Effects of Collaboration Scripts and Heuristic Worked Examples on the Acquisition of Mathematical Argumentation Skills of Teacher Students with Different Levels of Prior Achievement

Ingo Kollar[1], Stefan Ufer[2], Elisabeth Reichersdorfer[3], Freydis Vogel[1], Frank Fischer[1] & Kristina Reiss[3]

[1]*University of Munich, Chair of Education and Educational Psychology, Leopoldstraße 13, 80802 Munich, Germany*

[2]*University of Munich, Chair of Mathematics Education, Theresienstraße 39, 80333 Munich, Germany*

[3]*Technische Universität München, Heinz Nixdorf-Chair for Mathematics Education, Schellingstraße 33, Munich, Germany*

Corresponding author: Stefan Ufer, ufer@math.lmu.de, phone +49 89 2180 4451

This research is part of the project ELK-Math, funded by the Deutsche Forschungsgemeinschaft (DFG) under grant RE 1247/9-1 and FI 792/7-1. Ingo Kollar and Stefan Ufer contributed equally to this work as first authors. This contribution further contains work that is part of the doctoral dissertations of Elisabeth Reichersdorfer and Freydis Vogel under the supervision of Kristina Reiss and Frank Fischer, respectively.

Highlights

- MAS can be fostered with heuristic worked examples and collaboration scripts.
- There are only small indications of synergistic scaffolding when both scaffolds are combined.
- Effects of both scaffolds were dependent on general learning prerequisites.
- Combining instructional scaffolds runs the danger of overstraining learners.

Abstract

A challenging demand for mathematics teacher students is to produce acceptable scientific mathematical argumentations. We investigated to what extent mathematics teacher students who collaborated in dyads with different levels of prior achievement can be supported in their development of mathematical argumentation skills by two different instructional approaches that were systematically varied in a 2x2-factorial design: collaboration scripts (with vs. without) and heuristic worked examples vs. problem solving. An experimental study was run in the context of a two-weeks preparatory course for beginning mathematics teacher students (N = 101). Mathematical argumentation skills were conceptualized as consisting of an individual-mathematical and a social-discursive component. Results indicated positive effects of both scaffolds on the social-discursive component. Moreover, the effects of both scaffolds on both components were dependent on learners' prior achievement (high school GPA). Heuristic worked examples and collaboration scripts were particularly effective in the facilitation of mathematical argumentation skills for teacher students with higher general learning prerequisites. Possible process-based explanations for this pattern of results as well as ways to more specifically address the needs of teacher students with lower prior achievement are discussed.

Keywords

- Mathematical argumentation skills
- Collaboration scripts
- Heuristic worked examples
- Synergistic scaffolding
- General prior achievement
- Computer-supported collaborative learning

1. Introduction

The ability to construct arguments for and against mathematical claims and to generate or inquire mathematical conjectures has shifted into the focus of mathematics curricula worldwide during the last decade (e.g., National Governors Association Center for Best Practices, 2010). Mathematics teachers are thus demanded to help students acquire skills and competences related to mathematical argumentation. However, upon entering university education, many teacher students do not have the corresponding skills at their disposal to a sufficient extent. In other words, their capability to master mathematical argumentation varies with their overall *prior achievement*, that is with differences in the GPA they achieved at high school (e.g., Blömeke, Suhl, Kaiser, & Döhrmann, 2012). Since conveying mathematical argumentation skills (MAS) to mathematics teacher students is an important educational goal, it requires exploration about how to support them in the acquisition of MAS. Given the diversity of mathematics teacher students' learning prerequisites, it is also important to know to what extent instruction must to be tailored to the needs of students with lower vs. higher prior achievement.

We conceptualize MAS as the ability to inquire mathematical conjectures individually or in collaborative contexts, finally arriving at a proof or refutation for the conjecture (e.g., Koedinger, 1998). We propose to distinguish at least two components of MAS: a domain-specific, *individual-mathematical* and a domain-general, *social-discursive* component. The individual-mathematical component refers to the individual ability to generate arguments for or against a mathematical conjecture, to evaluate these arguments according to mathematical criteria, and to select and combine these arguments for a mathematical proof or refutation (Heintz, 2000; Reichersdorfer, Vogel, Fischer, Kollar, Reiss, & Ufer, 2012). The social-discursive component refers to the ability to participate in collaborative argumentation processes in social situations (Kollar, Fischer, & Slotta, 2007). Of course, expertise in MAS also includes domain-specific, social-discursive practices and skills (e.g., Yackel & Cobb, 1996) such as checking each others' arguments according to mathematical standards, which are at the interface of the two components described above. The current study was interested in contrasting domain-general and domain-specific aspects of MAS and will thus focus only on *individual-mathematical* and *social-discursive* aspects.

Over the past decade, a lot of research has investigated the effects of scaffolds directed at helping learners acquire social-discursive argumentation skills, especially in the context of Computer-Supported Collaborative Learning (CSCL). There, the collaboration script approach (e.g., Dillenbourg & Jermann, 2007; Fischer, Kollar, Stegmann, & Wecker, 2013) has been shown to be particularly effective (e.g., Noroozi, Weinberger, Biemans, Mulder, & Chirazi, 2013; Scheuer, McLaren, Weinberger, & Niebuhr, 2013; Rummel & Spada, 2005; Stegmann, Wecker, Weinberger, & Fischer, 2012; Wecker & Fischer, 2011). While such scripts have typically been effective in fostering social-discursive aspects of argumentation skills, they rarely had additional positive effects on domain-specific learning outcomes. If MAS are considered as including both a social-discursive and an individual-mathematical component, it therefore seems promising to combine the presentation of collaboration scripts with domain-specific scaffolding techniques. This, however, has hardly been investigated systematically in prior research, especially in mathematics.

One candidate for fostering the individual component of MAS are heuristic worked examples (e.g., Atkinson, Catrambone, & Merrill, 2003; Paas & van Merrienboer, 1994; Schwonke, Renkl, Krieg, Wittwer, Aleven, & Salden, 2009). In a study that aimed at fostering MAS in grade 8 students, Hilbert, Renkl, Kessler and Reiss (2008) showed that studying heuristic worked examples was more effective than studying an instructional text on geometry to foster conceptual knowledge on mathematical argumentation and individual MAS. However, although heuristic worked examples have already been used to support small groups of learners (Reiss, Heinze, Kessler, Rudolph-Albert, & Renkl, 2007), their effectiveness should be amplified when scaffolds that particularly aim at an improvement of social-discursive aspects of their argumentation support collaborative example elaboration.

This article reports an empirical study with mathematics teacher students at the start of their university education. The study investigated whether heuristic worked examples can successfully be combined with computer-supported collaboration scripts to foster students' MAS, with a particular focus on whether teacher students with different levels of prior achievement benefit equally from these two interventions.

1.1 The role of prior achievement as an individual learning prerequisite

Teacher students typically start academic education shortly after their secondary school degree, and there is considerable variance in their prior achievement, that is in their high school GPAs (Blömeke et al., 2012). Based on a meta-analysis of more than 800 studies to identify the main variables that affect later achievement, Hattie (2009) found prior achievement to be among the most influential predictors, with an average effect size of d = .67 (for the transition from high school to university, see Kuncel, Hezlett, & Ones, 2001). In research on Aptitute-Treatment-Interactions (ATI), this phenomenon is often interpreted as a "Matthew effect" meaning that students with higher prior achievement benefit more from a given kind of instruction than learners with lower prior achievement (see Stanovich, 1986). This can be explained by the assumption that prior achievement goes along with the level of prior knowledge a student has accumulated. Students with higher levels of prior knowledge have a higher chance to identify relevant information (Alexander & Jetton, 2003), to connect this information to existing schemata, and to integrate new information more easily into their existing knowledge structures in long-term memory. Even more detailed predictions can be derived from the Construction-Integration Model (Kintsch, 1998). As Scheiter and Gerjets (2007) point out, learners with low prior knowledge require instructional texts that present the micro- and macrostructure of a text very clearly. High prior knowledge students, in contrast, tend to benefit from less coherent texts. Nevertheless, Barab, Bowdish and Lawless (1997) argue that this prediction is restricted to learning tasks which require text comprehension and that dependence on prior knowledge should be less pronounced for tasks which require independent information processing and problem solving. In summary, these lines of reasoning lead to the hypothesis that the advantage of students with higher levels of prior achievement compared to learners with lower levels of prior achievement (Matthew effect) will be more pronounced when the learning environment requires to extract and integrate information from texts compared to environments that rely less on text comprehension. Applied to learning from heuristic worked examples and collaboration scripts, which both are typically presented in a textual format (Reiss & Renkl, 2002; Kollar et al., 2007), it may thus be expected that learners with higher prior achievement will be in an advantageous position compared to learners with lower prior achievement.

However, research on the “expertise-reversal effect” (Kalyuga, Rikers, & Paas, 2012) seems to suggest the contrary: For example, Rey and Buchwald (2011) have shown that more structured scaffolds (in their case a combination of text and animations that was presented to learners who were supposed to acquire knowledge on a mathematical optimization algorithm) were particularly effective for students with lower rather than high levels of prior knowledge. Learners with higher prior knowledge were better off when they were only presented with text (and no animation). The usual interpretation for such an effect is that if a learner already has the knowledge necessary to solve a certain type of tasks, information provided in a scaffold becomes redundant and produces extraneous load (Sweller, 2010) which is negatively related to knowledge acquisition. This line of reasoning would predict a negative influence of high prior knowledge on learning gain in learning environments with scaffolds that are textually represented, which is typical for worked examples and collaboration scripts.

Transferred to our study, it seems unclear what role prior achievement will play when students are provided with scaffolds targeting MAS: The Matthew effect argumentation predicts that providing students with collaboration scripts and heuristic worked examples will be especially beneficial for students with higher levels of prior achievement. Yet, the expertise-reversal position would predict the contrary. It should be noted that in research on the Matthew effect often quite general learning prerequisites like prior school achievement are considered. In contrast, more specific prior skills are usually considered in the worked example research tradition.

1.2 Facilitating MAS with collaboration scripts

The social-discursive component of MAS is necessary to communicate ideas and solutions for mathematical problems to others and to reach joint solutions based on group discussions. Collaborative learning is regarded a promising approach to foster the corresponding skills (e.g., Slavin, 1996). However, a wealth of evidence demonstrates that collaborative learning is often less effective than individual learning, especially when it is not structured appropriately (e.g., Gillies, 2004). In absence of guidance, collaborators often engage in low-level collaborative processes, which is reflected by producing few or only superficial questions and explanations (see King, 2007) or by showing low-level argumentation (e.g., Kollar et al., 2007).

One way to structure collaborative argumentation is to provide learners with collaboration scripts which are defined as interventions that specify, distribute and sequence learning activities and collaboration roles among the learners of a small group (Kollar, Fischer, & Hesse, 2006). According to the Script Theory of Guidance that was recently proposed by Fischer et al. (2013), such external scripts enable students to engage in learning activities at a level beyond their current abilities (external script guidance principle, p. 61). As research has shown, collaboration scripts can be tailored to evoke specific collaboration processes that stand in a positive relation to individual learning outcomes. For example, Rummel and Spada (2005; see also Rummel, Spada, & Hauser, 2009) developed a collaboration script that structured the collaboration of dyads with two learners of complementary expertise. Amongst others, this script prompted learners to explicitly coordinate their activities and told them when to work individually and when to share and discuss their thoughts to arrive at a joint conclusion for an authentic case that could only be solved successfully if the two learners combined their different areas of expertise. When compared to dyads who learned without the script, Rummel and Spada (2005) found that scripted learners reached higher-quality task solutions and acquired more collaboration skills.

In a study more specifically related to argumentation, Stegmann, Weinberger and Fischer (2007) used two scripts. Both scripts were integrated in an asynchronous discussion board environment in which triads of learners applied a psychological theory to authentic problem cases. The *script for the construction of single arguments* prompted learners to use data, grounds and warrants (Toulmin, 1958) while building arguments, whereas the *script for the construction of argument sequences* distributed the task to produce arguments, counterarguments and integrations (Leitao, 2000) among the learners. Both scripts had positive effects on the quality of argumentation during collaboration. Further, both scripts fostered the acquisition of domain-general knowledge on argumentation, and their effects on this measure added up. Also, studies by Hämäläinen, Oksanen and Häkkinen (2008), Noroozi et al. (2013), Schellens, van Keer, De Wever and Valcke (2007), Scheuer et al. (2013) and Schoonenboom (2008) showed that collaboration scripts effectively support collaboration, and typically, improved collaboration processes go along with an increased acquisition of domain-general skills.

With respect to the facilitation of domain-*specific* knowledge, empirical evidence for the effectiveness of collaboration scripts however is at best mixed: Several studies found positive effects on domain-specific learning outcomes (e.g., knowledge about the content of discussions; e.g., Wecker, Kollar, Fischer, & Prechtl, submitted; Weinberger, Ertl, Fischer, & Mandl, 2005). However other studies only found marginal or no effects at all concerning the acquisition of domain-specific knowledge (e.g., Asterhan, Schwarz, & Gil, 2012; Kollar et al., 2007; Stegmann et al. , 2007). These findings imply that collaboration scripts do not necessarily suffice to support domain-specific knowledge acquisition. Given the domain-unspecific independent character of collaboration scripts, this is not very surprising. Yet, the mixed empirical results underline the need to additionally explore further instructional interventions that address the peculiarities of the specific domain under study and might thus be used in combination with collaboration scripts to foster domain-specific learning outcomes (see also Vogel, Kollar, & Fischer, 2012).

Another avenue for future research comes from the lack of studies that checked the possibly diverse effects of collaboration scripts for learners with different levels of prior achievement (Hattie, 2009). As described earlier, it is an open question whether collaboration scripts will especially be effective for learners with high levels of prior achievement (supporting the Matthew effect argumentation) or whether they will especially support learners with lower levels of prior achievement, which would provide evidence for an expertise-reversal effect.

1.3 Facilitating MAS with heuristic worked examples

Research describing skill development as schema acquisition has substantiated the effectiveness of example-based learning to provide domain-specific scaffolding in individual settings (van Gog & Rummel, 2010). This comprises learning *modelling examples* with a human model (e.g., Bandura, 1977; Collins, Brown, & Newman, 1989; McLaren, Lim, Gagnon, Yaron, & Koedinger, 2008) or an electronic agent (Wouters, Paas, & van Merrienboer, 2010) as well as approaches based on studying *worked examples* (Atkinson, Derry, Renkl, & Wortham, 2000; Sweller, 2004; van Loon-Hillen, van Gog, & Brand-Gruwel, 2010). Worked examples present learners with a problem, a description of all solution steps, and the correct solution to the problem. Their superiority over unguided problem solving is usually explained by Cognitive Load Theory (CLT; Sweller, 2010). CLT assumes that individual working memory resources are restricted and can be allocated to three different types of cognitive load. (1) *Intrinsic load* refers to the information that is necessary to deal with the task at hand (e.g., a mathematical argumentation problem). Sweller (2010) conceptualizes it by the extent to which different memory elements have to be considered simultaneously (element interactivity) when dealing with the intrinsic complexity of the task and its solution. The amount of information that constitutes an element depends on the learner's prior knowledge. For example, it is well known that experts are able to encode related bits of information into single memory elements called "chunks" (Miller, 1956). Thus, Intrinsic load depends on the complexity of the task (which in turn is dependent on the learner's prior knowledge) and cannot be changed by instructional interventions. (2) *Extraneous load* describes working memory load caused by elements of the learning environment that are neither necessary to solve the task nor for schema acquisition. For example, when learners rely on weak problem solving strategies like means-end-analysis, they have to consider many different states of the problem space simultaneously. This consumes cognitive resources that cannot be invested in an acquisition of schemata (Sweller, 2010). Finally, cognitive load that is directed at the construction of new schemata is called (3) *germane load* (for recent critical views on the role of germane load, see Kalyuga, 2012). To be effective, instruction should thus be designed in a way that extraneous load is reduced and germane load is increased, e.g. by providing learners with worked examples instead of having them solve problems on their own (*worked example effect*; Sweller, 2010). Nevertheless, reducing extraneous load can provide additional resources, but additional means are necessary to ensure that these resources are allocated to schema construction. Research has shown that interventions such as self-explanation prompts (e.g., Bielaczyc, Pirolli, & Brown, 1995; Chi, de Leeuw, Chiu, & Lavancher, 1994; Renkl, 2002; Rittle-Johnson, 2006) can prevent superficial processing of worked examples. Completion gaps (Stark, 1999), which are similar to epistemic scripts that were used in more social domains (Weinberger et al., 2005), turned out to be less effective than self-explanation for heuristic worked examples (Hilbert, et al., 2008).

Much of previous research on the effectiveness of worked examples has been conducted in well-structured domains (e.g., algebra: Carroll, 1994; probability: Atkinson et al., 2003; geometry: Schwonke et al., 2009). Yet, when problems are less well-defined, like developing a medical diagnosis (Stark, Kopp, & Fischer, 2011) or constructing a mathematical proof, it is usually impossible to provide learners with a procedure that leads to a successful solution directly. Nevertheless, at least heuristics for choosing adequate principles to solve the task are usually available. Mathematical argumentation is considered an ill-defined domain that requires a whole set of different heuristics (Koedinger, 1998). Presenting a step-by-step solution does not address these heuristics.

One possibility to provide effective support in ill-structured domains is through *modelling examples*. Modelling examples demonstrate a typical problem solving behaviour or solution process (van Merrienboer, 1997; cf. Chinnappan & Lawson, 1996). Banduras social theory of learning (1977) assumes that by observing, learners build up a mental representation (schema) of the models' actions and imitates these actions in later problem solving situations. Thus, modelling examples do not convey schemata *to solve a particular problem*, but model the actions and strategies used *to find a solution.* Apart from highly structured mathematics tasks, research on modelling examples has also studied less structured tasks like writing, assertive communication, and also meta-cognitive skills (for an overview see van Gog & Rummel, 2010, cf. Schoenfeld, 1985 for mathematics).

In an effort to combine elements of worked examples and modelling examples, Reiss and Renkl (2002) proposed *heuristic worked examples*, which should be effective in rather ill-structured domains. Following Renkl, Hilbert, and Schworm (2009), heuristic worked examples present not only a solution to a concrete problem (exemplifying domain: e.g., how to prove a certain statement), but also domain-specific principles (learning domain: e.g., what constitutes a mathematical proof) and strategies that might be used to solve problems of a similar type (strategy level). The presentation of additional information elements on the strategy level however potentially introduces extraneous load (van Gog & Rummel, 2010). Accordingly Renkl et al. (2009) propose to focus learning on one type of information. Heuristic worked examples focus in particular on the strategy level, by explicating heuristic guidelines, e.g. following models of an expert's cognitive processes (Nadolski, Kirschner, & van Merrienboer, 2006). An example of such a model for mathematical argumentation is Boero's (1999) process model of mathematical proof. It distinguishes six phases, which cover explorative as well as systematizing processes, and each phase relates to specific heuristic strategies. Studying heuristic worked examples based on Boero's model in dyads was found to be more effective than studying an instructional text (Hilbert et al., 2008) and traditional classroom instruction (Reiss et al., 2007) to foster 8th graders' MAS.

As a side effect, studying heuristic worked examples may also be a way to facilitate social-discursive MAS, as they present learners also with information on what heuristics lead to impasses in relation to solving a particular problem. This may lead learners to think about the pros and cons of different solution paths, and thus imply an activation of dialectical argumentation strategies they have developed for different contexts but at first sight did not have expected to work in a mathematical context as well, which usually is regarded as a rather well-defined context with clear rules and algorithms. In other words, low level social-discursive argumentation in mathematics may actually be a transfer problem. As Fischer et al. (2013) argue in the context of what they call the *optimal external scripting level principle*, scaffolds that prompt learners to use a strategy they already possess in a new context should be more effective than prompting every single component skill of that strategy, and this might happen when learners process heuristic worked examples that point them to advantages and disadvantages of different solution strategies. Yet, there is no systematic research comparing the use of heuristic worked examples to a condition in which students solve equivalent problems collaboratively.

Recent research has indicated that, to describe collaborative learning, the mechanisms of CLT have to be extended and that results cannot be transferred easily from research in individual settings. Kirschner, Paas, Kirschner and Janssen (2011) postulate a *collaborative working memory effect* meaning that working memory load can be distributed over group members. If this distribution provides more working memory capacity than required for handling the group interaction, additional resources can be used for learning processes. On the other hand, efforts to coordinate group activities may pose additional extraneous load and reduce this positive effect of collaboration (Kirschner, 2009). However, results on collaborative learning from studying worked examples and problem solving are inconsistent. For a highly structured learning task, Kirschner et al. (2011) found problem solving to be more effective than studying worked examples. For ill-structured tasks, Rummel et al. (2009) found collaborative observation of a model for dyad collaboration to be more effective than collaborative problem solving. In particular for very complex learning tasks that are likely to overstrain also a groups' collaborative working memory, we can expect that studying worked examples will be superior to collaborative problem solving.

Another open issue is whether *heuristic* worked examples are effective independently from learners' prior achievement. CLT predicts that learners with high prior achievement will suffer from extraneous load caused by redundant information in the heuristic worked example (expertise reversal effect; Sweller, 2010). Yet, following the Matthew effect argumentation, learners' prior knowledge and achievement would be considered an important resource for integrating new information. This applies particularly to complex text-based scaffolds (Scheiter & Gerjets, 2007) such as heuristic worked examples, as compared to solving the corresponding problems.

1.4 Combining instructional interventions

Since collaboration scripts are powerful means to foster social-discursive aspects of argumentation (e.g., Rummel & Spada, 2005), and heuristic worked examples have been effective in fostering individual mathematical argumentation (Hilbert et al., 2008; Reiss, Heinze, Renkl, & Große, 2008), it seems promising to combine the two instructional approaches to foster both MAS components. Nevertheless, how to successfully combine different scaffolds is a delicate issue. As Tabak (2004) argues, the combination of two scaffolds can lead to a range of different effects which are not all educationally desirable: First, each scaffold might have effects on one of two different learning outcomes without producing combined effects on the other outcome measure, which would be an instance of *differentiated scaffolding*. Second, *synergistic scaffolding* is present when the two scaffolds mutually amplify each other's effects on a joint outcome, that is when a positive interaction effect between them occurs. Third, from a CLT perspective (Sweller, 2010), even a negative effect seems possible, since having to deal with two instructional interventions simultaneously may create extraneous load. In research on collaboration scripts, this phenomenon would be called "*over-scripting*" (see Dillenbourg, 2002). Yet, to explain the effects of two different interventions addressing two related but distinguishable learning outcomes simultaneously, it would be necessary to know how learners allocate their attention and working memory resources to learning each of the two outcomes. It must be noticed that neither research based on CLT nor on the Script Theory of Guidance has systematically dealt with this issue. Since no specific principles or guidelines are available within both theories, we base our study on the assumption that in this case, memory resources are equally distributed on learning processes related to each of the two scaffolds (*equal distribution assumption*).

2. The Present Study

We were interested in which of the aforementioned three effects (differentiated scaffolding vs. synergistic scaffolding vs. cognitive overload/over-scripting) would occur when mathematics teacher students with different levels of prior achievement are provided with a combination of collaboration scripts and heuristic worked examples to acquire MAS in a CSCL environment on mathematical proof problems. Our main questions were:

1. *Differentiated scaffolding effects:* What are the effects of a collaboration script vs. unstructured collaboration on the acquisition of the social-discursive component of MAS (RQ1a)? What are the effects of providing small groups with heuristic worked examples vs. problem-solving tasks on the acquisition of the individual component of MAS (RQ1b)?

In line with the external script guidance principle (Fischer et al., 2013) and prior research (e.g., Rummel & Spada, 2005), we expected positive effects of the collaboration script compared to unstructured collaboration on the acquisition of skills related to the social-discursive component (*hypothesis 1a*). The reason for this is that collaboration scripts help students perform skills on a higher level than they would be able to do without, which should in turn facilitate the acquisition of the corresponding skills. Further, worked examples have proven to reduce extraneous load and foster learning of domain-specific skills (Sweller, 2010) when compared to problem solving. Following the equal distribution assumption, learners should allocate significant portions of their cognitive resources not only to learning social-discursive aspects of MAS, but also to schema construction with respect to individual components of MAS. Thus, we expected a positive effect of heuristic worked examples compared to problem solving on the individual component of MAS (*hypothesis 1b*).

2. *Crossover effects:* What are the effects of studying heuristic worked examples vs. solving corresponding problems on the acquisition of the social-discursive component of MAS (RQ2a)? What are the effects of a collaboration script vs. unstructured collaboration on the acquisition of the individual component of MAS (RQ2b)?

Based on the Script Theory of Guidance (Fischer et al., 2013), it can be assumed that low argumentative performance will in some cases be due to a transfer problem: students may already possess an adequate argumentation strategy but do not see that they can apply it in a context that does not appear to be one in which this strategy would work. Thus, as heuristic worked examples provide collaborating learners with arguments for a certain strategy to solve a proof problem, but also with impasses, this might signal to learners that a dialectical argumentation strategy could be used also in this context. If a heuristic worked example is perceived this way, and in case the students have a dialectical argumentation strategy in their repertoire, positive effects of the heuristic worked example on the acquisition of social-discursive MAS may be expected (*hypothesis 2a*). Following the equal distribution assumption, also in our context with two competing outcomes a significant part of the memory resources provided by heuristic worked examples should be used to indeed activate these inactive strategies and adapt them according to the new context.s Further, collaboration scripts stimulate a more thorough argumentative utilization and elaboration of domain-specific information than unstructured collaboration, which should lead to higher levels of individual MAS. Likewise, in terms of CLT, collaboration scripts should reduce cognitive load caused by group coordination and increase germane load by stimulating content-related argumentation. We therefore expected positive effects of the script on individual MAS (*hypothesis 2b*). However, since only few previous studies found positive effects of collaboration scripts on domain-specific learning outcomes (e.g., Weinberger et al., 2005), we expected these effects not to be very pronounced.

3. *Synergistic scaffolding and cognitive overload/over-scripting effects:* What are the effects of studying heuristic worked examples vs. solving corresponding problems on the effectiveness of collaboration scripts with respect to the acquisition of the social-discursive component of MAS (RQ3a)? What are the effects of a collaboration script

vs. unstructured collaboration on the effectiveness of worked examples with respect to the acquisition of the individual component of MAS (RQ3b)?

Collaboration scripts have proved to be effective in fostering social-discursive skills in the past (e.g., Rummel & Spada, 2005). Following the equal distribution assumption, we expect learners to allocate a significant part of the memory resources provided by the heuristic worked example to process the collaboration scripts. Thus, we expect that when a collaboration script is combined to a heuristic worked example, this should lead to a larger positive effect of the collaboration script in the heuristic worked example condition compared to the problem solving condition (*hypothesis 3a*).

As argued above, collaboration scripts did not consistently prove to foster domain-specific learning in previous research. Vogel et al. (2012) found (sparse) positive effects especially when collaboration scripts were combined with domain-specific support. We expected that, compared to unstructured collaboration, providing learners with a collaboration script will positively affect the effectiveness of worked examples, since the script should free learners from efforts of group coordination. We expect that students use these additional resources at least partly to elaborate the contents of the heuristic worked examples more deeply (equal distribution assumption). Thus, also with respect to the individual-mathematical component, we expected students from the combined condition to outperform students from all other conditions (*hypothesis 3b*).

Regarding the role of prior achievement, note that heuristic worked examples and collaboration scripts contain information that does not exclusively refer to knowledge that is connected to the specific components of MAS studied here, but includes general, typically text-based information on mathematical heuristics and collaboration strategies. Thus, we assume that general prior achievement is a more valid indicator of students' success in extracting information from the learning environment as compared to pre-test achievement for the components of MAS.

4. *Role of prior achievement:* In what way do heuristic worked examples (compared to unstructured problem solving), collaboration scripts (compared to unstructured collaboration) and their combination have differential effects on the social-discursive (RQ4a) and the individual component (RQ4b) of MAS of mathematics teacher students depending on their prior achievement (as measured by high school GPA)?

Prior research offers reasons to assume that text-based scaffolds such as heuristic worked examples and collaboration scripts will especially support learners with higher levels of prior achievement due to their higher chance of extracting new information from the text and integrating it with their more extensive prior knowledge base (Collins & Loftus, 1975; Kintsch, 1998). Other research however indicates that more text-based instruction will harm the learning of high achievers due to the fact that instructional information is redundant for them (Kalyuga et al., 2012). Thus we did not set up a directed hypothesis with respect to this question.

To explore the mechanisms by which the two scaffolds influenced the two MAS components, we further checked the effects of the two scaffolds on three process indicators, namely overall cognitive effort, extraneous cognitive load during learning and the frequency of elaborations during collaboration. We expected that both the collaboration script and the heuristic worked examples would decrease extraneous cognitive load and increase the frequency of elaborations related to the learning material.

3. Method

3.1 Participants and design

After excluding 61 students who missed one or more learning or test sessions, a total of $N = 101$ beginning mathematics teacher students ($M_{Age} = 20.04$; $SD = 2.41$; 57 female, 44 male) were included in the analyses. Participants were divided into two groups according to prior achievement (high vs. low) based on a median-split of their overall high-school GPAs. The German averaged school grades were transformed to values from 0 (insufficient) to 5 (excellent) and rounded to the first decimal. The highest reported GPA was 5.0, the lowest was 2.5 (M = 3.91, SD = 0.58). Within each prior achievement group, participants were then randomly assigned to the four experimental conditions of a 2x2 factorial design with the factors (1) *collaboration script* (N_{with} = 48 vs. $N_{without}$ = 53) and (2) *heuristic worked examples* (N_{with} = 53 vs. $N_{without}$ = 48; the latter employing unguided problem-solving). $N_{combined}$ = 26 students were assigned to the combined condition with collaboration script and heuristic worked example. For each treatment session, each student was assigned to a new learning partner with a comparable level of prior achievement within each experimental condition. The reason for establishing homogeneous dyads was to reduce potential noise in the data that is produced if some students would collaborate with peers that were comparable to them while others would form dyads with students with considerably higher or lower GPA, which would later on be difficult to partial out (see, e.g., Webb, Nemer, & Zuniga, 2002). Although for the identification of prior achievement a median split procedure was used, for analyzing the effects of prior achievement, GPA was treated as a continuous variable.

3.2 Instructional setting and procedure

The experiment was embedded in a voluntary preparatory course for university students who started a mathematics teacher education program right after the course. The course was run at two universities in Germany, with all four conditions of the experimental design being realized at both places. During the first three days of the course, students participated in lectures and seminars about elementary number theory and other mathematical content. On the fourth day, the pre-tests for the experiment took place (see below). The experiment started on day 5 with a video introduction to the CSCL environment, followed by a first 45-

minutes treatment session. Two additional 45-minutes treatment sessions took place on the sixth and seventh day. On the eighth and ninth day, students completed the post-tests.

3.3 Learning environment

In the CSCL environment, dyads worked on mathematical argumentation problems (e.g., “Take five consecutive numbers and add them up. Repeat this and try to find regularities. Formulate a conjecture and prove it!”). The screen was divided into two areas. While the left side presented individual information to each participant, the right side featured a graphical chat area that could be used by both learners (see Fig. 1).

INSERT FIGURE 1 ABOUT HERE

The left side of the screen presented the problem formulation and (in the worked example condition) single steps of the heuristic worked example. Extra tools (e.g., calculator) were available on the bottom for all conditions. On top of the right side of the screen, a set of prompts implemented a common basic structure for the learning process for all four experimental conditions: Students were initially asked to think about the problem or the heuristic worked example individually. After that, one of them was requested to explain his or her ideas about the problem resp. the heuristic worked example (depending on the experimental condition). Then, the learners discussed their ideas before they went on with problem-solving resp. the heuristic worked example individually and repeated the same process. In the collaboration script conditions, these prompts were more elaborate and implemented the different script components. Buttons at the right bottom of the screen were activated depending on the current phase of the collaboration script to finish the phase by pressing the button. In the condition without collaboration script the learning environment did not show the buttons at the bottom but only one unspecific “finish” button to progress to next individual or collaborative learning phase. The middle part of the right side functioned as a communication tool (chat window with graphic and typing function) and was available for all conditions.

3.4 Independent variables

Collaboration script. In all collaborative learning phases, students provided with the collaboration script received prompts (e.g., “Please, formulate an argument supporting your position and share it with your learning partner.” or “Please listen critically to the argumentation of your learning partner.”) that sequenced their discussion in three phases of argumentation (see Leitao, 2000): “constructing a pro argument”, “answering with a counter argument”, “building a consensus”. Additionally, these students had access to a written explanation about the structure of good arguments (Toulmin, 1958; e.g., “To formulate your argument, first formulate a claim.”, or “Justify your claim with appropriate data.”, etc.). Participants without collaboration script did not receive prompts to structure their collaboration. Though, prior to the collaborative learning phase participants in both conditions watched a video that included the same information as the prompts that were later used in the collaboration script condition.

Heuristic worked example. The materials implemented on the left screen side varied according to whether a heuristic worked example of a mathematical argumentation problem was presented, or the same problem was posed to work on in the dyad. Three problems from elementary number theory were presented in the three treatment sessions (e.g., “Choose an odd amount of consecutive numbers, e.g. 3, 5 or 7 consecutive numbers. Sum up these consecutive numbers. Do you notice anything special? Find a conjecture and prove it!”). The heuristic worked examples described how a fictitious peer student proved a conjecture following six phases adapted from Boero’s (1999) process model of mathematical proof. They contained information from the exemplifying domain (mathematical theorems and operations applied by the student), the learning domain (principles of mathematical proof), and the strategy level (heuristic strategies applied by the students; Renkl et al., 2009). Self-explanation prompts focused only on the strategy level (Renkl et al., 2009) and were provided in every phase of the heuristic worked examples, asking students to reflect individually (e.g., "Why does the protagonist choose this approach?") and (in every second phase) with their partner why the peer in the heuristic worked example might have chosen this approach (e.g., after explaining the different ideas of the two worked examples "Please discuss advantages and disadvantages of the two approaches with your learning partner"). . To increase the need for discussion, the partners received different heuristic worked examples for the same problem that varied in the heuristic strategies applied. In the problem solving condition, without heuristic worked examples, dyads received the problem formulation and were instructed to find a solution together.

3.5 Dependent Variables

Social-discursive component of MAS. To measure the social-discursive component, we assessed students' acquisition of knowledge about the sequence of an argumentation process with a test that asked them to imagine participating in a discussion about physics phenomena (e.g., a discussion about the physical properties of light) and to describe up to five typical phases of such discussions, along with quality criteria to be taken into account for an optimal argumentation. The test was used both as a pre- and a post-test. Answers were coded for the appearance of elements included in the collaboration script (pro-argumentation, counter-argumentation, consensus building, and response to arguments, see Table 1 exemplifying an answer that was rated with a high value of the social-discursive component of MAS; answers that were rated with low values were for example "*Framing the topic in general. The frame of the discussion must be made clear.*", "*Examining the topic at the surface level. Finding different examples which are related to the topic.*", or "*Talking about one example more* *explicitly.*”). To check for inter-rater reliability, two trained raters independently coded 30 randomly selected answers from pre- and post-test (ca. 12 % of all coded answers). Cohen's Kappa for each single code reached good levels, on average ($M_{Cohen's\ \kappa}$ =.82; $\kappa_{minimum}$ = .76, $\kappa_{maximum}$ = .93). Students received ratings between 0 and 4 for naming relevant elements. On average, they reached values of M = 1.59 (SD = 1.27) in the pre-test and M = 2.22 (SD = 1.24) in the post-test.

INSERT TABLE 1 ABOUT HERE

In order to test for the validity of this measure, we asked students to work collaboratively on another mathematical proof task without any instructional support after the treatment phases. To analyze the students' collaboration process on the proof task we rated each turn if the students engaged in social-discursive argumentation by critiquing and integrating the contributions of their respective collaboration partners or not (interrater reliability regarding the identification of social discursive critique and integrations was sufficient, ICC = .68). A bivariate correlation between students' performance in the post-test for the social-discursive component of MAS and their individual engagement in social-discursive argumentation while solving a proof task collaboratively and unsupported integrations was significant (r = .22; p < .05), indicating that students who were good at naming elements of good social-discursive argumentation practice were also more likely to use these elements during their collaborative work on a mathematical proof task, compared to students who had a low post-test performance. We interpret this finding as evidence supporting the explanatory power of the knowledge test we used.

Individual component of MAS. Parallel pre- and post-tests with 17 open items each measured the individual component of MAS in three test parts, based on a model of mathematical proof skills (Heinze, Reiss, & Rudolph, 2005). Five items focused on schematic argumentation with elementary rules from number theory (e.g., “Show that for all natural numbers, a and b, the following statement is true: If 7 divides a+3b then 7 divides 2a+13b.”), which required transformations of the algebraic expression and application of rules from the courses’ number theory lectures. Proof skills in elementary number theory were examined by six items (e.g., “Prove the following statement: The sum of five consecutive numbers is divisible by five.”). Six items tested performance in open-ended argumentation problems (e.g., “Prove or refute the following statement for natural numbers a and b: If you multiply the sum of a and b with the difference of a and b, you will always obtain an even number.”). In addition, students had to explore and evaluate the conjectures and decide whether they were true or false. The coding procedure for all test parts was adapted from Heinze et al. (2005): 0.5 points were awarded for incomplete but partially correct solutions of an item, and one point if the item was solved correctly. Two trained, independent raters coded all items. Inter-rater reliability was good (Mean of $ICC_{unjust} = .79$). Where discrepancies remained, raters discussed them until they reached a consensus. Reliability was good for both tests (Cronbach’s alpha: $\alpha = .82$ for the pre-test, $\alpha = .80$ for the post-test). For all analyses, the mean scores for the three test parts (schematic argumentation, open-ended argumentation, and conjecturing) were summed up. Thus, the maximum score was 3 for pre- and post-test.

Low to moderate correlations between prior achievement and pre-test scores for each components of MAS (social-discursive: $r = .27$ $p < .01$, individual: $r = .51$; $p < .001$) indicate that general prior achievement can be differentiated from prior MAS skills.

3.6 Explorative process analyses

To judge what mechanisms contributed to learning concerning both MAS components during collaboration, we investigated the effects of the two scaffolds on three process indicators. First, to compare students’ general engagement in the learning environment, we used a measure of *overall cognitive effort* based on a ten point "thermometer" scale (values 1 (low effort) to 10 (high effort); adapted from OECD, 2007, p. 52f). Second, we used a measure of *extraneous cognitive load* proposed by Opfermann (2008) that asked learners to indicate the mental effort they had to invest while learning (e.g., "How hard or easy was it to identify all necessary information while studying the example/solving the problem?") on a nine-point scale (from "very easy" to "very hard"). Students had to indicate their effort retrospectively at the end of each treatment session. This measure is considered a good compromise between valid assessment of cognitive load and non-intrusive application during the learning phase (Opfermann, 2008). Third, to describe how intensively students' were actively engaged in learning, we used the *frequency of on-task utterances* during the collaborative learning process. Those included utterances in which students elaborated learning content or instructional information during collaboration. For economical and technical reasons, only the learners' chat utterances were included (that is, oral discussions which were possible due to the co-present learning setting were not analyzed). We trained two coders to independently code students' on-task utterances in the collaborative learning process while working on the proof tasks in the treatment phases. After training the coders reached a good level of interrater reliability based on a sample of 22 dyads ($ICC = 0.80$). The remaining process data were then coded by one of the two coders.

3.7 Statistical analyses

Hypotheses to RQs 1, 2 and 3 were tested using univariate analyses of covariance with collaboration script and heuristic worked examples as fixed factors, the post-test measures of social-discursive resp. individual MAS as dependent variable. To model learning gain, corresponding pre-test measures were entered as covariates. To investigate RQ 4, prior achievement was added as a further covariate, and regression analyses were run to estimate differential effects for learners with differing levels of prior achievement. The explorative process analyses were based on ANOVAs with the two scaffolds as independent factors and cognitive load resp. on-task utterances as dependent variables. For all analyses, the significance level was set to .05. As an effect size measure, partial η^2 was used, classifying values between .01 and .05 as weak effects, values above .06 as medium effects, and values of .14 or higher as large effects (Cohen, 1988).

4. Results

Pre-test performance on individual resp. social-discursive components of MAS had a rather low and insignificant correlation ($r = .13$, $p = .19$). Thus, we treated them separately in our analyses.

4.1 Social-discursive component of mathematical argumentation skills

Descriptively, the social-discursive component increased from pre- ($M = 1.59$, $SD = 1.27$) to post-test ($M = 2.22$, $SD = 1.24$) across conditions, $F(1,100) = 10.76$, $p < .01$, part. $\eta^2 = .10$. Pre-test differences just failed to reach significance, $F(3,97) = 1.96$, $p = .13$, part. $\eta^2 = .06$, suggesting to control for pre-test scores in subsequent analyses.

INSERT TABLE 2 ABOUT HERE

Effects of collaboration scripts and heuristic worked examples: To test our hypotheses, we first conducted a 2 X 2 factorial ANCOVA with the post-test measures of the social-discursive component as dependent variable, collaboration script and heuristic worked examples as independent variables, and pre-test performance on the social-discursive component as covariate (a significant correlation between the pre-test and post-test measures, $r = .41$, $p < .01$, substantiated to control for pre-test measures). Confirming hypothesis 1a, the collaboration script led to significantly higher gains than unstructured collaboration, $F(1,96)$

= 4.42, p = .04, part. η^2 = .04 (see Table 2). Also, in accordance with hypothesis 2a, we found a significant positive effect of heuristic worked examples compared to problem solving on post-test achievement, $F(1,96) = 9.68$, $p < .01$, part. η^2 = .09. Concerning hypothesis 3a, we found no significant interaction effect between the two interventions, $F(1,96) = 0.03$, p = .86, part. $\eta^2 < .01$. Rather, the effects of the two scaffolds on the social-discursive MAS added up.

Influence of prior achievement: To answer research question 4a that asked for possible differential effects of the treatments for learners with differences in prior achievement, we included learners' prior achievement (high school GPA, z-standardized) in the ANCOVA model (post-test performance on the social-discursive component of MAS correlated significantly with prior achievement, r = .27, $p < .01$). After that, learning with collaboration script vs. unstructured collaboration still had a positive effect on the acquisition of students' social-discursive MAS, $F(1,93) = 5.81$, p = .02, part. η^2 = .06. However, this result was qualified by a marginally significant interaction effect between collaboration script and prior achievement, $F(1,93) = 3.65$, p = .06, part. η^2 = .04. A comparison of the single regression lines of high-school GPA on the learning gain for both conditions revealed that in the unscripted condition, the regression slope was almost zero ($B < 0.01$, β = .002, p = .99), while it was positive (B = 0.56, β = .47, $p < .01$) for the scripted condition. This indicates that learning gain was dependent on prior achievement in the collaboration script condition, but not in the unscripted condition. Figure 2 indicates that learners with average and above average prior achievement showed a higher learning gain in the condition with collaboration script than in the condition without, while for learners with below average prior achievement the difference in the learning gain measure between the condition with collaboration script and the unscripted condition showed no consistent pattern.

Similar findings occurred for heuristic worked examples, when prior achievement was used as a covariate. Again, we found a significant main effect of heuristic worked examples, $F(1,93) = 10.82$, $p < .01$, part. η^2 = .10. However, between heuristic worked examples and prior achievement, a significant interaction occurred, $F(1,93) = 5.40$, $p = .02$, part. $\eta^2 = .06$. The comparison between the single regression lines within both experimental conditions (see fig. 3) revealed that the regression slope for students in the condition with problem solving was positive ($B = 0.51$, $\beta = .42$, $p < .01$) while it was almost zero ($B = 0.03$, $\beta = .03$, $p = .84$) for students in the condition with heuristic worked examples. This indicates that social-discursive learning gain was dependent on prior achievement in the problem solving condition, but not in the heuristic worked examples condition. Figure 3 indicates that learning with heuristic worked examples supported the acquisition of the social discursive component of MAS almost equally for all learners, while problem solving impaired it for students with low prior achievement but not for students with high prior achievement. Finally, after adding prior achievement as a covariate, we still found no significant interaction between both scaffolds, $F(1,93) = 0.01$, $p = .93$, part. $\eta^2 < .01$.

INSERT FIGURE 2 ABOUT HERE

INSERT FIGURE 3 ABOUT HERE

4.2 Individual component of mathematical argumentation skills

Table 3 shows the mean pre- and post-test scores for the test measuring individual MAS for all experimental conditions. Pre-test score differences did not reach significance, $F(3,97) = 1.34$, $p = .27$, part. $\eta^2 = .04$, but pre-test scores were controlled in subsequent analyses.

INSERT TABLE 3 ABOUT HERE

Effects of heuristic worked examples and collaboration scripts: We first conducted a 2 X 2 factorial ANCOVA with the post-test measures of the individual component of MAS as dependent variable, collaboration script and heuristic worked examples as independent variables, and pre-test performance on the individual component of MAS as a covariate (pre- and post-test performance were significantly correlated, r = .79, p < 0.01). No significant main effects of the *heuristic worked example,* F(1,96) = 0.39, p =. 53, part. η^2 < .01, and *collaboration script,* F(1,96) = 0.19, p = .67, part. η^2 < .01 occurred, and also no significant interaction, F(1,96) = 0.02, p = .88, part. η^2 < .01. Thus, hypotheses 1b, 2b and 3b were not confirmed by these first analyses.

Influence of prior achievement: Research question 4b concerned differential effects of the two treatments on the individual-mathematical component for learners with different levels of prior achievement. Therefore, we included learners' prior achievement (z-standardized high school GPA) in the ANCOVA model (post-test performance on the individual-mathematical component of argumentation skills correlated significantly with prior achievement, r = .50, p < .01). The interaction effect between *prior achievement* and *heuristic worked example* was significant, F(1,93) = 5.23, p = .02, part. η^2 = .05. To study the interaction effect in detail, separate regression analyses were calculated in both conditions. The regression slope between z-standardized high school GPA and learning gain on the individual-mathematical component was positive in the heuristic worked example condition (B = .13, β = 0.37, p < .01) and almost zero in the problem solving condition (B = -.03 , β = -.09, p = .54). This indicates that the acquisition of individual MAS depended significantly on high prior achievement in the heuristic worked example condition, but not in the problem solving condition. Figure 4 indicates that learning from a heuristic worked example fostered individual MAS of students with high prior achievement more than solving equivalent problems. Yet, heuristic worked examples turned out to be less effective than problem solving for students with low prior achievement. No significant interaction of prior achievement and collaboration script was observed $F(1,93) = 1.48$, $p = .23$, part. $\eta^2 = .02$. Taking also the missing main effect of the collaboration script on individual MAS into account, we found no indication that the collaboration script affected learners with lower nor with higher prior achievement with respect to the acquisition of individual MAS. The main and interaction effects of the collaboration script and the heuristic worked example remained unchanged when including prior achievement.

INSERT FIGURE 4 ABOUT HERE

4.3 Process analyses

Mental effort: The outcomes of our interventions were partly in contrast to our hypotheses and in general more complex than anticipated. To explain the underlying processes, we considered additional data from the treatment sessions. Firstly, these unexpected results might have been influenced by the extent to which the different learning environments evoked students' overall learning effort. Reported mean effort varied between 3 and 10 ($M = 6.61$; $SD = 1.40$). A 2 X 2 factorial ANCOVA with the two scaffolds as independent variables, prior achievement as covariate and the mean overall effort rating over three intervention sessions as dependent variable revealed no significant effects of the two scaffolds (heuristic worked example: $F(1,97) = 1.82$; $p = 0.18$; part. $\eta^2 = .02$, collaboration script: $F(1,97) = 0.23$; $p = 0.64$; part. $\eta^2 = .002$, interaction effect: $F(1,97) = 1.87$; $p = 0.18$; part. $\eta^2 = .02$).

Cognitive load: From the theoretical perspective of CLT, the main reason for the effectiveness of (heuristic) worked examples should lie in the reduction of extraneous load posed by the learning environment. We performed a 2 X 2 ANOVA for the participants' mean ratings of extraneous load with the two scaffolds as independent variables. Participants reported significantly less extraneous load when learning with heuristic examples (*M* = 3.59, *SD* = 1.20) compared to the problem solving condition (*M* = 4.18, SD = 1.02), $F(1,97) = 7.04$, $p < .01$; part. $\eta^2 = .07$. Neither a significant main effect of the collaboration script, $F(1,97) = 0.19$; $p = 0.66$; part. $\eta^2 = .002$, nor a significant interaction effect between script support and heuristic worked example, $F(1,97) = 2.22$; $p = 0.14$; part. $\eta^2 = .02$ occurred.

On-task utterances: When learning collaboratively – from problem solving or from worked examples – content-related discourse is regarded a main driver for individual learning. To study how the heuristic worked examples, the collaboration script and their combination influenced on-task utterances during the collaborative treatment phase, a 2 X 2 factorial ANCOVA with the two scaffolds as independent variables, the number of on-task utterances in the chat as dependent variable and prior achievement as covariate was conducted. Prior achievement had a medium positive effect on student's on-task utterances, $F(1,96) = 6.50$, $p = .01$, part. $\eta^2 = .06$, while there was no significant interaction between prior achievement and the two scaffolds. The results of the ANCOVA showed a weak positive main effect of learning with heuristic worked examples (*M* = 9.00, *SD* = 5.45) compared to problem solving (*M* = 7.08, *SD* = 3.56), $F(1,96) = 5.06$, $p = .03$, part. $\eta^2 = .05$. Also students learning with the collaboration script elaborated significantly more on the learning material (*M* = 9.57, *SD* = 5.33) compared to the learners that learned without the collaboration script (*M* = 6.75, *SD* = 3.68) with a medium effect size, $F(1,96) = 10.09$, $p < .01$, part. $\eta^2 = .10$. No significant interaction effect between the heuristic worked examples and the collaboration script occurred, $F(1,96) < 1$, $p = .86$, part. $\eta^2 <$ 01.

5. Discussion

In this study we investigated whether combining collaboration scripts and heuristic worked examples in a CSCL environment is an effective way to facilitate mathematics teacher students' acquisition of two components of MAS – a social-discursive component that refers to the ability to participate in a constructive collaborative argumentation process, and an individual-mathematical component necessary to generate arguments for or against a mathematical conjecture. The Script Theory of Guidance and CLT predicted (a) *differentiated scaffolding effects* (Tabak, 2004), meaning that the collaboration script would foster the social-discursive component and the heuristic worked examples would foster the individual-mathematical component, (b) *crossover effects*, which means that each scaffold would have positive effects on the component of MAS it was not primarily designed for, and (c) *synergistic scaffolding effects* (Tabak, 2004), which would occur when each scaffold amplifies the effectiveness of the other with respect to both MAS components, although also a negative interaction in terms of an cognitive overload or over-scripting effect would be conceivable. Furthermore, we were interested in the question whether the effectiveness of the two scaffolding approaches was dependent on the learners' prior achievement. The discussion is organized along these four aspects.

5.1 Differentiated scaffolding

Positive effects of collaboration scripts on social-discursive argumentation skills (e.g., Rummel & Spada, 2005) and of heuristic worked examples on individual-mathematical components (Hilbert et al., 2008) are well documented in the literature. In line with this and supporting hypothesis 1a, we found that the collaboration script supported the acquisition of the social-discursive component of MAS better than unstructured collaboration, which extends prior research on the positive effects of collaboration scripts on argumentation to a new domain, namely mathematics (e.g., Kollar et al., 2007; Rummel & Spada, 2005; Stegmann et al., 2007) and provides further support for the external script guidance principle that was proposed in the Script Theory of Guidance (Fischer et al., 2013). Thus, even in domains like mathematics, which learners might not regard as typical for social-discursive aspects of argumentation, it is possible to foster the development of social-discursive skills. Even more, the analyses of the collaborative learning process seem to indicate that this is due to an increase in on-task utterances that is caused by the collaboration script.

In contrast to hypothesis 1b, no significant advantage of studying heuristic worked examples over solving corresponding problems could be found for the individual component of MAS. This result is in contrast to evidence from previous studies that demonstrate the effectiveness of heuristic worked examples in fostering domain-specific skills (Reiss et al., 2007) and the effectiveness of worked examples in general (van Gog & Rummel, 2010). This is astonishing, since our process data indicate that the heuristic worked examples did indeed reduce extraneous load and increase on-task utterances. In the given learning context with two competing learning outcomes – social-discursive and individual MAS – the reason for this unexpected result might be that our underlying "equal distribution assumption" is false. In other words, our results seem to indicate that students did *not* allocate their available working memory resources to both outcomes equally. Kahneman (1973) hypothesized that memory allocation follows a set of so called memory allocation guidelines depending on the circumstances. If resources are not sufficient, storage and processing of newer information should for example have priority over older information, and individuals should allocate more resources to the task that is perceived as requiring more resources or the task that is perceived as more important. In line with this, Foos (1995) found that, when one of two simultaneous memory-intensive tasks (calculation and memorization in Foos, 1995) was presented as the more important one, individuals allocated memory resources primarily to that task. In our case, the collaborative intervention sessions were announced as opportunities to train collaborative learning, whereas domain-specific information was covered during the whole two-week preparatory course. Since this was also explicitly stated before each session, we may assume that participants perceived learning to collaborate as the major aim of these specific sessions. Following Kahneman's (1973) memory allocation hypothesis, students were then likely to allocate those cognitive resources that were set free by the heuristic worked example to schema acquisition for the social-discursive component of MAS, and less for the individual component.

Regarding the effects of problem solving and studying worked examples in collaborative settings, this result does not clarify the inconsistent findings summarized by van Gog and Rummel (2010). Nevertheless, our special context with two learning outcomes, which are treated simultaneously, renders a direct comparison with the results of studies like that of Kirschner (2009) problematic. Moreover, each learner in each dyad received one complete worked example (though, slightly different ones for each learner) in our study, while in Kirschner's study parts of one example were distributed over both learners. The positive effect of heuristic worked examples on social-discursive MAS, nevertheless, indicates that studying heuristic worked examples indeed provided more free memory resources than the distributed working memory effect while solving the problems collaboratively. Given the different domains in the studies of Kirschner (2009, heredity), Rummel et al. (2009, collaboration) and our study, more research is necessary to clarify what task features influence the effectiveness of studying (heuristic) worked examples and problem solving in collaborative settings.

5.2 Crossover effects

In addition, the *equal distribution assumption* resp. the optimal external scripting level principle that was formulated in the Script Theory of Guidance (Fischer et al., 2013) predicted positive *crossover effects* of each scaffold on the skill it was not directly addressed at. In other words, we expected a positive effect of studying heuristic worked examples on the social-discursive component and a positive effect of collaboration scripts on the individual-mathematical component. With respect to heuristic worked examples, the theoretical predictions were met, supporting hypothesis 2a. Obviously, the heuristic worked examples provided learners with more opportunities to argue about different approaches to solve a mathematical problem and to reflect on their argumentative discourse than actually having them solve such problems. Evidence for this interpretation comes from our process analyses that showed that studying heuristic worked examples significantly increased the frequency of on-task utterances, when compared to the problem-solving condition. Based on Fischer et al.'s (2013) optimal external scripting level principle, it seems that distributing slightly different heuristic worked examples that also included impasses in the described procedure towards solving the proof problem helped learners to activate a dialectical argumentation strategy that they already possessed (and even were informed about in the instructional video prior to the collaborative learning phase), despite a possibly low expectation that this strategy works in the rather-well structured domain of mathematics. This, in turn, might have increased students' knowledge about social-discursive argumentation. Another interpretation comes from cognitive load theory: Studying heuristic worked examples provided learners with additional free working memory resources, as process data indicates. Regardless of referring to the *equal distribution assumption* or assuming that learners devoted their resources primarily to social-discursive MAS (see 5.1), a part of these resources was indeed available to build up schemata of social-discursive MAS.

In contrast, we did not find evidence that supports the expected crossover effect from collaboration scripts on individual MAS (hypothesis 2b). Theoretically, this finding is surprising since it is often argued that engaging in high-level argumentation is related to learning about the content of argumentation (e.g., Andriessen, Baker, & Suthers, 2003). Nevertheless, previous studies (e.g., Kollar et al., 2007) have shown that it is far from easy to learn a new collaboration strategy and apply it right away to elaborate content information more deeply and that way acquire high levels of domain-specific knowledge. Our process analyses indicate the same even despite of an increase of on-task utterances in the collaboration script condition. Possibly, capitalizing on higher-level collaboration to acquire higher levels of domain-specific knowledge requires longer interventions. Another interpretation is that most learners in our sample already had at least implicit knowledge (that is, knowledge that learners were not able to express during pre-test) of the dialectical strategy imposed by the collaboration script, but that being confronted with a mathematical context hampered the application of this knowledge during collaboration, since it subjectively may not comply with the widely held view of mathematics as a rule-based, well-structured domain that does not leave much room for collaborative discourse. To make collaboration scripts become effective in this domain, the actual challenge is then to help learners understand that general dialectical strategies also work in the domain of mathematics.

5.3 Synergistic scaffolding effects

Regarding *synergistic scaffolding*, we expected – based on the equal distribution assumption – that each of the two scaffolds would increase the effectiveness of the other scaffold, resulting in significant interaction effects between the factors *heuristic worked example* and *collaboration script* for each component of MAS. Yet, none of these synergistic scaffolding effects occurred in our study, neither for the social-discursive nor the individual component of MAS. This underlines how difficult it is to combine different scaffolds in a way that would produce synergistic scaffolding (Tabak, 2004), although the combination of the two scaffolds at least worked better than each scaffold in isolation regarding the acquisition of social-discursive MAS. More positively speaking, our results indicate that the two scaffolds did not interfere with each other (producing an over-scripting/cognitive overload effect), which could have happened since the combination made the whole learning environment more cognitively demanding. The absence of synergistic scaffolding effects indicates that students in the combined condition had to separate their attention to two rather distinct tasks, namely to (a) follow interaction-related prompts provided by the script and (b) thoroughly process content-related and heuristic information presented in the heuristic worked examples, which overstrained learners in our sample. However, the combined condition did not stand out in learners' reported effort and extraneous load. At least the missing synergistic scaffolding effect for individual MAS can be explained, if – as hypothesized above – learners indeed allocated their working memory resources primarily to learning social-discursive skills and disregarded individual-mathematical schema construction as a further important task in the intervention setting. Apart from focusing learners' attention to relevant learning outcomes, nevertheless, it remains an open question under what circumstances synergistic scaffolding effects are likely to appear.

5.4 Dependence of the effects of collaboration scripts and heuristic worked examples on prior achievement

Several times, we found that the effectiveness of the two scaffolds we investigated depended on learners' general *prior achievement*. First, we found that learners with average and above average *prior achievement* showed a higher learning gain concerning social-discursive MAS when they were supported with the collaboration script compared to learners in the unscripted condition, while learners with below average prior achievement showed rather poor learning gains that were comparable across the two conditions. Similarly, the learning gain for individual MAS depended on prior achievement in the heuristic worked example condition, but not in the problem solving condition. While learners with low prior achievement gained less from studying heuristic worked examples than from problem solving, the results were reverse for learners with high prior achievement. Both effects are in line with the Matthew effect and predictions from text comprehension research (Barab et al., 1997; Scheiter & Gerjets, 2007): differences due to prior achievement especially materialized under the circumstances of highly instructed, text-intensive learning conditions, which means that the collaboration script and the heuristic worked examples helped learners with higher prior achievement to integrate new knowledge into their existing knowledge structures more efficiently. Simultaneously, this result runs counter to evidence on the expertise reversal effect (Sweller, 2010), and the learning processes we observed in the scripted vs. unscripted condition seem to imply in deed that the script primarily increased content elaboration (increase of on-task utterances) rather than extraneous load. In contrast to worked example research, we used high school GPA as a very general indicator of prior achievement, while worked example research usually works with more specific pre-test scores. Thus, high general prior achievement seems to be a good indicator of learning in the more text-based environments. High GPA scores are not associated with lower learning outcomes, as we would have expected for high specific prior skill levels (expertise-reversal). Nevertheless, this result demonstrates the need to investigate how collaboration scripts and heuristic worked examples can be designed to also support low prior achievers in their efforts of skill acquisition.

Interestingly though, for heuristic worked examples an opposite pattern was found with respect to the acquisition of social-discursive MAS: There, studying heuristic worked examples supported the acquisition of the social discursive component of MAS almost equally for all learners, while problem solving impaired it for students with low prior achievement but not for students with high prior achievement. In other words, at least with respect to the social-discursive component, students with low prior achievement who had to solve problems on their own were the most disadvantaged group in our sample. This is in line with cognitive load theory, as these students experienced high levels of extraneous load, most likely lack access to well-elaborated schemata to solve the mathematical proof problems and have to rely on strategies that cause a high working memory load (means-end-analysis). Thus, our results are consistent with the line of argumentation behind the Matthew effect and Kintsch’s (1998) CI Model.

5.5 Restrictions and perspectives

Even though evidence for different fields of research can be derived from our study, there are some limitations that have to be taken into account and that should be considered carefully in further research. First, by splitting MAS into two components, and operationalizing them separately by testing each individual separately, we can only provide indirect information about the learners' ability to really develop and (re-)construct mathematical knowledge *within a small group of learners*. The learners' behaviour within authentic collaborative mathematical argumentation situations should be taken into account not only as a process indicator, but also as an outcome measure in future research. Second, from a mathematics education perspective, a deeper analysis of the three subtests of the individual-mathematical component is at least as interesting as a deeper analysis of the argumentation processes within the learning environment (for first results see Reichersdorfer et al. 2012). Third, future studies should assess the social-discursive component of MAS by aid of a more performance-oriented measure in which students engage in a new collaborative argumentation, rather than by having them describe how they would engage in such a situation. Perhaps using a more performance-oriented measure would yield different effects on social-discursive MAS as the ones that were observed with respect to the measure used in this study. Fourth, with respect to the operationalization of the collaboration script, it should be noted that not only the students in the script condition, but also the students in the unscripted condition were initially informed about the structure of high-level argumentation (during the instructional videos). This is in contrast to many previous studies (e.g., Kollar et al. 2007; Stegmann et al., 2007) and may have led to less pronounced differences between the scripted and unscripted conditions. Fifth, some of the effect sizes we observed were rather low, which calls for future research to probe the generalizability of our findings especially to other domains. Also, the interaction effect between prior achievement and collaboration scripts on the social-discursive MAS component only reached marginal significance. To substantiate the interpretation that especially learners with more positive prior achievement are effectively supported by learning with collaboration scripts, more studies are needed that include prior achievement as a control variable in their experimental designs. Finally, it is also noteworthy that students from the control condition (problem-solving without collaboration script) made virtually no progress between pre- and post-test on social-discursive MAS. This seems to indicate how unfamiliar it is for students to apply social-discursive argumentation strategies in the context of mathematics; in other words, the participating students do not seem to have interpreted the learning scenario as one in which social-discursive argumentation skills can be practiced. Also, the high extraneous load during problem-solving may have aggravated this biased interpretation of the scenario.

Without ignoring these limitations, our study shows that several general principles of CLT (Sweller, 2010) and collaboration script (Fischer et al., 2013) research can be transferred to a setting in which two scaffolds address two components of a complex domain (MAS). One main new insight regards the combination of scaffolds to different outcomes within one learning setting. In this case our results indicate that it is crucial to consider how learners allocate working memory resources to these outcomes, since an “equal distribution assumption” is not necessarily adequate to predict students’ resource allocation and learning.

References

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.

Andriessen, J., Baker, M., & Suthers, D. (2003). Argumentation, computer support, and the educational context of confronting cognitions. In J. Andriessen, M. Baker, & D. Suthers (Eds.), *Arguing to learn: Confronting cognitions in Computer-Supported*

Collaborative Learning environments (pp. 1-25). Dordrecht: Kluwer. doi: 10.1007/978-94-017-0781-7_1.

Asterhan, C. S. C., Schwarz, B. B., & Gil, J. (2012). Small-group, computer-mediated argumentation in middle-school classrooms: The effects of gender and different types of online teacher guidance. *British Journal of Educational Psychology, 82*(3), 375-397. doi: 10.1111/j.2044-8279.2011.02030.x

Atkinson, R. K., Catrambone, R., & Merrill, M. M. (2003). Aiding transfer in statistical learning: Examining the use of conceptually-oriented equations and elaborations. *Journal of Educational Psychology, 95,* 762-773. doi: 10.1037/0022-0663.95.4.762

Atkinson, R.K., Derry, S.J., Renkl, A., & Wortham, D. (2000). Learning from examples: Instructional principles from the worked examples research. *Review of Educational Research, 70*(2), 181-214. doi: 10.3102/00346543070002181

Bandura, A. (1977). *Social learning theory.* New York: General Learning Press.

Barab, S. A., Bowdish, B. E., & Lawless, K. A. (1997). Hypermedia navigation: Profiles of hypermedia users. *Educational Technology Research & Development, 45,* 23-41. doi: 10.1007/BF02299727

Bielaczyc, K., Pirolli, P. L., & Brown, A. L. (1995). Training in self-explanation and self-regulation strategies: Investigating the effects of knowledge acquisition activities on problem solving. *Cognition and Instruction, 13*(2), 221-252. doi: 10.1207/s1532690xci1302_3

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).

Blömeke, S., Suhl, U., Kaiser, G., & Dörhmann, M. (2012). Family background, entry selectivity and opportunities to learn: What matters in primary teacher education? An international comparison of fifteen countries. *Teaching and Teacher Education, 28*(1), 44-55. doi: http://dx.doi.org/10.1016/j.tate.2011.08.006

Carroll, W.M. (1994). Using worked examples as an instructional support in the algebra classroom. *Journal of Educational Psychology, 86*(3), 360-367. doi: 10.1037/0022-0663.86.3.360

Chi, M. T. H., de Leeuw, N., Chiu, M., & LaVancher, C. (1994). Eliciting self explanations improves understanding. *Cognitive Science, 18,* 439-477. doi: 10.1207/s15516709cog1803_3

Chinnapan, M., & Lawson, M.J. (1996). The effects of training in the use of executive strategies in geometry problem solving. *Learning and Instruction, 6*(1), 1-17. doi: 10.1016/0959-4752(95)00013-5

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Collins, A., Brown, J.S., & Newman, S.E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 453-494). Hillsdale, NJ: Lawrence Erlbaum Associates.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82,* 407-428. doi: 10.1037/0033-295X.82.6.407

Dillenbourg, P. (2002). Over-scripting CSCL: The risks of blending collaborative learning with instructional design. In P. A. Kirschner (Ed.), *Three worlds of CSCL* (pp. 61-91). Heerlen, The Netherlands: Open University.

Dillenbourg, P., & Jermann, P. (2007). Designing integrative scripts. In F. Fischer, I. Kollar, H. Mandl, & J. Haake (Eds.), *Scripting Computer-Supported Collaborative Learning: Cognitive, computational, and educational perspectives* (pp. 259-288). New York: Springer.

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

Foos, P.W. (1995). Working memory resource allocation by young, middle-aged, and old adults. *Experimental Aging Research: An International Journal Devoted to the Scientific Study of the Aging Process, 21*(3), 239-250. doi: 10.1080/03610739508253983

Gillies, R. M. (2004). The effects of cooperative learning on junior high school students during small group learning. *Learning and Instruction, 14*(2), 197-213. doi: 10.1016/S0959-4752(03)00068-9

Hämäläinen, R., Oksanen, K., & Häkkinen, P. (2008). Designing and analyzing collaboration in a scripted game for vocational education. *Computers in Human Behavior*, *24*(6), 2496-2506. doi: http://dx.doi.org/10.1016/j.chb.2008.03.010

Hattie, J. (2009). *Visible Learning – A Synthesis of Over 800 Meta-Analyses Relating to Achievement.* New York: Routledge.

Heintz, B. (2000). *Die Innenwelt der Mathematik: Zur Kultur und Praxis einer beweisenden Disziplin.* Wien: Springer.

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

Hilbert, T., Renkl, A., Kessler, S., & Reiss, K. (2008). Learning to prove in geometry: Learning from heuristic examples and how it can be supported. *Learning and Instruction, 18*(1), 54-65. doi: 10.1016/j.learninstruc.2006.10.008

Kahneman, D. (1973). *Attention and Effort.* Englewood Cliffs, NJ: Prentice-Hall.

Kalyuga, S. (2012). Cognitive Load Theory: How many load types does it really need? *Educational Psychology Review, 23*(1), 1-19. doi:10.1007/s10648-010-9150-7

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

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

Kintsch, W. (1998). *Comprehension: A Paradigm for Cognition.* Cambridge: Cambridge University Press.

Kirschner, F. C. (2009). *United Brains for Complex Learning. A cognitive-load approach to collaborative learning efficiency.* Doctoral Dissertation. Open University of the Netherlands.

Kirschner, F., Paas, F., Kirschner, P. A., & Janssen, J. (2011). Differential effects of problem-solving demands on individual and collaborative learning outcomes. *Learning and Instruction, 21*(4), 587-599. doi: 10.1016/j.learninstruc.2011.01.001

Koedinger, K. R. (1998). Conjecturing and argumentation in high-school geometry students. In R. Lehrer (Ed.), *Studies in mathematical thinking and learning. Designing learning environments for developing understanding of geometry and space* (pp. 319-347). Mahwah, NJ: Erlbaum.

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., Fischer, F., & Slotta, J. D. (2007). Internal and external scripts in computer-supported collaborative learning. *Learning and Instruction, 17*(6), 708-721. doi:10.1016/j.learninstruc.2007.09.021

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2001). A comprehensive meta-analysis of the predictive validity of the graduate record examinations: Implications for graduate student selection and performance. *Psychological Bulletin 127*(1), 162-181. doi: 10.1037/003-2909.127.1.162

Leitao, S. (2000). The potential of argument in knowledge building. *Human Development*, *43*, 332-360. doi: 10.1159/000022695

McLaren, B. M., Lim, S., Gagnon, F., Yaron, D., & Koedinger, K. R. (2006). Studying the effects of personalized language and worked examples in the context of a web-based intelligent tutor. In M. Ikeda, K. D. Ashley, & T.-W. Chan (Eds.), *Intelligent tutoring systems 2006, lecture notes in computer science 4053* (pp. 318-328). Berlin: Springer.

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, 63(2), 81-97. doi:10.1037/h0043158

Nadolski, R. J., Kirschner, P. A., & van Merrienboer, J. (2006). Process support in learning tasks for acquiring complex cognitive skills in the domain of law. *Learning and Instruction, 16*(3), 266-278. doi:10.1016/j.learninstruc.2006.03.004

National Governors Association Center for Best Practices, Council of Chief State School Officers (2010). Common Core State Standards for Mathematics. Washington D.C: National Governors Association Center for Best Practices, Council of Chief State School Officers.

Noroozi, O., Weinberger, A., Biemans, H. J. A., Mulder, M., & Chizari, M. (2013). Facilitating argumentative knowledge construction through a transactive discussion script in CSCL. *Computers & Education, 61,* 59-76. doi: 10.1016/j.compedu.2012.08.013,

OECD (2007). *PISA 2006. Science Competencies for Tomorrow's World.* Volume 1: Analysis.

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

Paas, F. G. W. C., & van Merriënboer, J. J. G. (1994). Variability of worked examples and transfer of geometrical problem solving skills: A cognitive load approach. *Journal of Educational Psychology*, 86, 122-133. doi: 10.1037/0022-0663.86.1.122

Reichersdorfer, E., Vogel, F., Fischer, F., Kollar, I., Reiss, K., & Ufer, S. (2012). Different collaborative learning settings to foster mathematical argumentation skills. In T.-Y. Tso (Ed.), *Proceedings of the 36th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 3, pp. 345-352). Taipei: PME.

Reiss, K., Heinze, A., Kessler, S., Rudolph-Albert, F., & Renkl, A. (2007). Fostering argumentation and proof competencies in the mathematics classroom. In M. Prenzel (Ed.), *Studies on the educational quality of schools. The final report on the DFG Priority Programme* (pp. 251-264). Münster: Waxmann.

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: http://dx.doi.org/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), 29-35. doi: 10.1007/BF02655690

Renkl, A. (2002). Learning from worked-out examples: Instructional explanations support self-explanations. *Learning and Instruction, 12*(5), 529-556. doi: http://dx.doi.org/10.1016/S0959-4752(01)00030-5

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

Rey, G. D., & Buchwald, F. (2011). The expertise reversal effect: Cognitive load and motivational explanations. *Journal of Experimental Psychology: Applied*, 17, 33-48. doi: 10.1037/a0022243

Rittle-Johnson, B. (2006). Promoting transfer: Effects of self-explanation and direct instruction. *Child Development 77*(1), 1-15. doi: 10.1111/j.1467-8624.2006.00852.x

Rummel, N., & Spada, H. (2005). Learning to collaborate: An instructional approach to promoting collaborative problem solving in computer-mediated settings. *Journal of the Learning Sciences, 14*(2), 201-241. doi:10.1207/s15327809jls1402_2

Rummel, N., Spada, H., & Hauser, S. (2009). Learning to collaborate while being scripted or by observing a model. *International Journal of Computer-Supported Collaborative Learning, 4*(1), 69-92. doi: 10.1007/s11412-008-9054-4

Scheiter, K., & Gerjets, P. (2007). Learner control in hypermedia environments. *Educational Psychology Review, 19 (3)*, 285-307. doi: 10.1007/s10648-007-9046-3

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

Scheuer, O., McLaren, B. M., Weinberger, A., & Niebuhr, S. (2013). Promoting critical, elaborative discussions through a collaboration script and argument diagrams. *Instructional Science*. doi:10.1007/s11251-013-9274-5

Schoenfeld, A. (1985). *Mathematical Problem Solving.* New York: Academic Press.

Schoonenboom, J. (2008). The effect of a script and a structured interface in grounding discussions. *International Journal of Computer-Supported Collaborative Learning*, *3*, 327-341. doi: 10.1007/s11412-008-9042-8

Schwonke, R., Renkl, A., Krieg, C, Wittwer, J., Aleven, V., & Salden, R. J. C. M. (2009). The Worked-example Effect: Not an artefact of lousy control conditions. *Computers in Human Behavior*, 25, 258-266. doi:10.1016/j.chb.2008.12.011

Slavin, R. E. (1996). Research on cooperative learning and achievement: What we know, what we need to know. *Contemporary Educational Psychology, 21,* 43-69. doi:10.1006/ceps.1996.0004

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.

Stark, R. (1999). *Lernen mit Lösungsbeispielen: Einfluss unvollständiger Lösungsbeispiele auf Beispielelaboration, Lernerfolg und Motivation*. Göttingen: Hogrefe.

Stark, R., Kopp, V., & Fischer, M. R. (2011). Cased-based learning with worked examples in complex domains: Two experimental studies in undergraduate medical education. *Learning and Instruction*, *21*(1), 22-33. doi: 10.1016/j.learninstruc.2009.10.001

Stegmann, K., Wecker, C., Weinberger, A., & Fischer, F. (2012). Collaborative argumentation and cognitive elaboration in a computer-supported collaborative learning environment. *Instructional Science, 40*(2), 297-323. doi: 10.1007/s11251-011-9174-5

Stegmann, K., Weinberger, A., & Fischer, F. (2007). Facilitating argumentative knowledge construction with computer-supported collaboration scripts. *International Journal of Computer-Supported Collaborative Learning, 2*(4), 421-447. doi: 10.1007/s11412-007-9028-y

Sweller, J. (2004). Instructional design consequences of an analogy between evolution by natural selection and human cognitive architecture. *Instructional Science 32*, 9-31. doi: 10.1023/B:TRUC.0000021808.72598.4d

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

Tabak, I. (2004). Synergy: A complement to emerging patterns of distributed scaffolding. *The Journal of the Learning Sciences, 13*(3), 305-335. doi: 10.1207/s15327809jls1303_3

Toulmin, S. E. (1958). *The uses of argument.* Cambridge: Cambridge University Press.

van Gog, T., & Rummel, N. (2010). Example-based learning: Integrating cognitive and social-cognitive research perspectives. *Educational Psychology Review, 22*(2), 155-174. doi: 10.1007/s10648-010-9134-7

Van Loon-Hillen, N., Van Gog, T., & Brand-Gruwel, S. (2010). Effects of worked examples in a primary school mathematics curriculum. *Interactive Learning Environments, 18,* 1-11.

Van Merriënboer, J. J. G. (1997). *Training complex cognitive skills: A four-component instructional design model for technical training*. Englewood Cliffs: Educational Technology Publications.

Vogel, F., Kollar, I., & Fischer, F. (2012). Effects of computer-supported collaboration scripts on domain-specific and domain-general learning outcomes: A meta-analysis. In J. van Aalst, K. Thompson, M. J. Jacobson, & P. Reimann (Eds.), *The future of learning – ICLS 2012 Conference Proceedings (Volume 2 – short papers, symposia and abstracts)* (pp. 446-450). International Society of the Learning Sciences.

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.

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

Wecker, C., Kollar, I., Fischer, F., & Prechtl, H. (submitted). *Fostering scientific literacy in inquiry classrooms with continuous and faded computer-supported collaboration scripts.*

Weinberger, A., Ertl, B., Fischer, F., & Mandl, H. (2005). Epistemic and social scripts in computer-supported collaborative learning. *Instructional Science, 33*(3), 1-30. doi: 10.1007/s11251-004-2322-4

Wouters, P., Paas, F., & Van Merriënboer, J. J. G. (2010). Observational learning from animated models: Effects of studying-practicing alternation and illusion of control of transfer. *Instructional Science, 38,* 89-104. doi: 10.1007/s11251-008-9079-0

Yackel, E., & Cobb, P. (1996). Sociomathematical norms, argumentation, and autonomy in mathematics. *Journal for Research in Mathematics Education, 27*(4), 458-477.

Figure captions

Figure 1: Screenshot of the CSCL environment, showing the condition with heuristic worked example and with collaboration script.

Figure 2: Relationship of students' performance in the social-discursive component of MAS and prior achievement in the conditions with vs. without collaboration script (vertical axis: residual of post-test score under control for pre-test score and the heuristic worked example factor).

Figure 3: Relationship of students' performance in the social-discursive component of MAS and prior achievement in the conditions with heuristic worked example vs. problem solving (vertical axis: residual of post-test score under control for pre-test score and the collaboration script factor).

Figure 4: Relationship of students' performance in the individual component of MAS and prior achievement in the conditions with heuristic worked example vs. problem solving (vertical axis: residual of post-test score under control of pre-test score and script support factor).

ELK-Math

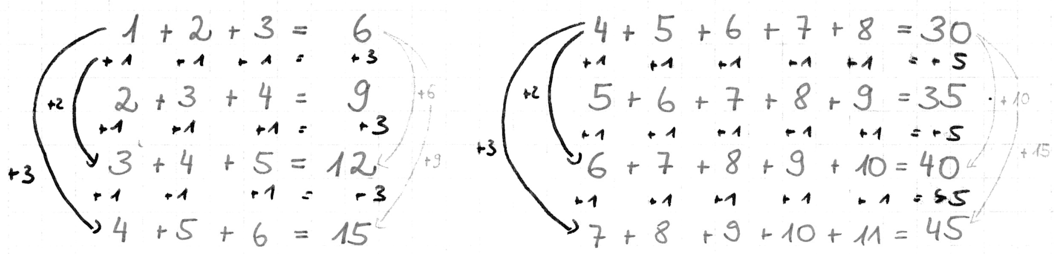
Please listen to Kathi's counterarguments!
If anything is unclear to you, ask her!
You can proceed as soon as Katrin pressed COUNTERARG Finished.

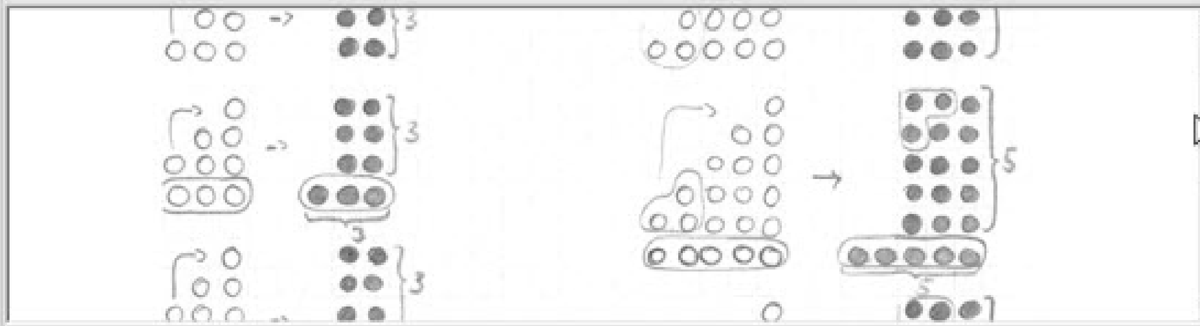
Exploration of the problem situation:

Finn starts to look at some examples:

3 consecutive numbers:	5 consecutive numbers:
$1+2+3=6$	$2+3+4+5+6=20$
$2+3+4=9$	$4+5+6+7+8=30$
$3+4+5=12$	$5+6+7+8+9=35$
$17+18+19=54$	$21+22+23+24+25=115$

Finn: It looks like, as if the result of 5 consecutive numbers is always divisible by 5. Perhaps I can find some structure in the examples:





calculator | lecture notes | Kathis approach | Argumentation

Manus explanation

Manu: the sum of three consecutive numbers is divisible by three,
the sum of five consecutive numbers is divisible by 5.

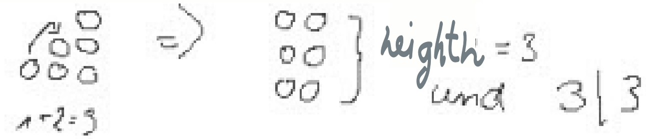
Manus Arguments

Manu: This approach is clearly represented and well structured,
so you can notice remarkable things

Kathis Counterarguments

Kathi: Tims approach is clearly structered and easy to understand. As he demonstrates his calculations with dots, you can arrange those to rectangles of which either the height or the width is divisible by the number of summands.
Kathi: see

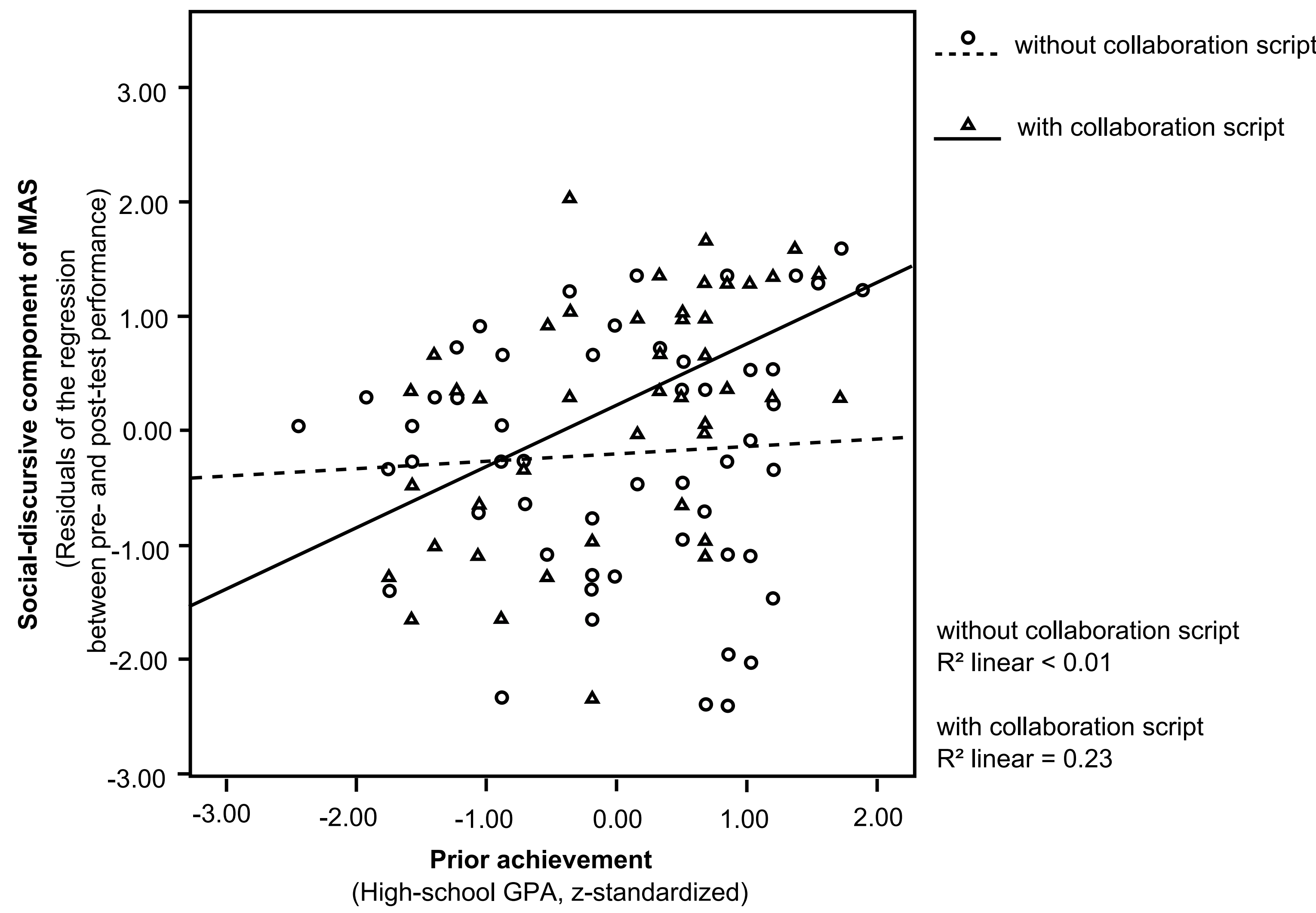
$1+2=3$



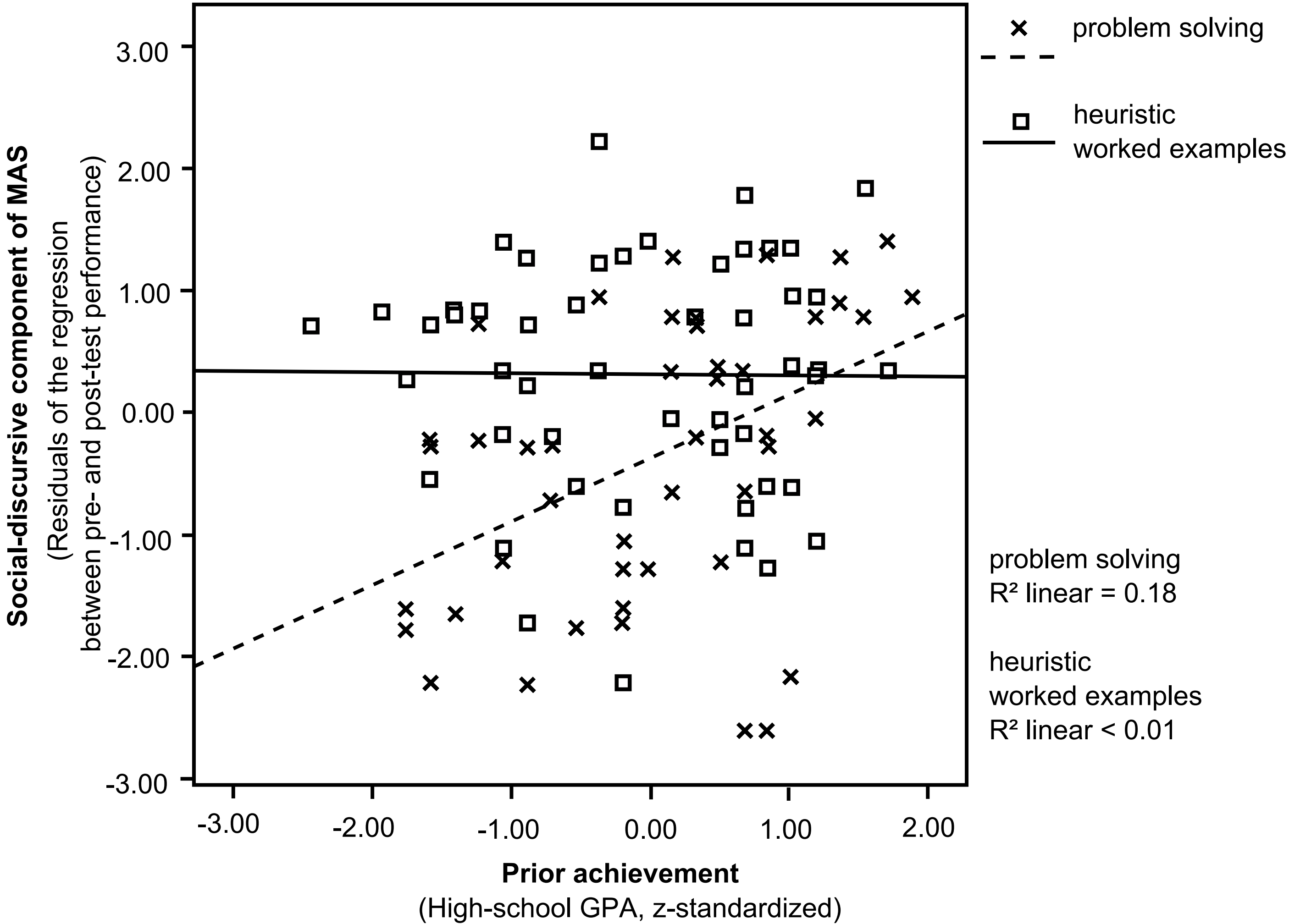
Kathi: But Finn calculates more examples than Tim.

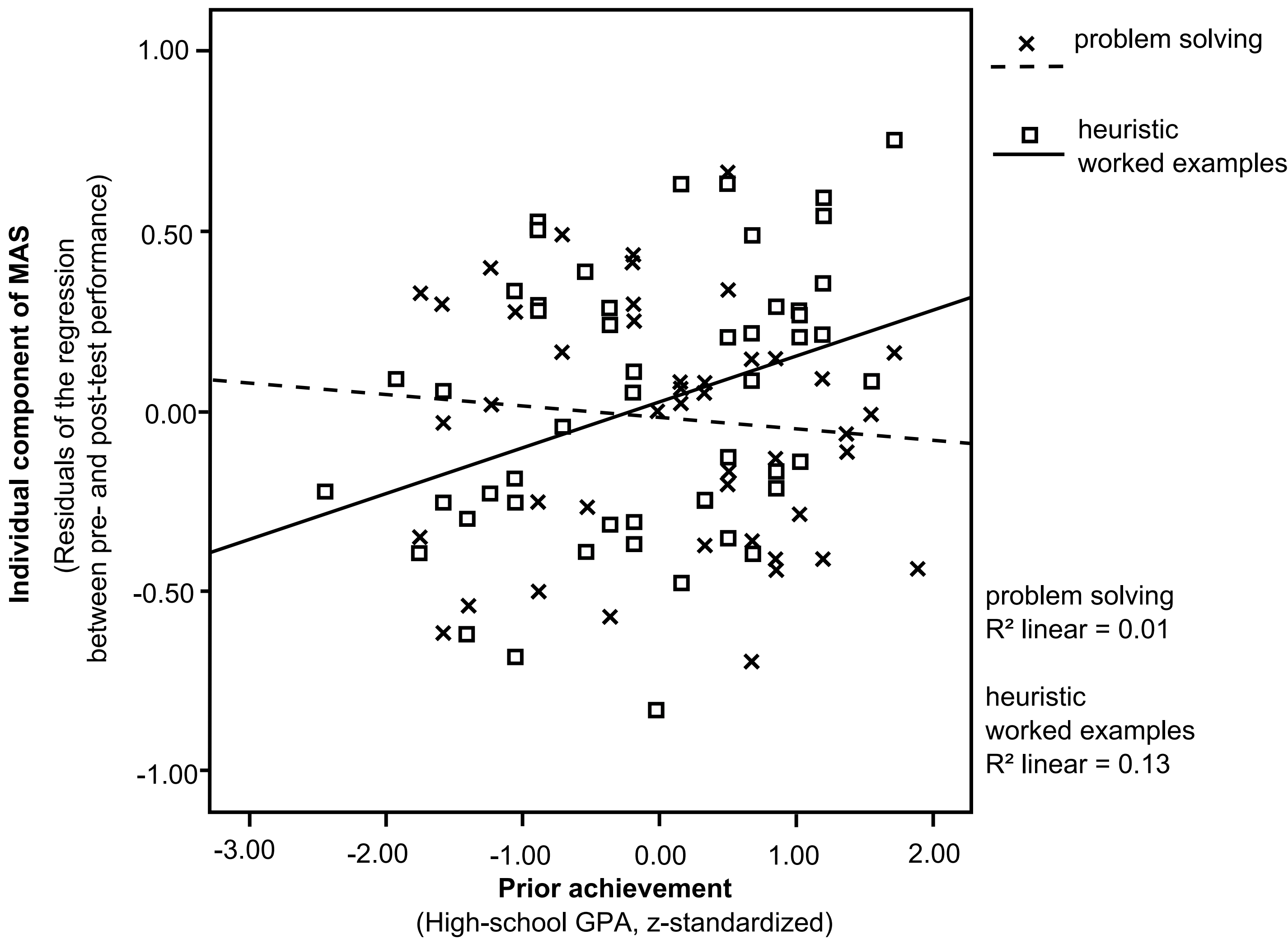
Page 1/1

Finish EXPLANATION | Finish ARGUMENT | Finish COUNTERARG | Finish SYNTHESIS | FINISH



without collaboration script
with collaboration script
Social-discursive component of MAS
(Residuals of the regression
between pre- and post-test performance)
3.00
2.00
1.00
0.00
-1.00
-2.00
-3.00
-3.00
-2.00
-1.00
0.00
1.00
2.00
Prior achievement
(High-school GPA, z-standardized)
without collaboration script
R² linear < 0.01
with collaboration script
R² linear = 0.23


problem solving
heuristic
worked examples
Social-discursive component of MAS
(Residuals of the regression
between pre- and post-test performance)
3.00
2.00
1.00
0.00
-1.00
-2.00
-3.00
-3.00
-2.00
-1.00
0.00
1.00
2.00
Prior achievement
(High-school GPA, z-standardized)
problem solving
R² linear = 0.18
heuristic
worked examples
R² linear < 0.01



problem solving
heuristic worked examples
Individual component of MAS
(Residuals of the regression between pre- and post-test performance)
1.00
0.50
0.00
-0.50
-1.00
-3.00
-2.00
-1.00
0.00
1.00
2.00
Prior achievement
(High-school GPA, z-standardized)
problem solving R² linear = 0.01
heuristic worked examples R² linear = 0.13

Table 1: Post-test answer that was rated with a high value for the social-discursive component of MAS

Student's answer	Coded elements
First phase: Everyone introduces her/his thesis or viewpoint. It is important to let the others finish their speech and to listen carefully to be able to refer to the others' thoughts later.	response to arguments
Second phase: Trying to convince the others by backing the own claims with arguments.	pro-argumentation
Third phase: Trying to respond to the other's arguments and formulating critique.	counter-argumentation, reponse to arguments
Fourth pase: The whole group tries to find a joint soulution by integrating the different arguments.	consensus building

Table 2: Mean pre-test and post-test scores (min. = 0, max. = 4; standard deviations in brackets) for the social discursive component of mathematical argumentation skill in the four experimental conditions.

	problem solving		heuristic worked examples	
	without collaboration script	with collaboration script	without collaboration script	with collaboration script
	M *(SD)*	M (SD)	M (SD)	M (SD)
Pre-test	1.65 *(1.52)*	1.27 *(0.70)*	2.04 *(1.40)*	1.35 *(1.16)*
Post-test	1.69 *(1.35)*	1.96 *(1.21)*	2.48 *(1.09)*	2.69 *(1.09)*

Table 3: Means (min. = 0, max. = 3; standard deviations in parentheses) of the pre-test and post-test scores in the tests on the individual-cognitive component of mathematical argumentation skill in the four experimental conditions.

	problem solving		heuristic worked examples	
	without collaboration script	with collaboration script	without collaboration script	with collaboration script
	M *(SD)*	M *(SD)*	M *(SD)*	M *(SD)*
Pre-test	1.46 *(0.73)*	1.16 *(0.65)*	1.17 *(0.52)*	1.34 *(0.55)*
Post-test	1.71 *(0.51)*	1.54 *(0.62)*	1.56 *(0.61)*	1.70 *(0.55)*