This is a post-peer-review, pre-copyedit version of an article published in International Journal of Computer-Supported Collaborative Learning (ijCSCL). The final authenticated version is available online at:
http://dx.doi.org/10.1007/s11412-017-9260-z

Please cite this article as:
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*(4). 281-305. doi: 10.1007/s11412-017-9260-z

How to combine collaboration scripts and heuristic worked examples to foster mathematical argumentation – when working memory matters

Matthias Schwaighofer[1], Freydis Vogel[2], Ingo Kollar[3], Stefan Ufer[4], Anselm Strohmaier[2], Ilka Terwedow[1], Sarah Ottinger[4], Kristina Reiss[2], & Frank Fischer[1]

[1]LMU Munich,
Chair of Education and Educational Psychology,
Leopoldstr. 13, 80802 Munich, Germany

[2]Technical University of Munich,
School of Education,
Arcisstr. 21, 80333 Munich, Germany

[3]University of Augsburg,
Universitätsstr. 10, 86159 Augsburg, Germany

[4]LMU Munich,
Chair of Mathematics Education,
Theresienstr. 39, 80333 Munich, Germany

Corresponding author:
Ingo Kollar: ingo.kollar@phil.uni-augsburg.de

Abstract Mathematical argumentation skills (MAS) are considered an important outcome of mathematics learning, particularly in secondary and tertiary education. As MAS are complex, an effective way of supporting their acquisition may require combining different scaffolds. However, how to combine different scaffolds is a delicate issue, as providing learners with more than one scaffold may be overwhelming, especially when these scaffolds are presented at the same time in the learning process and when learners' individual learning prerequisites are suboptimal. The present study therefore investigated the effects of the presentation sequence of introducing two scaffolds (collaboration script first vs. heuristic worked examples first) and the fading of the primarily presented scaffold (fading vs. no fading) on the acquisition of dialogic and dialectic MAS of participants of a preparatory mathematics course at university. In addition, we explored how prior knowledge and working memory capacity moderated the effects. Overall, 108 university freshmen worked in dyads on mathematical proof tasks in four treatment sessions. Results showed no effects of the presentation sequence of the collaboration script and heuristic worked examples on dialogic and dialectic MAS. Yet, fading of the initially introduced scaffold had a positive main effect on dialogic MAS. Concerning dialectic MAS, fading the collaboration script when it was presented first was most effective for learners with low working memory capacity. The collaboration script might be appropriate to initially support dialectic MAS, but might be overwhelming for learners with lower working memory capacity when combined with heuristic worked examples later on.

Keywords Mathematical argumentation skills * Collaboration scripts * Heuristic worked examples * Working memory capacity

Mathematical argumentations skills as an educational goal

Mathematical proof can be seen as a specific type of argumentation. Because proofs are central within mathematics as a science, mathematical argumentation is pivotal for mathematical activity (Hanna, 2000). In school curricula, meaningful practices such as constructing mathematical arguments and critiquing the reasoning of others are considered to be important goals of mathematics education (CCSSI, 2017). Constructing arguments is also an extensively studied and important goal in secondary and tertiary education (Schwaighofer, Fischer, & Bühner, 2015).

Mathematical argumentation skills (MAS) include not only domain-specific, i.e. genuine mathematical skills (e.g., Yackel & Cobb, 1998), but also knowledge and skills regarding social-discursive aspects of argumentation (Kollar et al., 2014). Social-discursive MAS are necessary, for instance, when different steps of a mathematical proof process are discussed, when an individual proof idea is explained, or when two learners try to jointly find solution steps for a proof. That way, social-discursive MAS serve two purposes: On the one hand, social-discursive argumentative activities may lead to cognitive elaboration of mathematical concepts that are required to solve proof tasks and thus optimally help learners acquire domain-specific skills ("arguing to learn"; Andriessen, Baker & Suthers, 2003). For instance, one learner may have to deeply elaborate on what the learning partner has formulated in order to be able to understand and criticize the other position. On the other hand, engaging in social-discursive argumentative activities may also help students acquire social-discursive MAS, as the repeated engagement in such activities should yield a practice effect ("learning to argue"; Andriessen et al., 2003). In the study at hand we focus on that "learning to argue" objective: We study to what extent students' social-discursive MAS can be enhanced by different scaffolds.

Within social-discursive argumentation, two different types of activities can be distinguished, namely dialogic activities and dialectic activities (Wegerif, 2008; Schwarz & Shahar, 2017). *Dialogic activities* are characterized by a joint conversation on the same arguments based on exchanging differences in a participatory way without overcoming these differences (Wegerif, 2008). I.e. two learners, while trying to find a solution for a task, work together to improve the joint argument by finding better reasons, explanations, further clarification, etc. In contrast, *dialectic activities* comprise counterarguments (e.g., challenges to arguments) and the integration of different arguments to arrive at a joint solution by explicating conflicting arguments, and by linking and weighing these arguments (e.g., by accepting parts of each learners' arguments; Schwarz, 2009).

Both an engagement in dialogic as well as in dialectic activities is assumed to be beneficial for learning (see Teasley, 1997). There however is some evidence that dialectic activities are even more important than dialogic ones in that regard, as was shown in studies by Asterhan and Schwarz (2007, 2009). More specifically, Vogel et al. (2016) demonstrated that the use of dialectic, but not dialogic activities improved learners' disposition to use argumentation skills (e.g., by providing counterarguments).

Difficulties in mathematical argumentation

Even though the importance of engaging in both dialogic and dialectic argumentation as an important motor for learning has often been acknowledged, learners often experience difficulties during mathematical argumentation. For example, they are often not able to produce logical chains of more than one argument (Heinze, Reiss, & Rudolph, 2005). Also, concerning dialectic activities, they often fail to engage in a comprehensive argumentative discourse cycle with counterarguments and integration of argumentation (Leitão, 2000), or they leave out important parts in their argumentation, such as justifications for their claims or responses to counterarguments (Jiménez-Aleixandre, Rodrígues, & Duschl, 2000; Sadler, 2004).

The transition to a university mathematics program poses specific challenges in this respect, since it includes the transition from the application-oriented school subject "mathematics" towards the scientific discipline "mathematics" (Rach & Heinze, 2016; Vollstedt, Heinze, Gojdka, & Rach, 2014), with its own values and norms regarding mathematical proof and argumentation (Dawkins & Weber, 2016). During their university studies, students are requested not only to find consistent lines of deductive arguments from a framing theory to validate specific hypotheses, but also to communicate these arguments according to mathematical standards (Vogel et al., 2016a). This transition is challenging (e.g., Hodds, Alcock, & Inglis, 2014). Therefore, supporting prospective university mathematics students to facilitate a successful transition to their study programs seems to be warranted. Preparatory courses and transition-to-proof courses are common to support students MAS in these settings (e.g., Bausch et al., 2014; Selden, Benkhalti, & Selden, 2014). However, the effectiveness of integrating promising scaffolds to foster MAS in preparatory courses has rarely been investigated systematically.

Fostering mathematical argumentation skills

Past research (Kollar et al., 2014) has shown that two promising candidates for fostering MAS are collaboration scripts and heuristic worked examples. Both scaffolds are subsequently described.

Collaboration scripts

Collaboration scripts support learners with respect to rather content-independent, social-discursive processes while being engaged in a collaborative task. For instance, these scripts may prompt learners to provide arguments for their positions and share them with their learning partner(s) (Kollar et al., 2014). That way, collaboration scripts specify and sequence learning activities and distribute them among the learners of a small group.

Optimally, the design of collaboration scripts is based on empirical research that demonstrated what collaborative activities go along with in-depth knowledge acquisition (e.g., explaining ideas and concepts, argumentation, resolving conceptual discrepancies). Since learners often do not spontaneously use the most beneficial strategies in collaborative learning (e.g., King, 2007), external support by means of collaboration scripts seems to be warranted.

Several studies in contexts other than mathematics have shown that learning with collaboration scripts may foster the acquisition of rather general collaboration skills, such as argumentation skills (e.g., Rummel, Mullins, & Spada, 2012; Schellens, Van Keer, De Wever, & Valcke, 2007; Weinberger, Stegmann, & Fischer, 2010). Collaboration scripts are also a promising scaffold to support the social-discursive aspects of MAS (Kollar et al., 2014, Vogel et al., 2016), as they may prompt learners to provide arguments, counterarguments and to integrate different arguments of learning partners. Thereby, they may especially facilitate dialectic activities. Dialogic activities may however also be induced when learners try to expand the arguments provided by a learning partner. However, the possibility of using collaboration scripts to foster dialogic and dialectic activities in the context of mathematical argumentation has not yet been systematically investigated. Due to their in principle content-independent nature, such collaboration scripts may however become even more effective when they are coupled with content-specific scaffolds such as heuristic worked examples (Reiss & Renkl, 2002).

Heuristic worked examples

Worked examples usually consist of a problem formulation, steps to solve the problem, and a final solution (e.g., Renkl, 2014). *Heuristic* worked examples do not only include solutions for particular problems in an exemplifying domain (e.g. elementary number theory), but also principles of a specific learning domain (e.g., how to formulate and prove a conjecture), and strategies to solve similar problems (Renkl, Hilbert, & Schworm, 2009). For this purpose, they may describe two fictitious learners trying to solve a mathematical problem with different approaches, thereby externalizing their strategies. The approaches of the fictitious learners can make strategic thinking visible.

In a collaborative learning process, heuristic worked examples may elicit both dialogic and dialectic activities. Heuristic worked examples rarely contain explicit debates about strategies. Rather, one or two fictitious learners argue along a consistent line of thought, modeling strategies that can be applied in the argumentation process. These strategies may support real learners to formulate arguments and to extend ideas of their learning partners. Thus, the heuristic worked examples include information that can be used for dialogic activities. In addition to dialogic MAS, they may also address dialectic activities in collaborative scenarios, e.g. when contrasting heuristic worked examples are distributed among the learners of a small group in order to increase the likelihood of socio-cognitive conflicts to emerge (Clark, D'Angelo, Meneske, 2009). Overcoming different viewpoints conveyed by contrasting heuristic worked examples may involve exchanging of arguments and counterarguments and attempts to come to an integration of the different viewpoints. Studies in the mathematical context (especially in geometry) have shown positive effects of learning with heuristic worked examples on mathematical argumentation and proof (e.g., Reiss, Heinze, Renkl, & Große, 2008) and social-discursive MAS (Kollar et al., 2014).

How to combine collaboration scripts and heuristic worked examples: Presentation sequence, fading, and the role of individual learner characteristics

A straightforward idea to supplement collaboration scripts with heuristic worked examples would be the simple combination of the two scaffolds. This combination might lead to synergistic scaffolding effects when both scaffolds mutually increase their effectiveness concerning a joint outcome (Tabak, 2004). However, prior research by Kollar et al. (2014) did yield evidence for a synergistic scaffolding effect, as learners who worked both with collaboration scripts and heuristic worked examples did not outperform students who had only received one of the two scaffolds.

It thus seems that certain conditions need to be met to reach synergistic scaffolding effects through a combination of collaboration scripts and heuristic worked examples. One idea might be to not present the two scaffolds simultaneously (as in the study by Kollar et al., 2014), but rather in a step-wise fashion. When doing so, three questions pop up: First, what scaffold should be presented first (*presentation sequence*)? Second, should the scaffold that is presented first still be available once the second scaffold is introduced or should it be faded out (*fading of scaffolds*)? And third, since presenting two scaffolds in combination – be it simultaneously or sequentially – is demanding for learners: What is the role of individual learner characteristics for the effectiveness of combining collaboration scripts with heuristic worked examples (role of individual learner characteristics)? In the following, these three questions are considered in more detail.

Presentation sequence of scaffolds

The temporal sequence by which scaffolds are presented may substantially influence learning outcomes (Renkl & Atkinson, 2007). Concerning collaboration scripts and heuristic worked examples, it though is not clear what scaffold should be presented first, and it seems possible to find arguments for both possible options (either presenting collaboration scripts first or presenting heuristic worked examples first). On the one hand, one might assume that it is more important to first receive content-specific support by heuristic worked examples in order to first help students construct content knowledge which in turn is a necessary basis for further argumentation processes. On the other hand, it may also be easier for learners to first learn about the general, cross-domain strategy of dialectic argumentation with a content-independent collaboration script before they apply that strategy in learning about the domain. The results of a study by Clarke, Ayres, and Sweller (2005) seem to be in accordance with this latter line of reasoning. The authors investigated whether spreadsheets (as a content-independent scaffold) to assist mathematics learning should be introduced before or concurrent with content-specific mathematical guidance. Introducing the content-independent scaffold first was superior – at least for learners with low prior knowledge regarding spreadsheets. Whether these results can be transferred to the combination of collaboration scripts and heuristic worked examples to foster students' dialogic and dialectic MAS is an open question.

Fading of scaffolds

Another question that needs to be answered when collaboration scripts and heuristic worked examples are presented in a step-wise fashion is whether the scaffold that is presented first should remain to be present once the second one comes into play. Based on prior research, both the fading-out of the first scaffold and the simultaneous availability of two scaffolds could be beneficial. On the one hand, learners may best be supported to integrate information provided by the two scaffolds, which would yield the hypothesis that the initially presented scaffold should still be available after the second one is introduced. For example, if heuristic worked examples are still available when introducing a collaboration script, learners may easily refer to the strategies conveyed by the previous heuristic worked examples for their argumentation about new examples. Also, taking away the previously presented scaffold may come too early for learners because they have to self-regulate their performance immediately with little previous practice (Wecker & Fischer, 2011). Especially learners with less favorable learning prerequisites (e.g., low prior knowledge concerning social-discursive MAS) may lack the skill of self-regulating dialogic and dialectic activities during the learning process. On the other hand, fading has been considered an important part of scaffolding (Pea, 2004) that affords learners to increasingly take control of their own learning activities. Further, knowledge about regulating the execution of skills can be acquired by repeatedly applying them in multiple contexts (e.g., Spiro, Coulson, Feltovich, & Anderson, 1988). Fading that scaffold that was presented first once the second one is introduced may thus enable learners to practice skills and thereby strengthen their dialogic and dialectic MAS.

The role of individual learner characteristics

The effectiveness of different scaffolds, especially when combined in one learning environment, may depend on specific individual learner characteristics. We specifically focus on two variables: prior knowledge and working memory capacity.

Prior knowledge. Prior knowledge has repeatedly been shown to be one of the most important factors influencing learning (Kalyuga, 2013). For example, it is predictive for learning in statistics (Leppink, Broers, Imbos, van der Vleuten, & Berger 2012), and performance in physics and mathematics (e.g., Hailikari, Nevgi, & Komulainen, 2008; Hudson & Rottmann, 1981).

Furthermore, prior knowledge is considered to be a potential moderator of the effectiveness of various kinds of scaffolds. However, whether high or low knowledgeable learners benefit most from instructional support seems to be unclear. Research using more general measures of prior knowledge (e.g., grade point average) has found that highly knowledgeable learners may benefit most from instruction. This finding has been termed *Matthew effect* (e.g., Stanovich, 1988). One explanation could be that learners with high prior knowledge are more likely to distinguish relevant from irrelevant information in texts (Alexander & Jetton, 2003) and are better able to integrate new information in existing schemata (Kollar, et al., 2014).

In contrast to research that hints towards a Matthew effect of scaffolding, some studies that usually use more specific instruments to assess prior knowledge (such as point scores in a content knowledge pretest) suggest that the effectiveness of scaffolds may decrease with increasing prior knowledge. This finding has been termed expertise reversal effect (e.g., Kalyuga, Rikers, & Paas, 2012). The explanation for the expertise reversal effect predominantly comes from cognitive load theory (e.g., Sweller, 2011). Accordingly, learners with high levels of prior knowledge have schemas which can be represented as single elements in working memory. Thus, these learners are likely to experience a low intrinsic cognitive load in working memory (i.e., working memory load due to the interacting elements in the learning material; Sweller, 2011). Contrary, for beginners, problem solving may induce a high cognitive load that is irrelevant for schema construction. This kind of cognitive load is called extraneous cognitive load. Worked examples may reduce extraneous cognitive load. Thereby, enough working memory capacity can be devoted to schema construction (Renkl, 2014). In contrast, supporting expert learners with information they already have in long-term memory may be redundant and cause additional extraneous cognitive load (e.g., Kalyuga, 2007). Applied to heuristic worked examples, the heuristics provided by the examples may interfere with learners' existing strategies (Reiss et al., 2008) so that students with low prior knowledge may not be able to use the support to engage in processes associated with schema construction (germane load), but instead be overwhelmed by having to coordinate the different kinds of support they are confronted with.

Working memory capacity. Working memory serves the function of temporarily storing and manipulating information (Baddeley, Allen, & Hitch, 2011). Several cognitive achievements depend on working memory such as problem solving performance (Bühner, Kröner, & Ziegler, 2008), math achievement (e.g., Peng, Namkung, Barnes, & Sun, 2016), and reading comprehension (Daneman & Merikle, 1996). Furthermore, working memory capacity is moderately correlated with fluid intelligence (e.g., Redick, Unsworth, Kelly, & Engle, 2012). Although working memory capacity presumably plays an important role for learning within cognitive load theory (e.g., Sweller, 2011), few studies investigating the effectiveness of worked examples have used objective and reliable measures of working memory capacity (for exceptions see de Jong, 2010; Schwaighofer et al., 2016) and instead relied on a subjective rating scale of cognitive load. However, concerns regarding the validity of the subjective rating scale exist (de Jong, 2010; Schwaighofer et al., 2016). For instance, Schwaighofer et al. (2016) found that the subjective rating of cognitive load did not correlate with working memory capacity measured with three reliable and valid tasks. Examples for such tasks are complex span tasks. In an operation span task, for example, participants receive a set of simple math tasks composed of three digits and two operations (e.g., "(2x2) + 5 = ?"; see Redick et al., 2012, p. 848) together with a suggestion for a solution and are asked

to hit “TRUE” or “FALSE” on a computer keyboard. After each task, the participant receives a letter she is asked to remember until the end of the trial. Working memory span is then operationalized via the number of the correctly remembered letters in serial order.
Concerning the combination of collaboration scripts with heuristic worked examples, learners with low working memory capacity might be overwhelmed when the two scaffolds are presented at the same time. Learners with high working memory capacity, in contrast, may be better able to integrate information from scaffolds that are presented simultaneously. Therefore, these learners might benefit from the simultaneous presentation of collaboration scripts and heuristic worked examples.
When the second scaffold is introduced, the availability of the first scaffold may induce a high extraneous cognitive load in working memory when information coming from the first scaffold is redundant to some extent. Again, especially learners with low working memory capacity may struggle with the high demand on working memory and not have enough working memory capacity to deal with information from the second scaffold. In addition, these learners might not be able to integrate relevant information from the two scaffolds. Therefore, fading of the initially presented scaffold could be effective particularly for learners with low working memory capacity.

Research Questions

Against the background of these considerations, this study investigates the effects of different step-wise combinations of collaboration scripts and heuristic worked examples on dialectic and dialogic MAS. The scaffolds were used for mathematical proof tasks that students were asked to solve in dyads in the context of a two-week preparatory course for mathematics freshmen at a German university. We asked the following research questions:

RQ1: What is the effect of the presentation sequence of the two scaffolds (collaboration script first vs. heuristic worked examples first), the fading of the primarily presented scaffold (fading vs. no fading) and their combination on learners’ acquisition of dialogic MAS (RQ1a) and dialectic MAS (RQ1b) during collaborative learning with mathematical proof tasks? Concerning the effect of the presentation sequence of the two scaffolds and the fading of the scaffold introduced first we described contradicting consequences that could be expected to happen. Learners might either benefit from learning with the rather content-independent scaffold or the content-specific scaffold first. Also fading of the primarily presented scaffold could either enhance learners' development of the faded components or fading could overwhelm learners. Therefore, we hypothesize effects of both the presentation sequence and the fading, but cannot determine the direction of the effects a-priori.

RQ2: To what extent is the effect of the presentation sequence of the two scaffolds (collaboration script first vs. heuristic worked examples first) and the fading of the initially introduced scaffold (fading vs. no fading) on learners' acquisition of *dialogic* and *dialectic* MAS moderated by learners' prior knowledge (RQ2a) and working memory capacity (RQ2b)?

For the moderation of the effects of the presentation sequence and the fading of the scaffolds on dialogic and dialectic MAS by learners' prior knowledge we argue that the Matthew effect would speak for learners with higher prior knowledge would benefit from no fading of either presentation sequence. In contrast, the expertise reversal effect would rather speak for learners with higher prior knowledge would benefit from fading of either presentation sequence. Therefore, we expect a moderation effect without a specific direction. For the moderation of the effects of the presentation sequence and the fading by learners' working memory capacity, we hypothesize that learners with higher working memory capacity might be less affected by the presentation sequence and fading while learners with lower working memory capacity might be affected by the presentation sequence and might benefit from fading either scaffold.

Method

Setting and sample

The study was conducted within a two-week preparatory course for prospective mathematics university students. The course was offered before the beginning of their first semester to support them in the transition from secondary school mathematics to mathematics at the university. It contained eleven lectures and eleven tutor exercises on elementary number theory and other basic mathematical topics (e.g., basic propositional and predicate logic, proof techniques, induction and recursion). Participation in the preparatory course was voluntary. Overall, $N = 108$ learners ($M_{\text{age}} = 18.99$, $SD_{\text{age}} = 1.89$; 45 female learners) were included in the analyses as they completed the course and took part in all treatment and test sessions.

Learning material

During the four treatment sessions, learners were seated in dyads collaborating on one mathematical proof task per session. The tasks were presented on a shared worksheet which also contained a coarse structure of the task process. Learners wrote down their ideas using Livescribe Smartpens with integrated microphones. The Smartpens recorded the dyad's talk in a digital video file, as well as their writing on the shared worksheet. Afterwards, each learner was asked to develop an individual solution based on the results of the collaboration. All dyads were provided with lecture notes that contained content from all lectures taught in the preparatory course. The collaboration script and the heuristic worked examples were presented depending on the experimental condition. Heuristic worked examples were provided in printed form. The collaboration script was implemented in the shared worksheet (see the description in the section about the operationalization of the collaboration script and heuristic worked examples below).

Design

The learners were randomly assigned to one of four experimental conditions of a 2 x 2 factorial design with the independent variables *presentation sequence of the scaffolds* (collaboration script first vs. heuristic worked examples first) and *fading of the initially introduced scaffold* (fading vs. no fading; see Table 1).

Table 1 Experimental conditions

		Presentation sequence of the scaffolds	
		Collaboration script first	Heuristic worked examples first
Fading of the initially introduced scaffold	Fading	$n = 31$	$n = 26$
	No fading	$n = 24$	$n = 27$

In the *collaboration script first* conditions, learners received the collaboration script in the first and second treatment session and the heuristic worked examples in the third and fourth treatment session. In contrast, learners in the *heuristic worked examples first* conditions received heuristic worked examples in the first and second treatment session and the collaboration script in the third and fourth treatment session.

Whether the initially introduced scaffold was still available in the third and fourth treatment session (i.e., when the second scaffold was presented) was determined by the second independent variable, *fading of the initially introduced scaffold*: The *fading* conditions did not receive the initially introduced scaffold in the third and fourth treatment session, while the *no fading* conditions received the initially introduced scaffold in the third and fourth treatment session in addition to the scaffold that was presented second.

Operationalization of the collaboration script and heuristic worked examples

Collaboration script. Before learners started to work on the mathematical proof tasks, the experimenter informed them about the structure of argumentation prompted by the collaboration script. This was to make sure that all learners understood how to use the collaboration script. Figure 1 shows the prompts of the collaboration script in the shared worksheets, which were intended to structure the discussion between the learning partners according to the three phases of argumentation proposed by Leitão (2000). These phases were adapted for the present study. Phase 1: presentation of arguments for a step in solving the problem (a step presented by the learner him- or herself when no heuristic worked example was simultaneously presented; or a step that was prestented by the fictitious learner the was described in the heuristic worked example when a heuristic worked example was simultaneously presented (see the description of the heuristic worked examples in the section below). Phase 2: critical evaluation of the arguments for the step in solving the problem (i.e., answering with a counterargument). Phase 3: building a synthesis for the arguments raised before. For example, in the condition with collaboration script and heuristic worked examples, the prompt related to the phase of building a synthesis was "Evaluate the pros and cons of the approaches by the fictitious learners and agree upon the best approach from your point of view". In the condition without heuristic worked examples the prompts referred to the real learning partner, see Figure 1). Especially in the last two phases, the prompts of the collaboration script focused on dialectic activities. Because integrating ideas (step 3) may also involve extending ideas of learning partners, the collaboration script prompts, however, also targeted dialogic activities to some extent. When no collaboration script was present, students were prompted to alternately work individually on the task and exchange their ideas collaboratively. Yet, there was no further structure given for the collaborative exchange of ideas.

<Proof task>

At first you should find an idea to this problem. Please start by working **ALONE** on the concept paper. Once you have an idea to discuss with your partner, please return to this shared worksheet (page X).

Now swap your ideas! Please shortly write down your most important considerations here!

1. **Presentation of different approaches**:
   Present your suggestions to the first processing step to each other and shortly write down your most important considerations!

Approach by learner A	**Approach by learner B**

<Proof task>

2. **Critical evaluation of the different approaches:**
   Evaluate each of the approach with your learning partner critically and justify your evaluation. Write down your most important arguments.

Evaluation of the approach by learner A	**Evaluation of the approach by learner B**

3. **Weighing and agreeing on the further procedure:**
   Weigh the pros and cons of the both approaches and agree on the best approach from your point of view. Note your most important considerations!

When you have finished the discussion, please take your concept paper and further develop a solution idea ALONE! Once you have an idea to discuss with your partner, please return to this shared worksheet (page X)!

Figure 1 Prompts of the collaboration script at the first page of the shared worksheets when students did not have a heuristic worked example.

Heuristic worked examples. Each heuristic worked example delineated how a fictitious learner tried to prove a conjecture for the given problem from elementary number theory according to the six phases adapted from Boero's (1999) process model of mathematical proof. One example for a problem from elementary number theory is: "Choose some square numbers and take differences of two square numbers. What do you notice? Formulate a conjecture and prove it!" To ensure that the learners understood how to work with heuristic worked examples, at the beginning of each treatment session the experimenter informed about the structure of a heuristic worked example, to track the solution processes in the examples, and to alternately work individually and collaboratively on the task. Figure 2 shows the third of six solution steps of a heuristic worked example related to the problem from elementary number theory described above.

Collecting ideas

Anna starts thinking about how to prove her conjecture and remembers helpful contents of the lecture.

Think about which contents of the lecture could be helpful for Anna`s idea.

First, she tries to find an appropriate illustration for the difference in two square numbers:

$$m^2 - n^2$$

Anna: This illustration does not take into account that the distance between two numbers to be squared plays a role. Maybe this illustration is better:

$$(n+i)^2 - n^2$$

Look at the first examples from Anna. What would be "i" in these examples?

Anna: Maybe I can simplify this:

$$(n+i)^2 - n^2 = n^2 + 2ni + i^2 - n^2 = 2ni + i^2$$

Anna: If I now know more about i, I can also say more about the partial terms. Oh yes, i is the distance between the two numbers which are squared and then subtracted. Therefore, the distance lies between n and n + i. And I think it`s crucial whether i is even or odd. Thus, I can write down formally my conjecture in this way:

Conjecture

$(n+i) - n^2 = 2ni + i^2$, whereby $2ni + i^2$

- Either is an odd number when *i is odd* or
- A number divisible by ← when *i is even*

Which meaning does this step ("Collecting ideas") have for Anna? According to which system did she proceed?

When you are finished, return to the shared worksheet on the pages 4-5.

Figure 2 Third of six solution steps for the task "Choose some square numbers and take differences for of two square numbers. What do you notice? Formulate a conjecture and prove it!"

The two learning partners in each dyad received heuristic worked examples on the same problem with different fictitious learners. The heuristic strategies of the fictitious learners in the worked examples differed to increase the need for discussion between the learning partners. Each solution step of a heuristic worked example contained prompts to reflect about the solution steps. For instance, learners were prompted to judge in which way the approach to the problem chosen by the fictitious learner might be beneficial to solve this and other problems, and to compare this strategy with that of the fictitious learner in the partner's worked example. After the first, the third, and the sixth solution step, participants were prompted to switch to the shared worksheet. These worksheets contained prompts to exchange ideas from the fictitious learners in the heuristic worked examples between the learning partners and to record the most important considerations on the sheet (either supported by the collaboration script or not). In addition, the worksheets contained prompts to return to the heuristic worked examples and work on the next solution steps after finishing the discussion. When no heuristic worked example was present, the learners were alternately asked to work individually on their idea for a step to come to a solution of the problem and collaboratively exchange their ideas

Procedure

The study contained two pretests, four treatment sessions, and a posttest during six consecutive weekdays. The posttest took place one day after the last treatment session. For each of the four treatment sessions, learners were randomly assigned to a new learning partner. Dyads were always homogeneous with respect to academic ability, which was realized by a median split of the final high school grade which was measured during pretest and by creating groups with either two high or two low ability learners. We decided to form homogeneous dyads to reduce further noise in the data, because dyads with comparable learning prerequisites might process learning materials differently than dyads with strongly different learning prerequisites (Webb, Nemer, & Zuniga, 2002). At the outset of the first treatment session, the experimenter explained the purpose and the procedure of the sessions and explained how to use the Smartpens. During each treatment session, the learners learned in dyads on a new mathematical proof task and received support by different scaffolds depending on their experimental condition.

Dependent variables

Dialogic and dialectic MAS. During pre- and posttest, participants worked on a test to measure their *dialogic and dialectic MAS*. The test asked them to describe phases and activities that appear in a prototypical talk between two individuals who have different positions regarding the question on how to best support learning motivation (pretest) and to what extent talent or practice accounts for a person's development of mathematical expertise. These questions were chosen in a way that should trigger the participants to describe the dialog, debate or discussion they would expect to appear. The students usually described a sequence of phases titled as "beginning of the talk", "stating arguments", "evidence", "discussion", "counterarguments", "critic", "conclusion", "compromise", "end of the talk", etc. Answers of participants were analyzed with respect to dialogic and dialectic activities. Dialogic activities included (1) agreements and (2) extensions of the other arguments, while dialectic activities comprised (1) critique, (2) counterarguments and (3) integrations of arguments and counterarguments. Concerning dialogic MAS, learners received one point each when they mentioned agreements or extensions of arguments. With respect to dialectic MAS, one point was awarded each when learners mentioned critique, counterarguments or integrations of arguments and counterarguments. Table 2 shows some examples of students' answers that were either rated with high or low values for dialogic and dialectic MAS.
For both kinds of MAS, we only rated if the single items appeared at all with one point for each and summed up the entries of dialogic and dialectic activities. This resulted in a range of 0 to 2 points for dialogic MAS and a range of 0 to 3 points for dialogic MAS. Two student assistants were trained to rate learners' answers for mentioning dialogic and dialectic activities with data that was not included in the study at hand. The rater training took four rounds of rating, discussing and adapting the coding scheme by including new examples etc. Each round took about one week to complete. After finishing the fourth round and consolidating the coding scheme, the two student assistants rated a random sample from the actual data of 26 pre-test answers and 25 post-test answers separately to calculate inter-rater reliability. Sufficient values of inter-rater reliability were reached for the about 23% of the ratings of the students' answers (Cohen's κ for dialogic MAS: $M = .71$, range = .68-.75; Cohen's κ for dialectic MAS: $M = .74$, range: .67-.83). Then, the data was evenly distributed between the two raters and each data set was rated by one of the two raters.

Table 2 Students' answers and coding of dialogic and dialectic activities

Student's answer	Dialogic activities	Dialectic activities
"The first interlocutor explains his arguments. The second interlocutor listens carefully and repeats the arguments of the first interlocutor in his own words to make sure he understood. He also adds his own ideas."	*high* (agreement, extension)	low (-)
"The first interlocutor poses his argument and an example. He states a hypothesis and tries to prove it with reasons. The second interlocutor poses a counterargument and an example. He shows his disagreement with an own hypothesis and proves it with reasons."	*low* (-)	*medium* (counter-argument)
"The first interlocutor collects the most important arguments. The second interlocutor extends the collection."	*medium* (extension)	*low* (-)
"The first interlocutor poses his hypothesis and arguments. The second interlocutor tries to find weaknesses in the argumentation of the first interlocutor and criticizes it. Then he poses counterarguments (…) in the end both interlocutors balance the different arguments and try to find a joint solution."	*low* (-)	*high* (critique, counter-argument, integration)

Control and moderator variables

Prior knowledge (dialogic and dialectic MAS). As described, we measured dialogic and dialectic MAS also during pretest (see section about the dependent varibale). The pretest scores were used as covariates in subsequent analyses.

Working memory capacity. Working memory capacity was measured in separate sessions during the preparatory course. Groups of students were invited into a separate room to complete the automated operation span task on a laptop computer (Unsworth, Heitz, Schrock, & Engle, 2005). In this task, participants have to alternately solve simple mathematical equations and memorize letters which have to be recalled at the end of a sequence. The sum of letters recalled in all sequences divided by all trials serves as an estimate of the participant's working memory capacity (Unsworth et al., 2005). The internal consistency (Cronbach's alpha) of the automated operation span was calculated by using the method of Kane et al. (2004) and yielded a value of $\alpha = .63$.

Fluid intelligence. We assessed *fluid intelligence* at the second pretest using the sum score of the short version of the Culture Fair Intelligence scale (CFT 20-R; Weiß, 2006). The short version comprised four subtests with 56 items in total. The reliability of the test was $\alpha = .74$.

Statistical analyses

The effects of the presentation sequence and fading of the two scaffolds on dialogic and dialectic MAS were analyzed using analyses of covariance controlling for prior dialogic or dialectic MAS, respectively.

Moderation analyses were conducted for prior knowledge and working memory capacity moderating the effects of the presentation sequence and fading of the scaffolds on the post test values of dialogic and dialectic MAS. These analyses were conducted with the SPSS macro PROCESS (Hayes, 2013). As proposed by Hayes (2012), heteroscedasticity-consistent standard errors were estimated. The influence of prior knowledge on the moderator and the dependent variable was controlled for when necessary. Applying the Johnson-Neyman technique (see Hayes, 2013) allowed us to quantify the effect of the independent variables on the dependent variables for different values of the respective moderator (prior knowledge or working memory capacity).

An alpha-level of 5% was used for all analyses.

Results

Preliminary analyses

Correlations among moderator, control and dependent variables. Dialogic and dialectic MAS were not correlated significantly, neither at pre-, nor at posttest. Dialogic MAS at pretest correlated with dialogic MAS at posttest, and dialectic MAS at pretest correlated with dialectic MAS at posttest (see Table 3). The fluid intelligence did not correlate with any of the other variables. The working memory only correlated significantly negatively with the dialectic MAS at posttest. In the subsequent analyses only for the prior knowledge significant effects on the dependent variables were found but not for fluid intelligence or working memory capacity. Therefore, in all subsequent analyses, prior knowledge concerning the respective dependent variable (i.e., dialogic or dialectic MAS) was included as covariate (if not already included as moderator). Neither fluid intelligence nor working memory capacity were included as covariate in the subsequent analyses.

Table 3 Correlations among moderator, control and dependent variables.

		Dialogic MS at pretest	Dialectic MAS at posttest	Working memory capacity	Fluid intelligence	Dialogic MAS at posttest	Dialectic MAS at posttest
Dialogic MAS at pretest	*r*	1					
	N	108					
Dialectic MAS at posttest	*r*	.18	1				
	N	108	108				
Working memory capacity	*r*	.06	-.02	1			
	N	97	97	97			
Fluid intelligence	*r*	-.136	.16	.16	1		
	N	106	106	96	106		
Dialogic MAS at posttest	*r*	.31**	-.03	.08	.02	1	
	N	108	108	97	106	108	
Dialectic MAS at posttest	*r*	.04	.29**	-.21*	.07	.01	1
	N	108	108	97	106	108	108

**$p < .01$, *$p < .05$ (two-tailed).

RQ1a: Effects of the presentation sequence and fading of scaffolds on the acquisition of dialogic MAS

Descriptively, the condition that was first presented with the collaboration script that was faded afterwards performed best in dialogic MAS, while the condition that was first presented with the heuristic worked examples that were not faded afterwards performed worst in dialogic MAS. Table 4 shows means and standard deviations for dialogic MAS for each experimental condition at posttest.

Table 4 Means and standard deviations (in parentheses) for the sum of dialogic activities mentioned by the learners in the individual posttest on dialogic MAS.

		Presentation sequence of the two scaffolds	
		Collaboration script first	Heuristic worked examples first
Fading of the initially introduced scaffold	Fading	0.35 (0.49)	0.23 (0.51)
	No fading	0.21 (0.42)	0.04 (0.19)

The results of the ANCOVA with the pretest dialogic MAS as covariate showed that overall there was no significant main effect of the presentation sequence of the two scaffolds on the acquisition of dialogic MAS ($F(1,103) = 1.81$, $p = .18$, partial $\eta^2 = .02$). In contrast, fading of the initially introduced scaffold had a significant positive effect on the acquisition of dialogic MAS, $F(1,103) = 6.63$, $p = .01$, partial $\eta^2 = .06$. No interaction effect between presentation sequence and fading of the two scaffolds occurred for the dialogic MAS ($F(1,103) = 0.19$, $p = .67$, partial $\eta^2 < .01$).

RQ1b: Effects of the presentation sequence and fading of the scaffolds on the acquisition of dialectic MAS

In line with the results for the dialogic MAS, descriptively the condition that was first presented with the collaboration script that was faded afterwards performed best in dialectic MAS. The condition that was first presented with the heuristic worked examples that were not faded afterwards performed worst in dialogic MAS. Table 5 shows means and standard deviations for dialectic MAS for each experimental condition at posttest.

Table 5 Means and standard deviations (in parentheses) for the sum of dialectic activities mentioned by the learners in the individual posttest on dialectic MAS.

		Presentation sequence of the two scaffolds	
		Collaboration script first	Heuristic worked examples first
Fading of the initially introduced scaffold	Fading	1.00 (0.89)	0.81 (0.69)
	No fading	0.92 (0.72)	0.67 (0.68)

An ANCOVA with the pretest dialectic MAS as covariate revealed no significant main effect on the acquisition of dialectic MAS, neither for the presentation sequence of the two scaffolds, ($F(1,103) = 1.92$, $p = .17$, partial $\eta^2 = .02$) nor for the fading of the initially introduced scaffold, $F(1,103) = 0.77$, $p = .38$, partial $\eta^2 = .01$. Also, no interaction effect was found, ($F(1,103) = 0.04$, $p = .84$, partial $\eta^2 < .01$).

RQ2a: Prior knowledge as moderator for the effects of the presentation sequence and fading of scaffolds on the acquisition of dialogic and dialectic MAS

Prior knowledge did not significantly moderate the effect of the presentation sequence of the scaffolds on *dialogic* MAS ($b = 0.06$, 95% CI [-0.39,0.51], $p = .78$), and neither the effect of fading of the initially introduced scaffold on dialogic MAS ($b = -0.05$, 95% CI [-0.38,0.28], $p = .75$).

Concerning *dialectic* MAS, prior knowledge did not significantly moderate the effect of the presentation sequence of scaffolds ($b = -0.18$, 95% CI [-0.49, 0.14], $p = .27$) and neither the effect of fading of the initially introduced scaffold ($b = 0.14$, 95% CI [-0.18, 0.45], $p = .40$).

RQ2b: Working memory capacity as moderator for the effects of the presentation sequence and fading of scaffolds on the acquisition of dialogic and dialectic MAS

The moderator analyses with prior knowledge as covariate revealed that working memory capacity did not significantly moderate the effect of the presentation sequence of the two scaffolds on dialogic MAS ($b = -0.03$, 95% CI [-1.56, 1.50], $p = .97$), and neither the effect of fading of the initially introduced scaffold on dialogic MAS ($b = 0.30$, 95% CI [-0.82, 1.43], $p = .59$).

Regarding dialectic MAS, the moderator analyses with prior knowledge as covariate showed that the working memory capacity was no significant moderator for the effect of the presentation sequence ($b = -0.26$, 95% CI[-2.92, 2.41], $p = .85$). Yet, the effect of fading of the initially introduced scaffold was significantly moderated by the working memory capacity ($b = 3.64$, 95% CI[1.30, 5.98], $p < .01$, increase in R^2 due to interaction = .08.).

More detailed moderator analyses revealed an interesting pattern: Within the two groups with different presentation sequences of scaffolds, the effect of fading on dialectic MAS was differentially moderated by working memory capacity: For learners who were initially presented with the collaboration script, the effect of fading the script on dialectic MAS depended significantly on working memory capacity ($b = 9.21$, 95% CI [5.82,12.59],

$p < .001$, increase in R^2 due to interaction = .26). Post-hoc power-analysis revealed a power of $1 - \beta = .98$. Applying the Johnson-Neyman technique indicated that learners with low working memory capacity benefitted most from fading of the collaboration script. In contrast, learners with very high working memory capacity benefitted from the simultaneous availability of to the two scaffolds (see Appendix). For learners who were initially presented with the heuristic worked examples, their working memory capacity did not significantly moderate the effect of fading the heuristic worked examples ($b = 1.04$, 95% CI [-2.97, 5.05], $p = .60$, increase in R^2 due to interaction = .01).

Discussion

The aim of this study was to investigate the effects of the sequence and the fading of a collaboration script and heuristic worked examples on learners' development of dialogic and dialectic social-discursive MAS during a preparatory course for mathematics students at the transition from secondary to tertiary education. Furthermore, we were interested in the role that prior knowledge and working memory capacity played for learners' benefit from learning with the differently sequenced and faded scaffolds. We conceived dialogic MAS as activities that build on the learning partner's contribution in a concordant way such as expanding ideas of the learning partner. In contrast, dialectic activities of MAS were conceived as activities involving controversial discussions between learning partners. Our measures of dialogic and dialectic MAS were not correlated indicating that dialogic and dialectic activities can be separated as proposed by other authors (e.g., Schwarz & Shahar, 2017; Wegerif, 2008).

No indication for a general effect of the presentation sequence of scaffolds on dialogic and dialectic MAS

The findings of this study indicate that the sequence of introducing the collaboration script and heuristic worked examples had no effect on students' acquisition of dialogic and dialectic MAS. In contrast to the findings of previous studies (Clarke et al., 2005), presenting content-specific scaffolds (heuristic worked examples) first or content-independent scaffolds (collaboration scripts) first seems not to make a difference with respect to the development of social-discursive MAS. Since we had contradicting hypothesis about the direction of the effect of the presentation sequence of scaffolds there might have been a balanced amount of participant the did or did not benefit from one or the other presentation sequence of scaffolds. Therefore it might be more interesting to explore for which types of participants one of the two presentation sequences were more beneficial. This might be uncovered by learners' different pre-requisites and will be discussed in the subsequent sections.

The general effect of fading of scaffolds on dialogic MAS

Moreover, the findings show that the fading of the initially introduced scaffold had a positive effect on dialogic MAS. This replicates existing findings that fading is an important mechanism of scaffolding for learning (Pea, 2004) yet, in another way as it might be expected. Because both scaffolds predominantly addressed dialectic activities, fading the initially introduced scaffold might have reduced the amount of irrelevant information for acquiring dialogic MAS. In addition, the collaboration script and heuristic worked examples may have fostered dialogic MAS to a similar extent when they were introduced as first scaffold. Therefore, introducing the second scaffold might have been redundant with respect to dialogic MAS (see Kalyuga, 2007). In accordance with this interpretation, both the collaboration script and heuristic worked examples involved prompts to foster the extension of arguments. For example, collaboration scripts prompted learners to integrate different arguments which could have involved at least in parts extending the views of the learning partner. Heuristic worked examples prompted participants to build upon the ideas of a fictitious learner. This line of reasoning is further corroborated by the finding that the effect of fading of the primarily introduced scaffold on dialogic MAS was not moderated by learning prerequisites. Regardless of their prior knowledge and working memory capacity, the availability of the initially introduced scaffold seems to be redundant for learners when the second scaffold is introduced.

A further explanation for the effect of fading on dialogic MAS might be that learners prefer to use dialogic activities. In contrast to dialectic activities, dialogic activities might be perceived as more socially accepted than dialectic activities that might uncover weaknesses in the learning partners' and own knowledge base. When being presented with scaffolds that predominantly address dialectic activities, learners may possibly fall back into the (preferred) use of dialogic activities once one scaffold is faded out. In other words, fading scaffolds which mainly address dialectic activities might reduce the threshold to engage in dialogic activities. Hence, if the goal is to support dialogic MAS, the results of the present study suggest that learning environments may be designed with consecutively introduced heuristic worked examples and collaboration scripts which are faded out during the learning process.

The effect of fading for different presentation sequences on dialectic MAS – working memory capacity matters

Regarding the effects of fading of the initially introduced scaffold on the *dialectic* activities of MAS, varying results occurred for learners with different learning pre-requisites. The learners' existing knowledge structures concerning dialectic activities might have been activated in the collaboration script first condition when the collaboration script was introduced as first scaffold (Fischer et al., 2013). Due to the learners' experience with dialectic activities during the time the initially introduced collaboration script was present, the script might have become increasingly irrelevant in later learning phases. When in the second phase heuristic worked examples were introduced, the components of the collaboration script may have already been internalized and subsequently activated. But, if they were then still externally present (i.e. when the collaboration script was still present, after the heuristic worked example was introduced), the support provided by the collaboration script may have been redundant and possibly have overwhelmed learners with low working memory capacity. Therefore, when introducing heuristic worked examples in the second phase and simultaneously fading the collaboration script, particularly learners with low working memory capacity may benefit from a reduction of the interacting elements (i.e., components of the script; e.g., Sweller, 2010) in working memory.

Applying the Johnson Neyman technique further indicated that learners with very high working memory capacity benefitted from the simultaneous availability of both scaffolds. It seems that these learners can handle the high demands on working memory capacity and focus their attention on the not yet internalized parts of the collaboration script and heuristic worked examples for acquiring dialectic MAS. In line with this suggestion, research indicates that control of attention is an important aspect of working memory capacity to maintain information in a short-term storage and retrieving information from long-term memory (Shipstead, Lindsey, Marshall, & Engle, 2014).

The moderating role of working memory capacity might also be related to a high motivation to work on complex tasks. Learners voluntarily took part in the preparatory course and were presumably highly motivated to work on mathematical tasks. The high motivation may have lead learners to put a high demand on their working memory which may have been too high for learners with low working memory capacity. De Jong (2010) suggested that overload may only occur when learners work under time pressure or when offloading working memory (e.g., by taking notes) is prevented. However, the present study points to additional factors which may cause overload in working memory, one of which might be a high motivation to work on complex tasks.

Nevertheless, working memory capacity was not a moderator concerning the fading of the heuristic worked examples in the heuristic worked examples first conditions. Apparently, the availability of the heuristic worked examples induced no detrimental demands on working memory when the collaboration script was introduced. Heuristic worked examples reduced problem-solving demands considerably by providing relevant information regarding processes to solve mathematical problems in all treatment sessions. Some learners might have found this information more useful for gaining knowledge about dialectic activities, while others might have found it less useful. However, this information did not seem to induce too much irrelevant working memory load for learners with low working memory capacity. Also, learners with high working memory capacity might not have been able to benefit from the continued availability of heuristic worked examples. This might explain the nonsignificant main effect of fading of the heuristic worked examples on dialectic activities.

Finally, the finding that prior knowledge had no moderating influence stands in contrast to research about the Matthew effect (e.g., Stanovich, 1986) and the expertise reversal effect (e.g., Kalyuga et al., 2012). Neither learners with high prior knowledge nor learners with low prior knowledge benefitted more from fading of the collaboration script or heuristic worked examples. At least for the effect of the fading of the collaboration script on dialectic MAS, working memory capacity seems to be the more important moderator. However, the variance in the lower range of values of prior dialogic and dialectic MAS was low, probably due to the small range of possible values (only integer values were achievable). Thus, the moderating role of prior knowledge might not have been established across a broad range of values in prior knowledge.

Limitations and directions for future research

Several limitations of our study need to be mentioned. To start with, fading was implemented with a rather low granularity by completely removing one of the two scaffolds after two treatment sessions. After removing one scaffold, learners worked without it in the last two treatment sessions. However, research suggests that gradually removing solution steps from worked examples with individual progress may be more effective than completely fading out the worked example at once (for an overview, see Renkl, 2014). Furthermore, fading of collaboration scripts may require additional monitoring of peers to be effective (Wecker &

Fischer, 2011). Future research should investigate the effects of a more gradual fading of one or both scaffolds when they are combined depending on individual knowledge or demands on working memory capacity. Additionally, future studies may investigate the role of peer monitoring when fading one of the two scaffolds.

Also, differences in the effects on dialogic and dialectic activities between the two types of presentation sequences of scaffolds might have been reduced by the design of the study and therefore hard to find. Even though the two scaffolds mainly were designed to trigger dialectic activities, they also involved characteristics that may have triggered dialogic activities. The collaboration script, although mainly focusing on dialectic sequence of argumentation, also included dialogic aspects that were supposed to help learners construct joint arguments (e.g., when asked to develop syntheses). Likewise, the heuristic worked examples for learning partners in a dyad were slightly different from each other and might thus easily trigger dialectic activities. Despite that, from what we observed they often led learners to a convergent understanding in the end. Thus, also dialogic activities might have been supported by these scaffolds. Against this background, finding similar patterns of effects for the two outcome measures may not have been very surprising. In addition, since all learners received both scaffolds (although at different time points), the overall differences between the four conditions might have been too small to cause detectable effects on social-discursive MAS.

The low variance of prior knowledge due to a small range of possible values is a further limitation of the present study. Future studies may include tests that assess dialogic and dialectic MAS in a more differentiated way with more items. Furthermore, the tests to assess dialogic and dialectic MAS assessed rather declarative knowledge because participants were asked to describe phases and activities that appear in a prototypical discussion about a question. Further investigations should explore the effects of collaboration scripts and heuristic worked examples on social-discursive MAS by aid of more procedural measures.

Another limitation concerns the measurement of working memory capacity with only one task. Therefore, task-specific influences due to the context or material of the task could not be eliminated. To handle this problem, several tasks should be used to measure working memory capacity on a latent variable level (see also Miyake & Friedman; 2012; Schwaighofer et al., 2017). Using several tasks to measure working memory capacity might also allow for a more reliable detection of moderation effects of working memory capacity by minimizing task-specific residual variance.

Yet, despite the relatively low reliability of the operation span task measuring working memory capacity, the reliability allowed to identify a large moderation effect with sufficient power. However, with respect to the moderation analyses, a limitation lies in the relatively small sample sizes for the comparisons of the conditions with or without fading of the initially introduced scaffold. Accordingly, the statistical power to detect small effects in addition to the large effect of working memory capacity might have been insufficient.
Also, since participation in the preparatory course was voluntary, it may well be that self-selection may have influenced our sample. In other words, we cannot rule out that our learners had particular cognitive abilities (e.g., a higher working memory capacity) or motivational preconditions (e.g., a higher interest in mathematics or different goal orientations) than students who did not choose to participate in the course. It might thus be fruitful to replicate our study in a context that leaves fewer opportunities for a self-selection bias.
Finally, as many other studies on CSCL scripts, we did not check how exactly the students understood the different script prompts. It may well be that different learners “appropriate” (Tchounikine, 2017) the script differently and these differences may yield differential effects on learning outcomes afterwards. It would be extremely interesting if future research would yield insights into how exactly such appropriation processes emerge during collaboration with a script.

Conclusion

The findings of this study reveal little support for the assumption that one specific sequence of introducing heuristic worked examples in addition to collaboration scripts in the context of mathematical argumentation and proof would be superior to another sequence. For designers of CSCL and non-CSCL environments, this finding might be welcome since it implies that pondering about the sequence of how different scaffolds are presented might be not particularly important. The findings do however support the claim that having two scaffolds available at a time can be overwhelming, and that this depends on an individual’s cognitive learning prerequisites. Thus, more support does not necessarily result in better learning of argumentation, and inter-individual differences in working memory capacity need to be considered. More specifically, this study showed evidence that learners with less favorable working memory capacity benefit when the more domain-general scaffold *collaboration* *script* is presented first and faded out when the more content related scaffold *heuristic worked examples* is presented in a second phase.

As a practical consequence, in order to individualize learning environments and adapt the environments to the learners' prerequisites, it would make sense to measure not only content-related learning prerequisites such as domain-specific prior knowledge, but also more domain-general pre-requisites such as working memory capacity. Knowing about these individual prerequisites of the learners is a necessary precondition to be able to provide adaptive support in the next step (e.g., Deiglmayr & Spada, 2010).

Appendix

Conditional effect of the fading of the collaboration script on dialectic MAS for different values of working memory capacity

Working memory capacity (raw score)	Effect of the fading of the collaboration script	SE	*t*	*p*	95% CI	
					LL	UL
.3421	-4.5289	1.1811	-3.8344	.0004	-6.911	-0.5593
.3743	-4.2284	1.1066	-3.821	.0004	-6.4601	-0.5173
.466	-3.9279	1.0322	-3.8052	.0004	-6.0096	-0.475
.4388	-3.6273	0.958	-3.7862	.0005	-5.5594	-0.4323
.4711	-3.3268	0.8841	-3.763	.0005	-5.1097	-0.3891
.533	-3.0262	0.8104	-3.7341	.0005	-4.6606	-0.3453
.5355	-2.7257	0.7372	-3.6975	.0006	-4.2123	-0.3006
.5678	-2.4251	0.6644	-3.6499	.0007	-3.7651	-0.2548
.6	-2.1246	0.5924	-3.5863	.0009	-3.3193	-0.2076
.6322	-1.824	0.5214	-3.4983	.0011	-2.8756	-0.1582
.6645	-1.5235	0.4519	-3.3713	.0016	-2.4349	-0.1058
.6967	-1.223	0.3847	-3.1791	.0027	-1.9988	-0.0489
.7289	-0.9224	0.3212	-2.8716	.0063	-1.5702	0
.7612	-0.6219	0.2642	-2.3537	.0232	-1.1547	0.0148
.7755	-0.4884	0.2422	-2.0167	.05	-0.9768	0.0889
.7934	-0.3213	0.2188	-1.4687	.1492	-0.7625	0.1784
.8257	-0.0208	0.1933	-0.1075	.9149	-0.4105	0.2876
.8579	0.2798	0.1956	1.4302	.1599	-0.1147	0.4181
.8724	0.4153	0.2059	2.0167	.05	0	0.5669
.891	0.5803	0.225	2.5797	.0134	0.1266	0.7035
.9224	0.8809	0.2727	3.2299	.0024	0.3309	0.7289
.9546	1.1814	0.331	3.569	.0009	0.5138	0.8998
.9868	1.4819	0.3952	3.7496	.0005	0.6849	1.0764

References

Asterhan, C. S. C. & Schwarz, B. B. (2007). The effects of monological and dialogical argumentation on concept learning in evolutionary theory. *Journal of Educational Psychology*, *99*, 626–639. doi: 10.1037/0022-0663.99.3.626

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. doi:10.1111/j.1551-6709.2009.01017.x.

Alexander, P. A., & Jetton, T. L. (2003). Learning from traditional and alternative texts: new conceptualization for an information age. In A. Graesser, M. Gernsbacher, & S. Goldman (Eds.), *Handbook of discourse processes* (pp. 199–241). Mahwah, NJ: Erlbaum.

Baddeley, A. D., Allen, R. J., & Hitch, G. J. (2011). Binding in visual working memory: The role of the episodic buffer. *Neuropsychologia, 49,* 1393–1400. doi: 10.1016/j.neuropsychologia.2010.12.042

Bausch, I., Biehler, R., Bruder, R., Fischer, P. R., Hochmuth, R., Koepf, W., … Wassong, T. (Eds.). (2014). *Mathematische Vor- und Brückenkurse [Mathematical preparatory and bridging courses]*. Wiesbaden: Springer Fachmedien Wiesbaden.

Boero, P. (1999). Argumentation and mathematical proof: a complex, productive, unavoidable relationship in mathematics and mathematics education. *International Newsletter on the Teaching and Learning of Mathematical Proof*, *7*(8).

Bühner, M., Kröner, S., & Ziegler, M. (2008). Working memory, visual–spatial-intelligence and their relationship to problem-solving. *Intelligence*, *36*(6), 672–680. doi:10.1016/j.intell.2008.03.008

Chi, M. T. H. & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, *49* (4), 219–243. doi: 10.1080/00461520.2014.965823.

Clark, D. B., D'Angelo, C. M., & Menekse, M. (2009). Initial Structuring of Online Discussions to Improve Learning and Argumentation: Incorporating Students' Own Explanations as Seed Comments Versus an Augmented-Preset Approach to Seeding Discussions. *Journal of Educational Science and Technology, 18*, 321–333. doi:10.1007/s10956-009-9159-1

Clarke, T., Ayres, P., & Sweller, J. (2005). The impact of sequencing and prior knowledge on learning mathematics through spreadsheet applications. *Educational Technology Research and Development*, *53*(3), 15–24. doi:10.1007/BF02504794

CCSSI. (2017). Common Core State Standards for Mathematics. Retrieved from http://www.corestandards.org/Math/Practice/#CCSS.Math.Practice.MP1

Collins, A., Brown, J. S., & Holum, A. (1991). Cognitive apprenticeship: Making thinking visible. *American educator*, *15*(3), 6-11. doi:10.1.1.124.8616

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review*, *3*(4), 422–433. http://doi.org/10.3758/BF03214546

Dawkins, P. C., & Weber, K. (2016). Values and norms of proof for mathematicians and students. *Educational Studies in Mathematics*. https://doi.org/10.1007/s10649-016-9740-5

Deiglmayr, A. & Spada, H. (2010). Developing adaptive collaboration support: the example of an effective training for collaborative inferences. *Educational Psychology Review, 22*(1), 103-113. doi: 10.1007/s10648-010-9119-6

de Jong, T. (2010). Cognitive load theory, educational research, and instructional design: some food for thought. *Instructional Science*, *38*(2), 105–134. https://doi.org/10.1007/s11251-009-9110-0

Dillenbourg, P. & Jermann, P. (2007). Designing integrative scripts. In F. Fischer, I. Kollar, H. Mandl, & J. M. Haake (Eds.), Scripting computer-supported collaborative learning. Cognitive, computational and educational Perspectives. (Vol. 6, pp. 275–301). New York, NY: Springer.

Dochy, F., Segers, M., & Buehl, M. M. (1999). The relation between assessment practices and outcomes of studies: The case of research on prior knowledge. *Review of Educational Research*, *69*(2), 145–186. doi:10.2307/1170673

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. doi:10.1080/00461520.2012.748005

Hailikari, T., Nevgi, A., & Komulainen, E. (2008). Academic self-beliefs and prior knowledge as predictors of student achievement in mathematics: a structural model. *Educational Psychology*, *28*(1), 59–71. https://doi.org/10.1080/01443410701413753

Hanna, G. (2000). Proof, explanation and exploration: an overview. *Educational Studies in Mathematics, 44,* 5-23. doi:10.1023/A:1012737223465

Hayes, A. F. (2012). PROCESS: A versatile computational tool for observed variable mediation, moderation, and conditional process modeling [White paper]. Retrieved from http://www.afhayes.com/public/ process2012.pdf

Hayes, A. F. (2013). Introduction to mediation, moderation, and conditional process analysis: A regression-based approach. New York, NY: Guilford Press.

Heinze, A., Reiss, K., & Rudolph, F. (2005). Mathematics achievement and interest from a differential perspective. *Zentralblatt für Didaktik der Mathematik*, *37*(3), 212–220. http://dx.doi.org/10.1007/s11858-005-0011-7.

Hodds, M., Alcock, L., & Inglis, M. (2014). Self-explanation training improves proof comprehension. *Journal for Research in Mathematics Education*, *45*(1), 62–101. doi:10.5951/jresematheduc.45.1.0062

Hudson, H. T. & Rottmann, R. M. (1981). Correlation between performance in physics and prior mathematics knowledge. *Journal of Research in Science Teaching*, *18*(4), 291–294. doi:10.1002/tea.3660180403

Jahnke, H. N., & Ufer, S. (2015). Argumentieren und Beweisen. In R. Bruder, L. Hefendehl-Hebeker, B. Schmidt-Thieme, & H.-G. Weigand (Eds.), *Handbuch der Mathematikdidaktik* (pp. 331–355). Berlin, Heidelberg: Springer Berlin Heidelberg. http://link.springer.com/10.1007/978-3-642-35119-8_12

Jiménez-Aleixandre, M.P., Rodríguez, A.B., & Duschl, R.A. (2000). ‘‘Doing the lesson’’ or ‘‘doing science:’’ Argument in high school genetics. *Science Education*, *84*, 757–792. doi:10.1002/1098-237X(200011)84:6<757::AID-SCE5>3.0.CO;2-F

Jonassen, D. H., & Kim, B. (2010). Arguing to learn and learning to argue: design justifications and guidelines. *Educational Technology Research and Development*, *58*(4), 439–457. doi: 10.1007/s11423-009-9143-8.

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review*, *19*(4), 509–539. doi: 10.1007/s10648-007-9054-3

Kalyuga, S. (2013). Effects of learner prior knowledge and working memory limitations on multimedia learning. *Procedia - Social and Behavioral Sciences*, *83*, 25–29. doi:10.1016/j.sbspro.2013.06.005

Kalyuga, S., Rikers, R., & Paas, F. (2012). Educational implications of expertise reversal effects in learning and performance of complex cognitive and sensorimotor skills. *Educational Psychology Review*, *24*(2), 313–337. doi: 10.1007/s10648-012-9195-x

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General*, *133*(2), 189–217. doi:10.1037/0096-3445.133.2.189

King, A. (2007). Scripting collaborative learning processes: a cognitive perspective. In F. Fischer, I. Kollar, H. Mandl, & J. M. Haake (Eds.), Scripting computer-supported collaborative learning: Cognitive, computational, and educational perspectives (pp. 13–37). New York: Springer.

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

Kollar, I., Ufer, S., Reichersdorfer, E., Vogel, F., Fischer, F., & Reiss, K. (2014). Effects of collaboration scripts and heuristic worked examples on the acquisition of mathematical argumentation skills of teacher students with different levels of prior achievement. *Learning and Instruction, 32*(1), 22–36. doi:10.1016/j.learninstruc.2014.01.003

Leitão, S. (2000). The potential of argument in knowledge building. *Human Development*, 43, 332e360. doi: 10.1159/000022695.

Leppink, J., Broers, N. J., Imbos, T., van der Vleuten, C. P. M., & Berger, M. P. F. (2012). Prior knowledge moderates instructional effects on conceptual understanding of statistics. *Educational Research and Evaluation*, *18*(1), 37–51. doi:10.1080/13803611.2011.640873

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

Peng, P., Namkung, J., Barnes, M., & Sun, C. (2016). A meta-analysis of mathematics and working memory: Moderating effects of working memory domain, type of mathematics skill, and sample characteristics. *Journal of Educational Psychology*, *108*(4), 455–473. /doi: 10.1037/edu0000079

Rach, S. & Heinze, A. (2016). The transition from school to university in mathematics: Which influence do school-related variables have? *International Journal of Science and Mathematics Education. doi:10.1007/s10763-016-9744-8.*

Redick, T. S., Broadway, J. M., Meier, M. E., Kuriakose, P. S., Unsworth, N., Kane, M. J., & Engle, R. W. (2012). Measuring working memory capacity with automated complex span tasks. *European Journal of Psychological Assessment, 28*(3)*,* 164–171. doi:10.1027/ 1015-5759/a000123

Redick, T. S., Unsworth, N., Kelly, A. J., & Engle, R. W. (2012). Faster, smarter? Working memory capacity and perceptual speed in relation to fluid intelligence. *Journal of Cognitive Psychology, 24,* 844–854. doi: 10.1080/20445911.2012.704359

Reiss, K., Heinze, A., Renkl, A., & Große, C. (2008). Reasoning and proof in geometry. Effects of a learning environment based on heuristic worked-out examples. ZDM. *The International Journal on Mathematics Education*, *40*(3), 455–467. doi:10.1007/s11858-008-0105-0.

Reiss, K., & Renkl, A. (2002). Learning to prove: the idea of heuristic examples. *Zentralblatt für Didaktik der Mathematik*, *34*(1), 29–3. doi:10.1007/BF02655690.

Renkl, A. (2014). Toward an instructionally oriented theory of example-based learning. *Cognitive Science*, *38*(1), 1–37. doi:10.1111/cogs.12086

Renkl, A., & Atkinson, R.K. (2007). An example order for cognitive skill acquisition. In F.E. Ritter, J. Nerb, E. Lehtinen & T.M. O'Shea (Eds.), *In Order to Learn. How the Sequence of Topics Influences Learning* (pp. 95-105). New York: Oxford University Press.

Renkl, A., Hilbert, T., & Schworm, S. (2009). Example-based learning in heuristic domains: a cognitive load theory account. *Educational Psychology Review*, *21*, 67–78. doi: 10.1007/s10648-008-9093-4.

Rummel, N., Mullins, D., & Spada, H. (2012). Scripted collaborative learning with the cognitive tutor algebra. *International Journal of Computer-Supported Collaborative Learning*, *7*(2), 307–339. doi:10. 1007/s11412-012-9146-z

Sadler, T. D. (2004). Informal reasoning regarding socioscientific issues: A critical review of research. *Journal of Research in Science Teaching*, *41*(5), 513–536. doi:10.1002/tea.20009.

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/edu0000115

Schwaighofer, M., Bühner, M., & Fischer, F. (2017). Executive functions in the context of complex learning: Malleable moderators? *Frontline Learning Research, 5*(1), 58-75. doi:10.1037/edu0000115

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

Schwarz, B. B. (2009). Argumentation and learning. In Muller-Mirza and A-N. Perret-Clermont (Eds.), *Argumentation and Education – Theoretical Foundations and Practices* (pp. 91-126). Springer Verlag.

Schwarz, B. B. & Shahar, N. (2017). Combining the dialogic and the dialectic: putting argumentation into practice for classroom talk. *Learning, Culture and Social Interaction, 12*, 113-132. doi:10.1016/j.lcsi.2016.12.003

Schellens, T., Van Keer, H., De Wever, B., & Valcke, M. (2007). Scripting by assigning roles: Does it improve knowledge construction in asynchronous discussion groups? *International Journal of Computer-Supported Collaborative Learning*, *2*(2–3), 225–246. doi: 10.1007/s11412-007-9016-2

Selden, J., Benkhalti, A., & Selden, A. (2014). An analysis of transition-to-proof course students' proof constructions with a view towards course redesign. *Proceedings of the 17thAnnual Conference on Research in Undergraduate Mathematics Education*.

Shipstead, Z., Lindsey, D. R. B., Marshall, R. L., & Engle, R. W. (2014). The mechanisms of working memory capacity: Primary memory, secondary memory, and attention control. *Journal of Memory and Language*, *72*, 116–141.doi: 10.1016/j.jml.2014.01.004

Sommerhoff, D., Ufer, S., & Kollar, I. (2015). Research on mathematical argumentation: A descriptive review of PME proceedings. In K. Beswick, T. Muir, & J. Wells (Eds.), *Proceedings of the 39th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 193–200). Hobart, Australia: PME.

Spiro, R. J., Coulson, R. L., Feltovich, P. J., & Anderson, D. K. (1988). Cognitive flexibility theory: Advanced knowledge acquisition in illstructured domains. In *Proceedings of the Tenth Annual Conference of the Cognitive Science Society* (pp. 375–383). Hillsdale, NJ: Erlbaum.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, *21*(4), 360–407. doi:10.1598/RRQ.21.4.1

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review*, *22*(2), 123–138. doi: 10.1007/s10648-010-9128-5

Sweller, J. (2011). Cognitive load theory. In J. Mestre, & B. Ross (Eds.), *The psychology of learning and motivation: Cognition in education* (Vol. 55, pp. 37 – 76). Oxford: Academic Press.

Tabak, I. (2004). Synergy: a complement to emerging patterns of distributed scaffolding. *The Journal of the Learning Sciences*, *13*(3), 305e335. doi: 10.1207/s15327809jls1303_3.

Tchounikine, P. (2016). Contribution to a theory of CSCL scripts: taking into account the appropriation of scripts by learners. *International Journal for Computer-Supported Collaborative Learning, 11*(3), 349-369. doi: 10.1007/s11412-016-9240-

Teasley, S. D. (1997). Talking about reasoning: How important is the peer in peer collaborations? In C. O'Malley (Ed.), *Disourse, Tools, and Reasoning: Situated Cognition and Technologically Supported Environments* (pp. 361–384). Berlin: Springer.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*(3), 498–505. doi:10.3758/BF03192720

Vogel, F., Kollar, I., Ufer, S., Reichersdorfer, E., Reiss, K., & Fischer, F. (2016a). Developing argumentation skills in mathematics through computer-supported collaborative learning: the role of transactivity. *Instructional Science, 44*(5), 477–500. doi:10.1007/s11251-016-9380-2

Vogel, F., Wecker, C., Kollar, I., & Fischer, F. (2016b). Socio-Cognitive Scaffolding with Computer-Supported Collaboration Scripts: a Meta-Analysis. *Educational Psychology Review*, doi:10.1007/s10648-016-9361-7

Vollstedt, M., Heinze, A., Gojdka, K. & Rach, S. (2014). Framework for examining the transformation of mathematics and mathematics learning in the transition from school to university. In S. Rezat, M. Hattermann & A. Peter-Koop (Hrsg.), *Transformation - A Fundamental Idea of Mathematics Education* (p. 29-50). New York: Springer.

Webb, N. M., Nemer, K. M., & Zuniga, S. (2002). Short circuits or superconductors? Effects of group composition on high-achieving students' science assessment performance. *American Educational Research Journal*, *39*(4), 943-989. doi:10.3102/00028312039004943

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

Wegerif, R. (2008). Dialogic or dialectic? The significance of ontological assumptions in research on educational dialogue. *British Educational Research Journal*, *34*(3), 347–361. doi:10.1080/01411920701532228

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. doi:10.1016/j.chb.2009.08.007

Weiß, R.H. (2006). Grundintelligenztest Skala 2 Revision, CFT 20-R. Göttingen: Hogrefe

Yackel, E., & Cobb, P. (1996). Sociomathematical norms, argumentation, and autonomy in mathematics. *Journal for Research in Mathematics Education*, *27*(4), 458–477. doi:10.2307/749877