually or in dyads. Firstly, after the students arrived individually or in pairs to the study, they filled out an informed consent and a survey on demographic variables. Secondly, they were given a card sorting task to measure their problem solving scripts (see below). Thirdly, they read five printed out presentation slides (took about 5 minutes) that carried scientific content information originating from their introductory psychology class and encompassed short descriptions of theories and concepts (e.g., on strategic use of short-term memory and a classification of learning strategies). Fourthly, participants were presented an authentic problem from their (future) professional practice, which was: "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 spends enough time on her homework and studies. You as a teacher, please find possible reasons and maybe solutions to the problem". Participants were asked to find reasons and possible solutions (individually or as pairs depending on the condition) for this problem in about 10 minutes. Data from this problem-solving process were used to measure reasoners' engagement in the epistemic activities of scientific reasoning and their use of scientific content. The whole data collection took about one hour.

Dependent variables

Dependent variables were collected during the problem-solving phase that had students (dyads or individuals) solve an authentic problem case from their future professional practice (see above). To make scientific reasoning processes visible, we asked students in the individual condition to think aloud and students in the dyadic conditions to verbally discuss how they would solve the problem. All verbal data were transcribed and segmented into propositional units (Chi, 1997). The proportion of agreement between two segmenters showed reliability with a lower bound of 79.73% (Coder 1) and an upper bound of 85.09% (Coder 2; Strijbos, Martens, Prins & Jochems, 2005). We developed coding schemes to assess students' engagement in (a) epistemic activities of scientific reasoning and (b) scientific content use.

Epistemic activities

All segments were coded with the aid of a coding scheme developed to capture the eight epistemic activities of scientific reasoning proposed by Fischer et al. (2014). Since it proved impossible to reliably separate "evidence generation" from "evidence evaluation", we merged these two categories into one: evidence evaluation. After

coding each segment, the numbers of segments that fell in the same category were summed for each epistemic activity (problem identification, questioning, hypothesis generation, constructing artefacts, evidence evaluation, drawing conclusions, communicating and scrutinizing) and for every transcript (individual or dyadic). The resulting sum scores were included in the statistical analyses. Two independent coders coded 10% of the data for the identification of epistemic activities. Inter-rater reliability was sufficient (Cohen's $\kappa = .68$).

Scientific content use

A second coding scheme was developed to capture for each segment whether or not participants used scientific content. Segments were coded as "scientific content" if speakers referred to scientific theories, concepts or methods. Specifically, we used the following five categories: "Learning strategy" was applied when participants mentioned learning strategy or memory coding-retrieval related topics. "Anxiety" was applied when participants referred to test anxiety or emotional pressure. "Motivation" was used when participants talked about motivation. As the last scientific content code we used the category "Other" such as "Self-fulfilling prophecy"; "Mobbing"; "Mind-map". If the above codes did not apply, the segment was coded as non-scientific content related. For each transcript (individual or dyadic) we merged all content categories to calculate engagement in scientific content use. 5% of the segments were coded by two independent coders with an agreement of Cohen's $\kappa = .82$.

Independent variables

Learning setting: Collaborative vs. individual reasoning

Learning setting was varied by assigning students to the problem-solving phase either as individuals or dyads.

Dyadic composition: Homo- vs. heterogeneity of problem solving scripts within dyads

We defined homo- and heterogeneous dyadic composition by comparing dyadic members' problem solving scripts that were measured during the card sorting task that preceded the problem-solving phase. During the card sorting phase, students were first presented the practice-related problem case described above. Then they were asked to use a set of prefabricated activity cards available on a MS PowerPoint slide to indicate what (epistemic) activities they would perform while solving the presented problem. The eight epistemic activities from Fischer et al. (2014) and five additionally selected activities (e.g., "Giving feedback", "Improvising") were written on the activity cards that were presented to the participants. Besides, five blank cards were provided to give participants the opportunity to note down further activities if they wanted to. Resulting activity sequences were then coded in the following way: We summed those epistemic activity cards that represented activities 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 inquiry. Then, we calculated a disagreement on sequentiality index (DSI) between dyadic members by calculating how many activities would need to be switched in position so as both members' selection shows the same sequence. Then, we calculated a pooled knowledge component index (PKCI) on scientific inquiry by summing the number of epistemic activities of the Fischer et al. (2014) model that at least one dyadic member had selected. Finally, a homogeneity index was calculated as (SKCI - DSI)/PKCI to account for agreements and, at the same time, controlling for disagreements between dyadic members. A median split on the resulting values divided the sample in homo- and heterogeneous groups: dyads above a homogeneity index above .50 ($N = 11$) were considered as homogeneous groups, dyads with a value of .50 ($N = 9$) or below ($N = 10$) were considered heterogeneous dyads.

Results

Quantitative analysis

Overall, dyads ($M = 138.50$, $SD = 40.16$) talked more than individuals ($M = 88.19$, $SD = 27.74$) did, $F(1, 44) = 19.93$, $p < .001$, partial $\eta^2 = .31$. Thus, total talk was used as a covariate in all subsequent analyses. To investigate RQ1, we calculated a MANCOVA that included learning setting (individual vs. dyadic) as independent variable and all the eight epistemic activities as dependent variables. This MANCOVA revealed a main effect of learning setting, $V = .33$, $F(7, 37) = 2.59$, $p < .05$, partial $\eta^2 = .33$. Follow-up ANCOVAs revealed significant effects of learning setting on the engagement in hypothesis generation, $F(1, 43) = 5.68$, $p < .05$, partial $\eta^2 = .12$; as well as a close-to-significant effect on the engagement in solution generation $F(1, 43) = 3.35$, $p = .07$, partial $\eta^2 = .07$: Dyads engaged more in hypotheses generation (adj. $M = 29.18$, $SE = 2.24$) than individuals (adj. $M = 19.03$, $SE = 3.25$). Yet, individuals engaged longer in solution generation (adj. $M = 48.41$, $SE = 5.5$) than dyads (adj. $M = 35.21$, $SE = 3.8$). Regarding scientific content use, an ANCOVA with learning setting as independent variable and

the engagement in scientific content use as dependent variable revealed no significant effect of learning setting on the engagement in scientific content use, $F(1, 43) = .04$, $p = .84$.

Regarding RQ 2, we preliminarily found a significant main effect the of dyadic composition on total talk, $F(2, 43) = 10.35$, $p < .001$, partial $\eta^2 = .33$. Subsequent post hoc tests revealed that both homogeneous ($M = 146.55$, $SD = 38.47$) and heterogeneous ($M = 133.84$, $SD = 41.39$) dyads talked more than individuals ($M = 88.19$, $SD = 27.74$) did, $p < .001$. Yet, there was no difference between homogeneous and heterogeneous dyads in total talk, $p = .36$. In light of these results, amount of total talk was used as a covariate for all analyses regarding RQ 2. To investigate RQ2, a MANCOVA with dyadic composition as independent variable, all epistemic activities as dependent variables and total talk as covariate revealed a significant multivariate main effect of dyadic composition on the engagement in epistemic activities, $V = .56$, $F(14, 74) = 2.04$, $p < .05$, partial $\eta^2 = .28$. Follow-up ANCOVAs showed significant effects of dyadic composition on the engagement in hypothesis generation, $F(2, 42) = 3.85$, $p < .05$, partial $\eta^2 = .16$; and in solution generation, $F(2, 42) = 4.87$, $p < .05$, partial $\eta^2 = .19$. LSD pairwise comparisons showed that heterogeneous dyads (adj. $M = 31.21$, $SE = 2.67$) engaged more in hypothesis generation than individuals (adj. $M = 19.33$, $SE = 3.22$), $p < .01$. Moreover, heterogeneous dyads (adj. $M = 29.39$, $SE = 4.31$) engaged significantly less in solution generation than homogeneous dyads (adj. $M = 46.52$, $SE = 5.85$) and individuals (adj. $M = 47.56$, $SE = 5.22$), $p < .05$. No other differences were significant, $p \geq .11$. Finally, an ANCOVA including total talk as covariate showed no effect of dyadic composition on scientific content use, $F(2, 42) = .06$, $p = .94$.

Qualitative analysis

The quantitative analysis revealed a different focus between dyads and individuals: dyads engaged longer in hypothesis generation and less in solution generation than individuals did. The comparison of heterogeneous dyads, homogeneous dyads and individuals led to the observation that the main differences were present between heterogeneous groups and individuals Therefore, in the following, we are presenting examples from a (heterogeneous) dyadic discourse and an individual think aloud in order to briefly introduce the special characteristic of dyadic verbalizations that might have led to a different focus from the individual think alouds (names occurring in the excerpts are fictitious). The dyadic discussion stems from the conversation of a female (age = 20, elementary school teacher-to-be) and a male (age = 21, high school teacher-to-be) participant. The individual think aloud comes from a female (age = 18, high school teacher-to-be).

Excerpt 1: Heterogeneous dyadic reasoning

Sarah has initiated the conversation after about 10 seconds of the recording. After 30 seconds she is starting expressing her explanations for the problem case:

Sarah:	So I could imagine, that maybe it's exam fever that she has… when she is writing it on the paper, and maybe she cannot ask when she does not understand something or…
Tim:	As for me, I think one difficulty, [as I am] just reading, could even be a wrong learning strategy. She seems to study at home. However, for a teacher, it is not always totally visible how the student learns when it is a big class. Yes, I would also start with first, to think it over, where exactly the difficulties can occur. So what could be responsible for it after all.

Sarah has come up with the idea of exam fever and exemplifies it with a hypothetical episode how she thinks exam fever might influence the students' performance. Tim refers to the content of the slides that describe learning strategies and introduces a second plausible hypothesis. He evaluates the case information they have about the student by arguing with possible limitations of the observations. Then, Tim further opens the discussion on hypotheses by moving one step back in exploration: "what could be responsible for it after all". By these moves, Tim has changed the reference of exploration from Sarah's initial idea on exam fever. Indeed, Sarah continues the conversation from this altered point:

Sarah:	Yes, maybe the case is that she learns only to short term memory at home, and she tries to learn a large amount of the materials again short before the exam, and with that a lot of time is passing by… well, is lost. That's also apparently why the parents say that she learns fluently, and that is however,

<blockquote>
not so effective. When she is sitting at the exam, she is confused or is forgetting it again.
</blockquote>

Tim: Ok, how will we try to find it out?

Sarah elaborates on the idea that Tim has brought up beforehand and she also takes into account case evidence to justify her point. This accommodation regarding the topic and reasoning may be a coordination attempt to reduce ambiguity and to develop shared understanding. Tim is turning back to the original question, how they could make sure about the reason for the problem, causing a temporary confusion for Sarah who cannot answer his question. Shortly after the previous episode she comes up with an idea on a solution for the problem:

Sarah: Maybe you can also see that you describe another learning strategy to her, and look if she learns differently for the following exam, and maybe it turns out indeed better.

Tim: Okay, so to reflect then, if it has brought something. Yeah, ok.

[~40 seconds talk on the ideas how to solve the problem]

Tim: I think, what would be very important too, to go in the direction, if she might have totally different reasons, personal or so, yeah.

The initial problem solving phase here breaks by jumping back to the unclosed hypothesis generation phase. Tim's continuous turning back to the uncertainty of the explanations of the problem is a critical characteristic of this conversation as it does not let ideas for solutions evolve. After this episode, they continue discussing possible explanations for the problem and planning ways to generate more evidence to be able to find out the "real" explanation. Only in the very end, they reach the point of solution generation and do not elaborate on that longer than 45 seconds until the end of the discussion:

Sarah: Or as a teacher you can also try to give feedback to the student, that she does not need to be anxious from the writing, because her oral performance is very, very good, and [this way] you may take a bit the anxiety from the writing away, and say: try to imagine yourself, that I set the questions to you orally, and you write then down what you would say otherwise.

Tim: Yeah, you can also, if it is… it can depend also on the time pressure. [It] would also be a possibility. That you then simply take it away by letting all the students write longer. I cannot let [only] one student write longer.

The overall conversation took about seven and a half minutes and it seems that the dyad could not establish a common ground for further collaborative engagement in solution generation. Challenging questions seemed to lead to enhanced productivity from one side (considering engagement in hypothesis generation), but it seemed to hinder constructive collaboration (building up on each other's initial ideas) for solution generation.

Excerpt 2: Individual reasoning
Lisa starts 20 seconds after the recording and immediately considers explanations for the problem case:

Lisa: Well, I think that the problem for sure does not lie in learning, rather maybe in the pressure that sometimes comes from the parents, or from friends, or from external influences. So it can be that the student is simply anxious, she has exam fever, and at all of the exam situations she feels herself overloaded, and maybe from this pressure from the parents or from friends who have better grades than she, [she] is pressed or feels pressed. Of course, it can also be that her learning strategy is not really good, rather she learns like, she memorizes, and with that she has not understand it well, rather – like it is nowadays said – bulimic learnt. She learns everything and forgets everything.

During the first minute of her talk Lisa generates two potential hypotheses about the student's problem and gives brief explanations about these ideas. She seems to consider both ideas as equally possible ("Of course, it can also be…") without weighting their plausibility or, in contrast with Sarah and Tim, at least considering further information collection so as to be able to reason in a more evidence-based manner. In fact, her ideas are merely descriptive, consisting of claims and clarifications without engaging in deeper reasoning. After one minute she continues with solution generation:

> Lisa: Therefore, it could make sense […] to talk with the student about certain learning strategies, what options are there, how can someone really keep the stuff in the head, and not to forget it again, so how it is anchored in the mind. Those were such strategies…uhm…yeah, where you like, where she herself, what is really important, has a motivation to learn. That means that she is interested in the topic.

During the second minute, Lisa tries to solve the problem by building on her second hypothesis (learning strategy problem). When thinking about strategies for deeper knowledge processing, she mentions that the student should be motivated, should realize the value of the topic. This does not sound like a consequent solution to the problem of "bulimic learning" though, especially because the engagement or motivation of the student was not mentioned by Lisa earlier, when she was thinking of the potential explanations for the problem. It is a question though, whether engagement in deeper reasoning during the hypothesis generation phase would have triggered the appearance of additional ideas such as "motivation" and, consequently, lower the risk of incoherency between hypotheses and solutions.

In the remaining about five and a half minutes, Lisa's focus is on how to solve the problem. Although through some segments she is examining the potential of involving other parties (parents, school psychologist) to find out more information about the student and what her problem might be, this happens in the context of solutions. As a matter of fact, 75% of the remaining segments were coded as solution generation.

Conclusions

This study aimed to address two questions. Our first research question targeted if collaborative or individual setting is more beneficial for engagement in scientific reasoning processes and knowledge utilization. Our second research question was concerned with whether heterogeneous or homogeneous dyads show different engagement in scientific reasoning processes and knowledge utilization.

Quantitative analysis of the scientific reasoning processes showed a different pattern for collaborative vs individual reasoning. Dyads engaged in hypothesis generation longer (Okada & Simon, 1997), but at the same time in solution generation shorter than individuals. Separating dyads based on homogeneity vs heterogeneity of dyadic members' scripts proved to be meaningful. Heterogeneous dyads showed longer engagement in hypothesis generation but to a less extent in solution generation compared to individuals and homogeneous groups. However, to investigate where this difference originates from, we compared excerpts of a heterogeneous dyadic discourse with an individuals' think aloud.

The qualitative analysis demonstrated that the pattern of the amount of talk found in the quantitative analysis reflected also on the time of talk participants discussed hypotheses and solutions. Moreover, the way they used case evidence or planned the collection of further evidence was also in accordance with such an epistemic focus. We believe that dyads' focus on hypothesis generation can at least partially be explained by a coordination process that is necessary, especially for heterogeneous groups, to establish a common ground (Roschelle & Teasley, 1995; Schwartz, 1995). Coordination demands emerged in the initial stage of the conversation when participants also dealt with the explanations of the problem. Developing hypotheses may particularly afford the opportunity for developing shared meaning in dyads. Earlier studies also concluded that constructive engagement in hypothesis generation or experimentation might originate in the aim of knowledge co-construction (Okada & Simon, 1997).

Besides the advantage of dyads in more elaborated engagement in hypothesis generation, their solutions seemed to be shorter. Yet, they showed more coherence with their initial ideas than the individual excerpts. It is a matter of further analysis to investigate whether dyadic solutions - although they were shorter - showed more epistemic stability in terms that they emerged from well-evaluated and established hypotheses.

In the context of professional development, evidence-based collaborative problem solving is an important skill to acquire (Rummel & Spada, 2005). Engaging teacher students in complex problem solving processes in a collaborative setting might be beneficial to develop such skills. Furthermore, problem solving can benefit from

the diversity of collaborating partners if collaboration is appropriately structured (Rummel & Spada, 2005). For such efforts, it seems to be important to establish a grounding phase where participants can explore the problem, develop explanations and evaluate their own ideas (Asterhan & Schwarz, 2009). However, collaborators might need further scaffolding on their engagement in a solution generation phase, to build on the productive advantage of their initial exploration phase. Also, for teachers, who solve problems individually, it is important to reason critically enough while thinking of potential explanations of a problem and do not necessarily jump into conclusions by jeopardizing coherency and validity of their ideas.

References

Asterhan, C. S., & 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.

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.

Canham, M. S., Wiley, J., & Mayer, R. E. (2012). When diversity in training improves dyadic problem solving. *Applied Cognitive Psychology, 26*(3), 421-430.

Fischer, F., Kollar, I., Stegmann, K., & Wecker, C. (2013). Toward a Script Theory of Guidance in Computer-Supported Collaborative Learning. *Educational Psychologist, 48*(1), 56-66.

Fischer, F., Kollar, I., Ufer, S., Sodian, B., Hussmann, H., Pekrun, R., Neuhaus, B., Dorner, B., Pankofer, S., Fischer, M., 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.

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

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

Klahr, D., & Dunbar, K. (1988). Dual Space Search During Scientific Reasoning. *Cognitive Science, 12*, 1-48.

Lazonder, A. W. (2005). Do two heads search better than one? Effects of student collaboration on web search behaviour and search outcomes. *British Journal of Educational Technology, 36*(3), 465-475.

Métrailler, Y. A., Reijnen, E., Kneser, C., & Opwis, K. (2008). Scientific problem solving in a virtual laboratory: A comparison between individuals and pairs. *Swiss Journal of Psychology, 67*(2), 71-83.

Okada, T., & Simon, H. A. (1997). Collaborative discovery in a scientific domain. *Cognitive Science, 21*(2), 109-146.

Paulus, P. (2000). Groups, teams, and creativity: The creative potential of idea-generating groups. *Applied Psychology, 49*(2), 237-262.

Roschelle, J., & Teasley, S. (1995). The construction of shared knowledge in collaborative problem solving. In C. O'Malley (Ed.), *Computer-Supported Collaborative Learning* (pp. 69-197). Berlin, Germany: Springer.

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

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

Schwartz, D. L. (1995). The emergence of abstract representations in dyad problem solving. *The Journal of the Learning Sciences, 4*(3), 321-354.

Strijbos, J. W., Martens, R. L., Prins, F. J., & Jochems, W. M. (2006). Content analysis: What are they talking about? *Computers & Education*, *46*(1), 29-48.

Teasley, S. (1995). The role of talk in children's peer collaborations. *Developmental Psychology, 31*(2), 207-220.

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.

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.

Wiley, J. & Jolly, C. (2003) When two heads are better than one expert. In R. Alterman D. Hirsh (Eds.), *Proceedings of the Twenty-Fifth Annual Conference of the Cognitive Science Society* (pp 1242-1246). Boston, Massachusetts: Lawrence Erlbaum Associates Inc.

Zimmerman, C. (2000). The development of scientific reasoning skills. *Developmental Review, 20*(1), 99-149.

Acknowledgments