Computer-Supported Collaborative Learning
DOI 10.1007/s11412-011-9122-z

Are two heads always better than one? differential effects of collaboration on students' computer-supported learning in mathematics

Dejana Mullins · Nikol Rummel · Hans Spada

Received: 18 August 2010 / Accepted: 16 May 2011

Abstract While some studies found positive effects of collaboration on student learning in mathematics, others found none or even negative effects. This study evaluates whether the varying impact of collaboration can be explained by differences in the type of knowledge that is promoted by the instruction. If the instructional material requires students to reason with mathematical concepts, collaboration may increase students' learning outcome as it promotes mutual elaboration. If, however, the instructional material is focused on practicing procedures, collaboration may result in task distribution and thus reduce practice opportunities necessary for procedural skill fluency. To evaluate differential influences of collaboration, we compared four conditions: individual vs. collaborative learning with conceptual instructional material, and individual vs. collaborative learning with procedural instructional material. The instruction was computer-supported and provided adaptive feedback. We analyzed the effect of the conditions on several levels: Logfiles of students' problem-solving actions and video-recordings enabled a detailed analysis of performance and learning processes during instruction. In addition, a post-test assessed individual knowledge acquisition. We found that collaboration improved performance during the learning phase in both the conceptual and the procedural condition; however, conceptual and procedural material had a differential effect on the quality of student collaboration: Conceptual material promoted mutual elaboration; procedural material promoted task distribution and ineffective learning behaviors. Consequently, collaboration positively influenced conceptual knowledge acquisition, while no positive effect on procedural knowledge acquisition was found. We discuss limitations of our study, address methodological implications, and suggest practical implications for the school context.

D. Mullins · H. Spada (✉)
Institute of Psychology, University of Freiburg, Engelbergerstr. 41, 3. OG, 79085 Freiburg, Germany
e-mail: spada@psychologie.uni-freiburg.de

D. Mullins
e-mail: dejana.mullins@gmail.com

N. Rummel
Institute of Education, Ruhr-Universität Bochum, Bochum, Germany
e-mail: nikol.rummel@rub.de

Keywords Computer-supported collaborative learning · Learning in mathematics · Procedural and conceptual knowledge acquisition · Tutored problem-solving

Introduction

Current standards for teaching mathematics emphasize the importance of collaborative learning for students' knowledge acquisition (KMK 2004; NCTM 2000). Indeed, many studies have demonstrated the potential effectiveness of collaboration for improving problem-solving and learning (Berg 1994; Ellis et al. 1993; Slavin 1996). The positive effect of collaboration can be explained by the promotion of elaborative meaning-making activities. In a collaborative setting, students provide *explanations* to their partners (cf. Hausmann et al. 2004; Webb 1989); this requires them to make their thinking explicit and verbalize their knowledge. Often they have to reformulate and clarify their statements if their partner has difficulties in understanding their explanations. This verbalization and reformulation of knowledge demands elaboration of the learning content (O'Donnell 1999) and thus can promote knowledge acquisition. Furthermore, *joint elaboration* of the learning material can promote learning. Particularly in the domain of mathematics, knowledge co-construction has been shown to yield improved student achievement (Berg 1994). Finally, students can learn by *asking for help and receiving explanations* from a partner (Webb 1989). For instance, clarification questions enable the student to fill knowledge gaps and correct misconceptions.

Nevertheless, beneficial effects of collaboration on knowledge acquisition cannot always be found (e.g. Souvignier and Kronenberger 2007). Lou et al. (1996) evaluated the impact of collaboration in a meta-analysis. Although most results were in favour of collaborative learning, about a fourth of the results showed none or even negative effects when compared to individual learning. In earlier studies, we found indications that the impact of collaboration on mathematical knowledge acquisition may depend on the type of knowledge that students are trying to acquire during collaboration (Diziol et al. 2007, 2009). When students collaborated on conceptual problem-solving steps, they talked to each other and provided mutual explanations. This positive collaborative behavior yielded improved learning outcome in a conceptual post-test when compared to individual learning (Diziol et al. 2007). However, when students collaborated on procedural problem solving-steps, they didn't engage in mutual elaboration. Instead, they often took turns in solving the different problem-solving steps. In other words, the differences in the learning material seemed to trigger different types of collaborative behavior that were not equally effective for promoting student learning.

While the observations collected in these earlier studies suggested that the type of knowledge that is targeted by the learning material may affect the success of collaborative learning, we had not yet investigated the differential impact of collaboration on knowledge acquisition experimentally. The present study aims at increasing our understanding of differential effects of collaboration on learning in mathematics by empirically comparing individual and collaborative learning with conceptual and procedural instructional material. The instruction was computer-supported and provided adaptive feedback in the form of error-flagging and hint messages. The learning environment automatically recorded students' problem-solving in a logfile and thus enabled us to analyze the learning processes in detailed fashion. In the following sections, we will give a short overview of the distinction between conceptual and procedural knowledge acquisition in algebra, the mathematical domain of our study. Then we will discuss results regarding these two knowledge types from the literature on collaborative learning. We will conclude the theoretical background with an overview of our hypotheses and dependent variables.

Conceptual and procedural knowledge

Literature on knowledge acquisition in mathematics often distinguishes between conceptual and procedural knowledge. Conceptual knowledge is described as the understanding "of the principles that govern a domain and of the interrelations between pieces of knowledge in a domain" (Rittle-Johnson and Alibali 1999, p. 175). Particularly important concepts in the area of algebra, the domain of the present study, are the equation, the variable, and the constant term. These concepts can be represented in different formats: verbally in a story problem ("they earn $2 per glass sold"), graphically in a coordinate plane, algebraically in an equation ("+ 2*x*"), or in a table (cf. Brenner et al. 1997). One important aspect of students' conceptual understanding is reflected in their ability to flexibly translate between these representations (Brenner et al. 1997; Mevarech and Stern 1997).

Procedural knowledge can be defined as students' ability to execute stepwise action sequences to find the solution to a problem (Rittle-Johnson and Alibali 1999). By repeatedly solving tasks that require these procedures, students can gain skill fluency. Typical examples from algebra are manipulation problems such as solving equations for *x* (Brenner et al. 1997; Nathan et al. 1994). If students know the relevant procedures, they can easily solve these tasks.

The influence of collaboration on conceptual and procedural knowledge acquisition

For several reasons, research on collaborative learning so far does not support definite conclusions concerning the differential influence of collaboration on conceptual and procedural knowledge acquisition. The already mentioned meta-analysis by Lou et al. (1996) showed that positive results of collaboration can mainly be found in studies that provide additional instruction to collaborative conditions that is not given to students learning individually. Thus, it is unclear if the positive effect is due to the collaboration or due to the additional instruction. For instance, in a study by Berg (1994), a collaboration script supported dyadic problem-solving and prompted students to engage in mutual explanations. Post-test comparisons showed that students who learned collaboratively outperformed individual learners. However, as the script instructions were not provided to students learning individually, the positive effect of collaboration could also be ascribed to the instruction to elaborate on the underlying mathematical background.

Another area of confusion concerns the test items used for assessing learning. Often, the test material does not separately assess the two knowledge types, but both conceptual and procedural knowledge are required to solve the problems (e.g. Diziol et al. 2007). Thus, it is not clear from the test results if collaboration had a positive influence on either conceptual or procedural knowledge, or both. The present study aims at solving these confusions by distinguishing more clearly between conceptual and procedural knowledge both in instructional and test materials.

We hypothesize that conceptual and procedural instructional material elicits different types of collaborative learning processes, and that the elicited learning processes are not equally effective in promoting student learning. Conceptual instructional material elicits elaborative meaning-making processes. Particularly the translation between different conceptual representations is challenging for students (Brenner et al. 1997), thus students have to reason about the learning content in order to solve problems and to increase their understanding (Hiebert and Wearne 1996; Nokes and Ross 2007). For instance, when students solve algebra word problems, they have to reflect on the translation of the verbal problem description into the algebraic equation. Thereby, the application of simple translation rules based on keywords may be misleading (cf. Nathan et al. 1992; e.g., "the depth increases by 3 m/h" may have to be translated to "$-3x$", even though the word "increase" normally refers to a positive variable term). Instead, students have to correctly represent the problem scenario described, extract the important information, and transform this information into a different, that is, a mathematical representation format (Staub and Reusser 1995). Collaborative learning settings have the potential to increase beneficial elaborative learning mechanisms as students have to make their thinking explicit to their learning partner (Teasley 1995). Therefore, collaborative learning can be expected to promote learning with conceptual instructional material and to yield improved conceptual knowledge acquisition when compared to individual learning.

In contrast, procedural instructional material focuses students' attention on step-wise problem-solving procedures. In a collaborative setting, the step-wise procedures entail the danger that students will take turns in solving the problem-solving steps: As soon as one student knows the solution for a problem-solving step, he or she may enter it in the system. In other words, collaborative learning with procedural instructional material may lead to a division of practice opportunities between partners. However, as practice and the application of the problem-solving procedures is crucial to gain procedural skill fluency (Anderson 1983), the reduced amount of practice in a collaborative setting may be harmful for procedural knowledge acquisition.

Hypotheses

To assess the effect of collaboration on conceptual and procedural knowledge acquisition, we compared four conditions: individual versus collaborative learning with conceptual instructional material, and individual versus collaborative learning with procedural instructional material. The instruction was implemented in a computer-supported environment. Addressing the critique that previous research on collaborative learning in mathematics did not distinguish between conceptual and procedural knowledge in the test material, we assessed the effect of the four conditions on both conceptual and procedural knowledge acquisition.

Our main hypothesis concerns the differential impact of collaboration: We hypothesize that collaborative learning with conceptual instructional material elicits mutual *elaboration* on mathematical concepts and thus promotes students' *conceptual* understanding when compared to individual learning. In contrast, we expect that collaborative learning with instructional material that focuses on practicing procedures may promote *task distribution* and thus yield similar or less *procedural* skill fluency than individual learning.

Furthermore, we expect a condition specific main effect of the instructional material on students' knowledge acquisition; in other words, conceptual instruction should mainly improve conceptual knowledge acquisition, while procedural instruction should mainly improve students' procedural knowledge acquisition. This hypothesis also serves as manipulation check to evaluate the effectiveness of the instructional material.

We investigated the effect of collaboration on learning in mathematics at different levels. *Student performance during the learning activity* is usually the first observable indicator for the effectiveness of collaboration in the school setting, and thus is often used by teachers to decide whether to use a collaborative learning setting or not. However, from an educational viewpoint, testing their *individual knowledge acquisition* is also of great importance in order to determine if students are able to apply their knowledge subsequently. Furthermore, to better understand possible differential effects of collaboration on student learning, we also have to evaluate their *learning and interaction processes*, analyze how these processes relate to the learning outcome, and investigate under which conditions collaboration increases beneficial learning processes.

Method

Participants and study design

Seventy-nine students participated in the study. Participants were recruited from two local high schools on a voluntary basis and got paid for their participation. As one of the schools was a girls' school, we restricted participation to female students in order to avoid a confounding of gender and school. Students were in grade 8 (age M=13.18, SD=.50) and had already basic experience with the task domain. A two-factorial design was implemented (see Table 1): *instructional material* (conceptual vs. procedural) and *setting* (individual vs. collaborative). Prior to the study, we asked students which class mate they would particularly like to work with if they were selected for one of the collaborative conditions. Then, we randomly assigned these potential pairs to the four conditions, distributing students from the two schools evenly across study conditions (block randomization). This resulted in the following numbers: conceptual individual learning (19 students), conceptual collaborative learning (20 students), procedural individual learning (20 students), and procedural collaborative learning (20 students). In the collaborative conditions, students collaborated with the partner they had chosen; in the individual conditions, both students of a potential pair worked individually.

In order to enable us to compare the learning processes in the individual and the collaborative conditions, half of the students in both the conceptual individual and the procedural individual condition were randomly selected and asked to think aloud while solving the problems. We recorded audio and video during the learning phase: in the individual conditions, we recorded individual students thinking aloud, and in the collaborative conditions, we recorded students interacting with each other in dyads. To reduce the risk of student reactivity, the think aloud directions followed the guidelines described in Ericsson (2003). Students first received a short instruction to the think-aloud method that asked them to simply verbalize each thought that emerges. To familiarize them

Table 1 Study design and procedure

Instructional material	Conceptual		Procedural	
Setting	Individual N=19	Collaborative N=20	Individual N=20	Collaborative N=20
Pre-test	Individual problem-solving (paper-pencil, order counterbalanced across conditions): conceptual and procedural problem-set			
Learning phase	Individually or in dyads: conceptual instruction (tutored learning environment)		Individually or in dyads: procedural instruction (tutored learning environment)	
10-min break				
Post-test	Individual problem-solving (paper-pencil, order counterbalanced across conditions): conceptual near and far transfer problem-set procedural near and far transfer problem-set			

with the method, they practiced thinking-aloud while solving a sorting task that was not related to mathematics and the learning content of the study. In the sorting task, a picture story had been mixed up, and students were asked to find the correct order of the pictures. If students stopped talking, the experimenter reminded them to continue verbalizing their thoughts. Statistical comparison of students' performance during the learning phase and of their learning outcome (see list of dependent variables, Table 2) confirmed that thinking aloud did not influence student performance and learning outcome: Neither in the procedural individual nor in the conceptual individual condition did we find differences between students thinking aloud and the other individual students (for all analyses, $p > .10$). We therefore combined the think aloud students and non-think aloud students within the respective individual conditions for the quantitative analyses.

Procedure

The study procedure consisted of three parts: pre-test, learning phase, and post-test (see Table 1). In order to assess prior knowledge, participants first worked individually on a pre-test that contained conceptual and procedural problems. The test was delivered in paper and pencil fashion. For the learning phase, students moved to the computer where they received instruction according to their condition. In the collaborative conditions, two students worked together on one computer to solve the tasks (i.e., face-to-face interaction). After the learning phase, there was a short break before students took the post-test. As was the case for the pre-test, the post-test was solved individually on paper. It consisted of four problem-sets: a near and a far transfer problem-set for each of the two knowledge types. Students solved the problems at their own pace both during pre- and post-test and during instruction. In total, the experiment lasted about 140 min.

Learning environment and instructional material

We implemented the instruction during the learning phase in a computer-supported learning environment. This implementation enabled us to provide tutoring support to students' problem-solving actions, a form of instructional support that has been shown to be particularly beneficial for student learning. A particularly prominent example for the success of tutoring environments are the Cognitive Tutors for mathematics instruction (e.g.

Table 2 Overview of dependent variables

	Dependent variable	Operationalization	Data source
Learning phase[a]	Performance	Error rate	Logfiles
	Learning process	Time before action	
		Time after error	
		Elaboration after errors	Audio recordings
		Student interaction after errors	
Test phase	Conceptual knowledge acquisition	Conceptual near transfer	Post test scores
		Conceptual far transfer	
	Procedural knowledge acquisition	Procedural near transfer	
		Procedural far transfer	

[a] Depending on the condition, students learned either with conceptual or procedural instructional material

Algebra, Geometry and Integrated maths) that were developed at Carnegie Mellon University, Pittsburgh. These tutoring curricula are widely used in regular classrooms across the US to teach mathematics at the high school level and have been shown to improve knowledge acquisition when compared to traditional classroom instruction (e.g. Koedinger et al. 1997). Their success is based on an evaluation of the student's knowledge that enables adaptive support tailored to the student's needs. The Tutors provide immediate error feedback, answer to help requests, and select problems that target skills that are not yet mastered by the student.

Similar to the Cognitive Tutors, the learning environment in our study was designed to provide adaptive support to students. We implemented our learning environment with the Cognitive Tutor Authoring Tools (CTAT; Aleven et al. 2009), a software that enables researchers and teachers to author intelligent tutoring behavior. The learning environment provided immediate feedback to student actions by marking errors in red and correct answers in green. Furthermore, it automatically launched a hint after the third incorrect student attempt to ensure that students would not get stuck during problem-solving (see Fig. 2). The hint message told students the correct solution to the problem-solving step. To prevent students from exploiting this help functionality, they were not told about it. In contrast to the Cognitive Tutors, a functionality to ask for help and an automatic selection of problems was not implemented in our environment. The tutored problem-solving was alternated with worked example study. The learning environment automatically logged all student actions to allow a detailed analysis of the learning processes.

The task domain of the study was algebra, more specifically linear functions. The learning material in the conceptual and procedural conditions differed in the following way: In the conceptual conditions, students were asked to derive linear equations from story problems. For instance, in the story problem in Fig. 1, Peter is scuba-diving and students were requested to find an algebraic equation that represented his depth. They were,

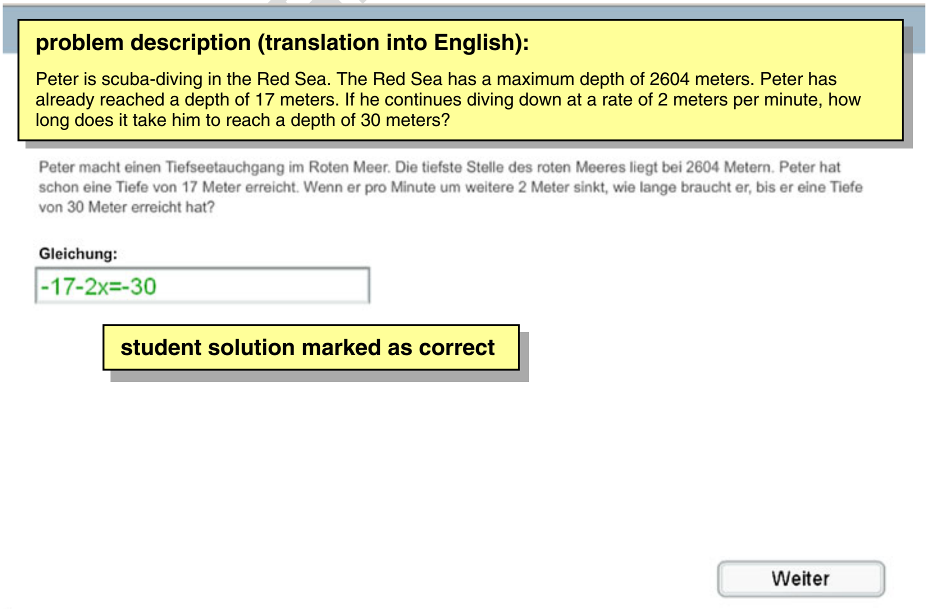


Fig. 1 Screenshot of the conceptual learning environment

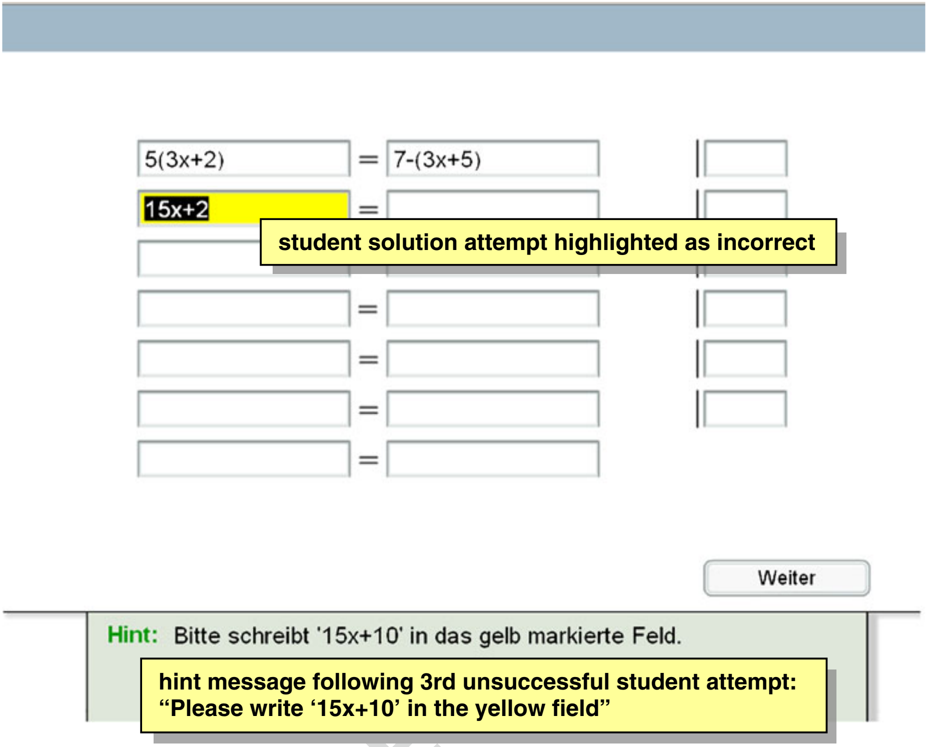


Fig. 2 Screenshot of the procedural learning enviornment

however, not asked to solve the equation. The problems were of increasing difficulty, reaching from simple story problems that only contained a constant term to story problems with variable and constant terms, several variable terms, negative constant or variable terms, and brackets. Students in the conceptual conditions received one worked example for each level of difficulty and altogether solved 15 problems on their own. The conceptual worked examples focused on the translation of verbal concept representations into algebraic concept representations.

In the procedural conditions, students practiced solving linear equations (see Fig. 2). Again, the problems had increasing difficulty, reaching from simple equations with one variable and one constant term to equations with negative constant terms, negative variable terms, several variable terms (e.g. $8x + 5 + 6x = 12$), and subtraction and multiplication brackets. As in the conceptual conditions, students received one worked example for each level of difficulty and altogether solved 15 problems on their own. The worked examples focused on the procedures necessary to solve the equations. In both the conceptual and the procedural conditions, students could only proceed to the next problem once they had correctly solved the problem at hand.

Dependent variables

To gain a deeper understanding of the effects of collaboration, we evaluated the effects of our experimental conditions at several levels based on different data sources: logfiles, audio recordings, and post-test score (for an overview, see Table 2). The *performance* of students during the learning activity served as first indicator for the effectiveness of collaboration when compared to an individual setting. However, good collaborative performance may not necessarily promote individual knowledge acquisition. Therefore, our study also evaluated the impact of collaboration on *students' learning processes* during instruction, and on their *learning outcome* (i.e., conceptual and procedural knowledge acquisition) as measured by the post-test. Students' prior knowledge was analyzed as a control variable based on the pre-test score. The following sections describe the operationalization of these dependent variables in more detail.

Student performance during learning phase: Error rate As a first step in evaluating the influence of the learning setting (individual vs. collaborative) on knowledge acquisition in mathematics, we assessed the performance during the learning phase based on the variable *error rate* extracted from the log data. This variable measures the relative number of errors on the first attempt to solve a problem-solving step. An error rate of 1 indicates that a student solved each step incorrectly on the first attempt; an error rate of 0.5 indicates that on average, half of the steps were solved incorrectly, half were solved correctly on the first attempt; and an error rate of 0 indicates that all steps were solved correctly on the first attempt.

Learning processes: Time variables To validate the process model that underlies the hypothesized differential effect of collaboration, we analyzed student learning processes in more detail. Particularly, we were interested in assessing if collaboration increased beneficial elaboration behavior, or rather promoted task distribution. As a first step to answer this question, we evaluated the average *time* spent *before an action* and the average *time* spent *after an error* (measured in seconds). As elaboration takes time, the analysis of these variables can serve as indicators of cognitive processes in problem-solving (cf. Diziol et al. 2009). Thus, in a collaborative condition longer times before an action could indicate mutual elaboration, whereas shorter times could indicate task division. These variables are highly objective and can easily be assessed automatically; on the other hand, they leave a lot of room for speculation about what actually happened during these times. In a second step, we therefore analyzed the actual individual and collaborative learning processes in order to disambiguate what was going on.

Learning processes: Coding analysis of learning from errors To shed further light on the results of the log data analyses, we evaluated relevant aspects of the think aloud recordings of individual students and of the dialogue of collaborating dyads, using a coding scheme. As the analysis of verbal data is very time consuming (Chi 1997; Reimann 2007), we concentrated our analysis on one aspect of student learning that has been shown to be a particularly important predictor of student learning in intelligent tutoring systems: learning processes following errors. Earlier studies have shown that student behavior after errors can be critical for successful knowledge acquisition (e.g. Baker et al. 2004). When students elaborate on an error and its correction, they can increase their understanding. However, when they engage in trial and error behavior, that is, try several different answers until the learning environment marks one answer as correct, they cannot capitalize on the learning opportunity. We analyzed students' learning processes around errors, taking into account two aspects: *elaboration processes* and *task distribution* when trying to correct the errors (see also Diziol et al. 2010b). For the analyses, we devised a coding scheme and implemented it using the Activity Lens software (Avouris et al. 2007). The software, Activity Lens, supports researchers in the analysis of collaborative learning and interaction.

Different data sources—for instance, audio, video, and log data—can be entered and synchronized. For our analysis, we linked log data from the learning environment with video recordings from individual or collaborative problem-solving. The synchronization of the data sources enabled us to navigate to relevant sequences of the video (e.g. student behavior after errors) for the process analysis.

In the analysis of *elaboration processes* after errors, we distinguished between two types of errors: errors that were corrected in the subsequent step (*error corrected*) and errors that were followed by a subsequent error (*next step incorrect*). The following three codes were used to specify how errors were corrected: If students elaborated on the error to find the correct solution, their problem-solving action was coded as *elaboration*. If students did not verbally elaborate on the error, but remained silent for a while before they corrected the error, the action was coded as *no elaboration*; this code was also applied to utterances where the student repeated the problem description aloud or verbalized his or her suggestion for the next step without further explanation. If students immediately corrected the error without providing an explanation, the action was coded as *immediately corrected*. As several studies by Webb and colleagues have shown (for an overview, see Webb 1989), the latter behavior is often detrimental for the partner's knowledge acquisition in a collaborative setting, as she may not understand the error correction without further explanation. Similarly, we used three codes to specify student behavior after errors that were followed by a subsequent error (*next step incorrect*). The first and second code, *elaboration* and *no elaboration*, correspond to the codes for errors corrected; the third code *trial and error* was applied if students exhibited trial and error behavior. To check the inter-rater reliability, a second coder reanalyzed eight of the 20 individuals thinking aloud and eight of the 20 collaborating dyads, respectively. The inter-rater reliability for the elaboration dimension was κ=.77.

Furthermore, with the variable *task distribution* (inter-rater reliability κ=.68) we evaluated if the two students worked together on getting past the error, or if they distributed the task between them: Did students collaborate to correct the error (*both*), did they distribute the task, thus only one student was responsible for the action following the error (*one*), or did they not discuss the error correction at all (*none*)? This variable was only evaluated for the collaborative conditions. If a high amount of behavior after errors were coded as *both,* this would indicate collaborative interaction that could be beneficial for learning. If, on the other hand, a high amount of behavior after errors were coded as *one* or *none,* this would indicate a task distribution that could have a negative impact on the individual knowledge acquisition.

Learning outcome assessed in the post-test After the learning phase students solved a post-test on paper. We adapted the test material from an earlier study (Diziol et al. 2009). The test consisted of four problem-sets: conceptual near and far transfer and procedural near and far transfer. The *near transfer problems* were structurally equivalent to the problems solved during the learning phase; however, now students had to solve the problems on paper without receiving tutoring support. For conceptual near transfer, students had to derive linear equations from story problems; for procedural near transfer, students were asked to solve linear equation problems.

The problems in the *conceptual far transfer problem-set* required a reverse translation between representations: Students received an equation and several keywords; they were instructed to use the keywords to formulate a story problem corresponding to the given linear equation. Conceptual understanding should enable students to verbalize the functional relationship represented in the equations, that is, to translate the algebraic problem representation into a verbal representation (Brenner et al. 1997). We evaluated the concordance between the linear equations and the story problems written by the students with scores ranging from 0 to 3. Students received a score of one if the story problem contained all relevant values, but there were major errors concerning their functional relationship (e.g. if students confused the variable and the constant term), a score of two if the story problem contained all relevant values, but there were minor errors concerning their functional relationship, and a score of three if the story problem was concordant with the algebraic equation. The scoring system was based on the cognitive processes involved in story problem solving which are described in Staub and Reusser (1995).

In the *procedural far transfer problem-set*, students received erroneous problem-solutions of a fictitious student and were asked to find the errors. The problem-solutions contained typical computational errors such as combining constant and variable terms when solving equations for x. Procedural knowledge should help students find these errors.

To evaluate inter-rater reliability, a second coder analyzed a quarter of the tests, yielding good agreement on all scales (for the conceptual near transfer problem-set and the procedural problem-sets, κ=.88 each; for the conceptual far transfer problem-set, $ICC_{2,1}$ r=.97). For each of the four problem-sets, we added the points a student had achieved in the single tasks to one score. The maximum score that could be reached differed between problem-sets. To support the reader's understanding, we use percentages of the maximum score that were reached by the students to report the results.

Prior knowledge as covariate We evaluated prior knowledge in algebra with a pre-test. The pre-test consisted of a conceptual and a procedural problem-set and was solved on paper. The problems were structurally equivalent to the problems of the learning phase, but had a lower difficulty level to avoid de-motivating and frustrating students. We added the z-transformed conceptual and procedural pre-test scores to a combined measure of prior knowledge in algebra. Conditions did not differ concerning their prior knowledge. As prior knowledge correlated significantly with students' performance during the learning phase and with their learning outcome in the post-test, we included it as covariate (see also results of the covariance analyses, Tables 3, 4, 5, 6).

Table 3 Learning phase: comparison of conditions with conceptual instructional material

Setting:	Individual	Collaborative		F	p	$\eta 2$
Performance[a]						
Error rate						
M(SD)	.44 (.17)	.31 (.07)	Prior knowledge	12.37	<.01**	.32
			Setting	3.37	.08+	.12
Time variables[b]						
Time before action						
M(SD)	64.16 (18.66)	84.50 (26.74)	Setting	5.76	.02*	.18
Time after error						
M(SD)	44.79 (20.18)	74.30 (47.38)	Setting	5.59	.03*	.17

For error rate, less means better

[a] df = 1,26 for performance variables. [b] df = 1,27 for time variables

**p<.01; *p<.05; +p<.10

Table 4 Learning phase: comparison of conditions with procedural instructional material

Setting:	Individual	Collaborative		F	*p*	η2
Performance[a]						
Error rate						
M(SD)	.15 (.09)	.09 (.04)	Prior knowledge	7.43	.01**	.22
			Setting	4.80	.04*	.15
Time variables[b]						
Time before action						
M(SD)	17.70 (3.83)	14.70 (3.06)	Setting	4.64	.04*	.14
Time after error						
M(SD)	22.10 (7.53)	16.80 (11.54)	Setting	2.30	.14	.08

For error rate, less means better

[a] df = 1,27 for performance variables. [b] df = 1,28 for time variables

$^{**}p<.01$; $^{*}p<.05$; $^{+}p<.10$

Results

Learning phase

We evaluated both performance and process data from the learning phase. The instructional material in the conceptual and the procedural conditions was not directly comparable (e.g. different type of tasks, different number of steps per problem, …). We therefore compared individual and collaborative learning separately within the conceptual conditions and within the procedural conditions. For the collaborative conditions, the analyses were based on dyadic student data (i.e. one data point per dyad).

Performance during the learning phase We employed an analysis of variance to evaluate the impact of collaboration on student performance (error rate) during the learning phase.

Table 5 Post-test: comparison of students' conceptual knowledge acquisition

Instructional material:	Conceptual		Procedural			*F*	*p*	*η2*
Setting:	Ind	Coll	Ind	Coll				
Conceptual near transfer								
M	39%	51%	23%	22%	Prior knowledge	34.00	<.01**	.32
SD	14%	17%	23%	15%	Factor instruction	41.62	<.01**	.36
					Factor setting	.72	.40	.01
					Interaction	4.46	.04*	.06
Conceptual far transfer								
M	57%	63%	45%	38%	Prior knowledge	17.13	<.01**	.19
SD	26%	24%	22%	19%	Factor instruction	13.07	<.01**	.15
					Factor setting	.34	.56	.00
					Interaction	2.78	.10+	.04

df = 1,74 for all analyses

$^{**}p<.01$; $^{*}p<.05$; $^{+}p<.10$

Table 6 Post-test: comparison of students' procedural knowledge acquisition

Instructional material:	Conceptual		Procedural			F	*p*	η2
Setting:	Ind	Coll	Ind	Coll				
Procedural near transfer								
M	56%	59%	72%	71%	Prior knowledge	45.83	<.01**	.38
SD	20%	29%	18%	22%	Factor instruction	16.26	<.01**	.18
					Factor setting	.33	.57	.00
					Interaction	.76	.39	.01
Procedural far transfer								
M	67%	74%	89%	79%	Prior knowledge	20.52	<.01**	.22
SD	26%	32%	13%	32%	Factor instruction	8.11	.01*	.10
					Factor setting	.77	.38	.01
					Interaction	3.08	.08+	.04

df = 1,74 for all analyses
***p*<.01; **p*<.05; +*p*<.10

As performance was significantly related to prior knowledge, we included prior knowledge as a covariate. As mentioned above, we conducted two separate analyses with the independent variable learning setting: conceptual individual vs. conceptual collaborative, and procedural individual vs. procedural collaborative.

In both the conceptual and the procedural conditions, students who worked in a collaborative setting showed better performance during the learning phase than students who solved problems individually (see Tables 3 and 4). In the *conceptual conditions,* we found a marginally significant effect of the setting; descriptively, dyads made fewer errors than students who learned individually. In the *procedural conditions*, we found a significant difference between conditions; again, dyads had a lower error rate than students working individually.

Learning processes: Time variables The time variables served as a first indicator for learning processes. Again, we employed an analysis of variance with learning setting as the independent variable. In addition, we correlated the time variables with students' outcome in the respective near transfer problem-set as we wanted to see if the learning processes were related to students' learning outcome as assessed in the post-test. For the correlation analyses, we will only report significant results.

Depending on the type of instruction, the learning setting influenced the time variables in opposite directions. In the *conceptual conditions*, dyads spent significantly more *time before actions* and *time after errors* than individuals (Table 3). The time variables were positively related to the conceptual understanding in the post-test: Students who had spent more *time before actions* and more *time after errors* during the learning phase showed better learning outcomes in the conceptual near transfer problem-set (r=.47, p<.01 and r=.61, p<.01, respectively). This suggests that collaborative learning with conceptual instructional material may have increased elaborative learning processes that promoted conceptual understanding.

In contrast, in the *procedural conditions* dyads spent less *time before actions* than students working individually (Table 4). While the analysis of the variable *time after error* did not reach significance, the result pointed in the same direction. This indicates that collaboration may not have promoted mutual elaboration on the procedural instructional material. Neither *time before action* nor *time after error* correlated with the learning outcome in the procedural post-test (for both analyses, $p>.10$).

Learning processes: Elaboration dimension As discussed above, the time variables are highly objective, but can only provide first indications for the learning processes of students. To better understand the differential influence of the setting depending on instructional material, we also analyzed think aloud protocols of individuals and the dialogue of collaborating dyads.

We compared the process variables *elaboration processes* and *task distribution* after errors with chi square statistics (unit of analysis: occurrence of errors). Furthermore, we correlated the learning process codings with the learning outcome in the respective near transfer problem-set of the post-test. As the analysis of the error rate had indicated significant differences in the number of errors between the individual and collaborative condition, the correlation analysis was based on the proportional occurrence of the respective behavior to avoid confounding. For the correlation analyses, we will only report significant results.

The comparison of elaboration processes during *conceptual instruction* revealed a significant effect of the setting for *error corrected*, $\chi^2(2)=9.39$, $p=.01$, and a marginally significant effect on student behavior for *next step incorrect,* $\chi^2(2)=4.87$, $p=.09$ (see Fig. 3). The descriptive comparison of the individual and collaborative condition showed that collaboration increased *elaboration* both for errors that were corrected and errors that were followed by a subsequent error while reducing the percentage of *no elaboration* when compared to the individual condition. Thereby, *elaboration* was positively related to the learning outcome in the near transfer test (elaboration when next step incorrect—conceptual near transfer: $r=.63$, $p<.01$), while *no elaboration* was negatively related to student learning (no elaboration when next step incorrect—conceptual near transfer: $r=-.58$, $p<.01$). Thus, collaboration with conceptual instructional material promoted effective learning processes and reduced ineffective learning behavior.

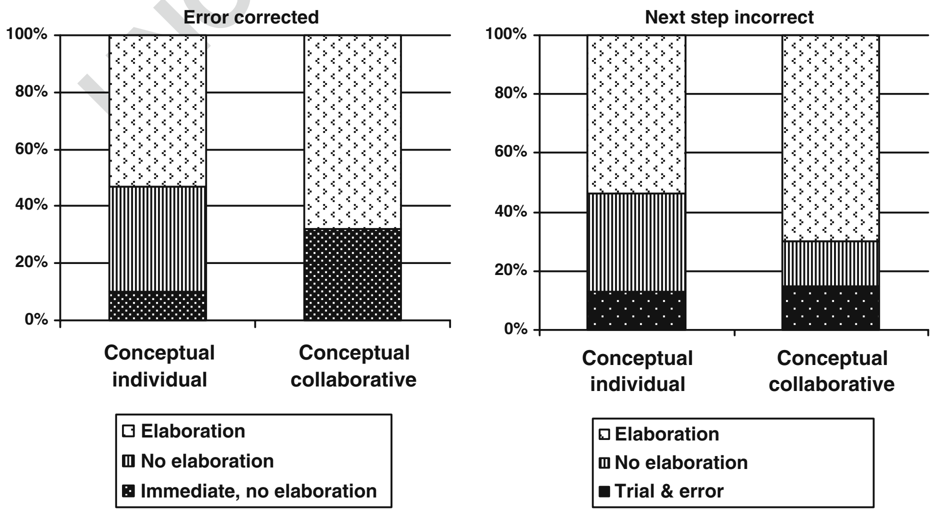


Fig. 3 Learning processes following errors in the conditions with conceptual instructional material

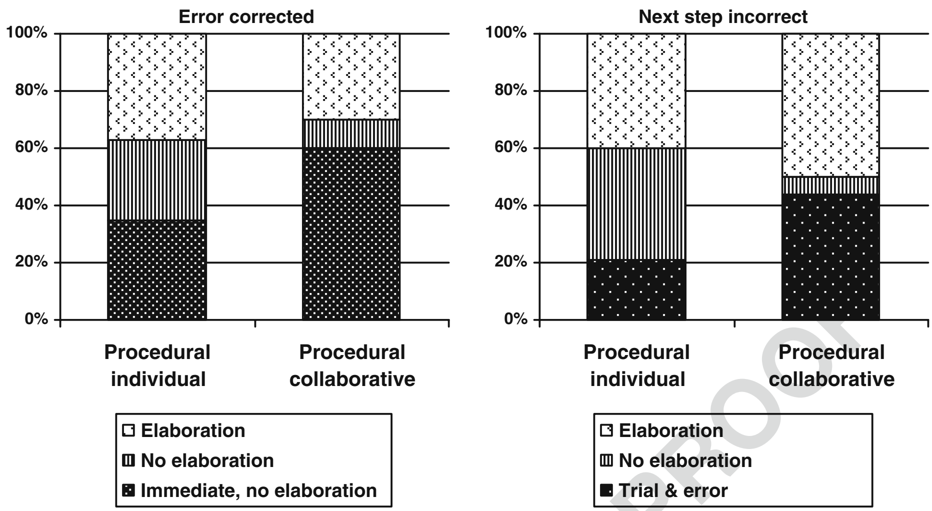
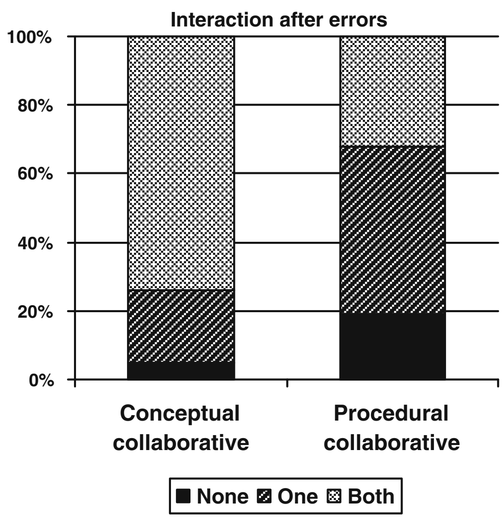


Fig. 4 Learning processes following errors in the conditions with procedural instructional material

Also in the *procedural conditions*, individual and collaborative learning processes differed significantly for *errors corrected,* $\chi^2(2)=12.77$, $p<.01$, and *next step incorrect,* $\chi^2(2)=7.04$, $p=.03$ (see Fig. 4). The descriptive comparison of the conditions revealed that collaboration increased *immediate error correction* for errors corrected; however, in contrast to the conceptual conditions, it reduced *elaboration* when compared to individual learning. In other words, students hardly explained the error correction to their learning partner. For next step incorrect, collaboration more than doubled the percentage of *trial and error* behavior (21% in the individual condition, 44% in the collaborative condition). As in the conceptual conditions, *elaboration* after errors positively correlated with procedural knowledge at post-test (elaboration when next step incorrect—procedural near transfer: $r=.49$, $p=.03$), while *trial and error* behavior showed a negative correlation with the post-

Fig. 5 Task distribution following errors: Percentage of errors that were discussed by none, one or both students of a dyad



test results (trial & error—procedural near transfer: $r=-.42$, $p=.06$). Thus, collaborative learning with procedural instructional material did not improve the learning processes, but increased the application of ineffective trial and error behavior.

Learning processes: Task distribution The comparison of the conceptual collaborative and the procedural collaborative condition revealed a significant difference in the amount of task distribution during error correction ($\chi^2=25.92$, $p<.01$, see Fig. 5). The descriptive comparison shows that in the conceptual collaborative condition, mostly *both* students were engaged in error correction (74% of errors) while in the procedural collaborative condition, the dyad partners tended to divide labor after errors: Most of the time, only *one* partner took responsibility for the next solution step (49% of errors), and frequently, dyads did not talk about the following step at all (*none* for 19% of errors). The consequential decrease of practice in the procedural collaborative condition was related to a lesser learning outcome: The percentage of errors corrected by the learning partner negatively correlated with student performance in the procedural near transfer test ($r=-.47$, $p=.04$).

Post-test performance

During the learning phase, students in the conceptual conditions and in the procedural conditions had worked with different instructional material. In contrast, in the test phase, every participant solved both the conceptual and the procedural problem-set. This enabled us to evaluate the impact of our four study conditions on conceptual and procedural knowledge acquisition with a two-factorial covariance analysis with instructional material (conceptual vs. procedural) as factor one, setting (individual vs. collaborative) as factor two, and prior knowledge as a covariate. The analysis of factor one can serve as manipulation check (did conceptual instruction improve the outcome in the conceptual post-test when compared to procedural instruction and vice versa?). The analysis of factor two evaluates if collaboration has a general effect as compared to individual learning. Finally, the interaction effect evaluates if collaboration has a specific effect on knowledge acquisition depending on the type of instructional material.

A problem often raised concerning the analysis of collaborative learning outcomes is the possible interdependence of data points: The individual post-test results of students who collaborated during the learning phase may be more similar than the test results of two independent learners, yielding an analysis bias (cf. Cress 2008). To address this issue, we analyzed the intraclass-correlations between individual post-test scores of dyad partners. For three of four outcome measures, we could not find an indication of an interdependency of the dyadic values; only for the variable conceptual near transfer, the analysis revealed a consequential non-independence (i.e., an intraclass correlation between dyad partners that is higher than $r=.45$ and significant at an alpha level of .20, as defined by Kenny et al. 1998). To keep the analyses of the different post-test sets comparable, we did not account for this correlation and included both dyad partners in the analysis individually.

Conceptual near and far transfer The analysis of the *conceptual near transfer* problem-set revealed a positive effect of conceptual instruction on the learning outcome (see Table 5): Students in the conceptual conditions were better at deriving equations from story problems than students in the procedural conditions (manipulation check). More importantly, the positive effect of learning with conceptual instructional material was particularly found for students in the conceptual collaborative condition as revealed by the significant interaction effect. In other words: collaboration improved students' conceptual knowledge acquisition. No significant general effect of the factor setting was found.

Similarly, we found a significant influence of the factor instructional material on the *conceptual far transfer problem-set* with higher test scores in the conceptual conditions (manipulation check). While the interaction effect was only marginally significant, the descriptive comparison again indicates that conceptual instruction was particularly effective for students who had learned in a collaborative setting. The factor setting did not show a significant effect.

Procedural near and far transfer Students in the procedural conditions reached a significantly higher test score in the *procedural near transfer* problem-set than students in the conceptual conditions (factor instructional material, i.e., manipulation check; see Table 6). However, although descriptively the best results were achieved by students in the procedural individual condition, neither the factor setting nor the interaction effect were significant.

Also in the *procedural far transfer* problem-set (see also Table 6), the factor instructional material had the expected specific effect: Students in the procedural conditions detected significantly more computational errors than students in the conceptual conditions. The interaction effect was only marginally significant, showing a trend for students who had practiced procedures individually to detect more errors than students who had practiced procedures together with a learning partner. No significant general effect of setting was found.

Discussion

Summary and discussion of study results

So far, research findings concerning the effect of collaboration on student learning in mathematics have been inconsistent: While some studies found positive effects, others found none or negative effects of collaboration on learning (Lou et al. 1996). Upon closer inspection, previous studies often confounded conceptual instruction and collaborative learning in their learning material and did not distinguish between conceptual and procedural knowledge acquisition at post-test. With the aim to increase our understanding of when and why collaboration is beneficial, the present study attempted to distinguish more clearly between the two knowledge types in both instructional and test material. The importance of this differentiation is confirmed by our post-test results: The type of instruction had a specific effect on student knowledge acquisition; in other words, conceptual instructional material improved conceptual knowledge acquisition, and procedural instructional material improved procedural knowledge acquisition.

Furthermore, we had hypothesized that the type of instruction would influence the quality of collaboration and its effectiveness for promoting learning. The results of our study partly support this assumption, and the process analyses helped to better understand the processes underlying this effect. The analysis of student collaboration confirmed that *conceptual instructional material* was able to stimulate mutual elaboration and explanation giving. Under this condition, we found that usually both learning partners were engaged in the collaborative activity, while division of labor was rare. The collaboration yielded a reduced number of errors during the learning phase as compared to individual learning. But more importantly, collaboration also improved the learning processes. Dyads in the conceptual collaborative condition showed an increased amount of elaboration of the underlying mathematical concepts as indicated both by the time variables and by the analysis of student dialogue after errors. Furthermore, dyads rarely engaged in negative learning processes such as trial and error behavior. The correlation analyses confirmed that this positive collaboration behavior was beneficial for students' conceptual knowledge acquisition. The positive effect of collaboration was also confirmed by a comparison of the post-test results: The conceptual collaborative condition reached the highest test scores in the conceptual near and far transfer problem-sets.

In contrast, collaborative learning with procedural instructional material did not have the same positive effect on students' learning processes and their learning outcome. The dependent variables draw the picture of a typical collaboration when practicing to learn procedures: Instead of mutual elaboration, collaboration on procedural instructional material promoted ineffective learning behavior such as trial and error. Furthermore, dyads often took turns in solving the different problem-solving steps and in correcting errors, in other words, the student who knew how to solve or correct a problem-solving step did so without conferring with his or her partner. Although distributing the task of error correction in this way may have contributed to the reduced amount of errors and to the reduced amount of time in the collaborative condition, students could not sufficiently benefit from the learning opportunities due to the lack of explanations by their partner as confirmed by the correlation analyses: When a student's learning partner corrected most of the errors, the student herself showed lower results at post-test. In line with the results of the process analyses, we could not find a positive effect of collaboration on the learning outcome: Students who had practiced procedures together with a learning partner showed comparable or even lower procedural knowledge acquisition than students of the procedural individual condition. To conclude, the results of our study revealed that collaboration is particularly beneficial for knowledge acquisition in mathematics if the learning material does not so much emphasize stepwise problem-solving, but requires elaborative learning activities and thus benefits from mutual explanations and joint discussions (see also Renkl 2008).

In our study we aimed at clearly distinguishing between conceptual and procedural knowledge both in the learning and test material. However, it is important to note that the two knowledge types are not totally independent (Hiebert and Wearne 1996)—and that it is often the goal of instruction to particularly strengthen their dialectic relationship. For instance, a high understanding of underlying concepts can help to monitor the appropriateness and execution of procedures, thus conceptual knowledge can influence the performance in procedural tasks. On the other hand, the execution of procedures can positively influence students' conceptual understanding if the students engage in active learning processes and try to understand the underlying principles (Rittle-Johnson 2006). Rittle-Johnson et al. (2001) therefore describe conceptual and procedural knowledge as two ends of a continuum that influence each other in an iterative way, in other words, improvement in one knowledge type can result in improvement in the other knowledge type (see also Perry 1991; Rittle-Johnson and Alibali 1999). In our study, we also found support for an interrelation between the two knowledge types (small to medium correlations: conceptual near transfer—procedural near transfer $r=.33$, $p<.01$; conceptual near transfer—procedural far transfer $r=.23$, $p=.04$; conceptual far transfer—procedural near transfer $r=.25$, $p=.03$; correlation between the two far transfer tests not significant). Thus, it may be an interesting endeavour for future research to evaluate the effect of collaboration on this relationship in more detail. Regarding our post-test scales, it is important to note that the correlations *within* each knowledge type were higher than *between* conceptual and procedural knowledge, thus supporting the differentiation we made between conceptual and procedural knowledge acquisition (for the conceptual post-tests, $r=.59$, $p<.01$; for the procedural post-tests, $r=.57$, $p<.01$).

Limitations of the study results and outlook

The current study investigated the differential effect of dyadic collaboration for a specific domain and in a specific, computer-supported setting. Future studies will have to evaluate if the established effects can be generalized to other areas in mathematics, such as geometry or arithmetic, to other domains, such as physics or chemistry, and to other settings. Indeed, ample research has shown that these factors can affect the impact of collaboration on knowledge acquisition. For instance, the meta-analysis by Lou and colleagues (1996) revealed that collaborative learning is more effective in mathematics and science instruction than in reading or arts, and more effective for dyads and small groups compared to groups of five or more learners.

When considering the limitations of the study, it can be helpful to consider the generalizability of our findings separately for conceptual versus procedural knowledge acquisition. For *conceptual knowledge*, elaborative learning processes are central to increase knowledge acquisition. As collaboration can particularly promote student elaboration, it is likely that different instructional materials and different settings will still yield similar results.

In contrast, several studies indicate that collaboration may not always hamper *procedural knowledge* acquisition. First, *variations to the task material* could increase positive effects of collaboration. For example, a study by Rittle-Johnson and Star (2007) revealed that individual learners can increase their procedural flexibility by comparing the effectiveness of different solution procedures; if two students engage in mutual elaboration when comparing different solution approaches, these positive effects may increase. Second, *collaboration training* or *support*, for instance through a collaboration script (e.g., Dillenbourg and Jermann 2007), could support positive effects of collaboration on procedural knowledge acquisition. Along these lines, Walker and colleagues investigated the effect of a peer tutoring script for learning literal equation solving. In a first study (Walker et al. 2009) they were not able to establish a positive effect of the script on students' learning outcome. However, in a follow-up study with improved script support (Walker et al. 2011), they found a positive script effect. The revised script comprised sophisticated adaptive collaboration support that encouraged peer tutors to explain tutee errors and to provide elaborative help. The results by Walker and colleagues suggest that collaboration support can promote procedural knowledge acquisition if it is successful at promoting the right types of interaction amongst students.

The generalizability of our results may also be influenced by the characteristics of our *learning setting*. Several researchers (e.g., Gweon et al. 2007; Lou et al. 2001) have hypothesized that corrective feedback as provided by our learning environment may eliminate positive effects of collaboration. Krause and Stark (2004) ascribe this effect to an "excess supply" of instruction: Receiving feedback by the learning partner is a major factor for the success of collaboration; if the feedback is already provided by the system, the feedback by the learning partner may no longer be necessary and elaborative meaning making processes may thus be reduced. In our study, the interface in the procedural conditions may have particularly provided such an excess supply due to the high level of support it provided (i.e., it contained a higher number of text boxes and more feedback opportunities per problem compared to the conceptual interface). It is possible that collaborative learning with procedural instructional material would have been more beneficial if no (or less) error feedback had been provided. However, it is important to note that research findings on the complex interaction of (computerized) feedback and collaboration are so far inconsistent, and final conclusions cannot yet be made. For instance, in contrast to the studies mentioned above, a study by Ellis et al. (1993) could only establish a positive effect of collaboration over individual learning when collaboration was combined with corrective feedback; however, when dyads did not receive corrective feedback, individuals and dyads reached comparable results.

While the aspects discussed in the previous sections point at limitations in the generalizability of our study results, several studies indicate that our results may be generalized to *other domains* such as physics. For instance, Jonassen (2003) has shown that the difficulties students encounter when solving story problems in physics are quite similar to their difficulties in mathematics. Often, students find it particularly challenging to understand the underlying concepts, while they are able to memorize equations and perform the correct problem-solving procedures. Along these lines, a study on learning in physics by Gadgil and Nokes (2009) revealed that collaborative learning with worked examples was particularly effective in improving conceptual understanding, while procedural fluency did not increase.

The results of our study have important methodological and practical implications. The *methodological implications* concern the question about which dependent variables provide valid conclusions on the success of collaborative learning. Researchers and teachers might often be tempted to evaluate collaboration based on group performance during collaboration as this is the first observable indicator for the success of a collaborative activity. However, as our results show, focussing merely on group performance may be misleading: Even though collaboration improved the group performance during the learning phase in both the conceptual and the procedural conditions, we only established a positive effect of collaboration on conceptual knowledge acquisition, while collaboratively practicing procedures did not increase procedural skill fluency. In contrast, the analysis of student activities during critical situations of the problem-solving process showed particularly valuable to indicate the success of collaboration in our study. We evaluated the quality of students' learning processes based on time variables and coding variables. While the coding variables are more meaningful and can thus yield a more detailed understanding of the dyadic learning processes that are responsible for the effectiveness of collaboration, the time variables are easy to assess and can even be analyzed “on-line”. Particularly the latter aspect can open up interesting opportunities for future research. For instance, the automatic analysis of the time variables may enable scientists to develop collaboration support that is adaptive to the dyad's needs (cf. Diziol et al. 2010a, c). As an example, it would be possible to automatically detect if students proceeded too quickly in error correction, and to subsequently encourage them to explain the error correction to their learning partner. This could reduce trial and error behavior and thus increase the effectiveness of collaborative learning with procedural instructional material.

Practical implications of our study concern guidelines for the implementation of collaboration in school settings. The results of our study show that it is crucial to increase teachers' awareness of the fact that collaborative performance does not necessarily yield improved individual learning outcomes, and to provide them with pedagogical knowledge of when and why collaboration can be beneficial (cf. Krauss et al. 2008). Particularly, knowledge on factors that influence the effectiveness of collaboration can help teachers to better match the learning setting they choose to the type of instructional material and the goals of instruction. Along these lines, our findings can help to support a teacher's decision on whether to implement an individual or a collaborative learning setting: If the goal is for students to acquire conceptual understanding, collaboration can be beneficial; if the goal is to support students' skill fluency, an individual learning setting may be superior. Furthermore, our study results provide valuable indicators for teachers to evaluate the success or failure of the collaborative activity: If students engage in mutual elaboration, they are on the right track; however, if the teacher observes a high amount of task distribution between students, he should intervene and encourage them to interact more.

Acknowledgements This research was supported by the Virtual PhD Programm, VGK (DFG) and by the Landesstiftung Baden-Württemberg. We are grateful to the CTAT team for their support in the development of our learning environment, and to the research group of Nikolaos Avouris that we could use their tool ActivityLens to analyze our data. Many thanks go to our student research assistants Jan Koch-Weser and Martina Rau for their help with the implementation of the tutoring problems, and to our student research assistants Stephanie Haug, Marlene Herrberg, Katharina Westermann, and Michael Wiedmann for their help during data collection and data analysis. Furthermore, we want to thank Lars Holzaepfel for his advice on the development of the study material from a mathematics education perspective.

References

Aleven, V., McLaren, B. M., Sewall, J., & Koedinger, K. R. (2009). A new paradigm for intelligent tutoring systems: Example-tracing tutors. *International Journal of Artificial Intelligence and Education, 19*(2), 105–154.

Anderson, J. R. (1983). *The architecture of cognition*. Cambridge: Harvard University Press.

Avouris, N., Fiotakis, G., Kahrimanis, G., Margaritis, M., & Komis, V. (2007). Beyond logging of fingertip actions: analysis of collaborative learning using multiple sources of data. *Journal of Interactive Learning Research JILR, 18*(2), 231–250.

Baker, R. S., Corbett, A. T., & Koedinger, K. R. (2004). *Detecting student misuse of intelligent tutoring systems.* Paper presented at the Proceedings of the 7th International Conference on Intelligent Tutoring Systems.

Berg, K. F. (1994). *Scripted cooperation in high school mathematics: Peer interaction and achievement.* Paper presented at the Annual meeting of the American Educational Research Association, New Orleans, Louisana.

Brenner, M. E., Mayer, R. E., Moseley, B., Brar, T., Durán, R., Reed, B. S., et al. (1997). Learning by understanding: The role of multiple representations in learning algebra. *American Educational Research Journal, 34*(4), 663–689.

Chi, M. T. H. (1997). Quantifying qualitative analyses of verbal data: A practical guide. *The Journal of the Learning Sciences, 6*(3), 271–315.

Cress, U. (2008). The need for considering multi-level analysis in CSCL research. An appeal for the use of more advanced statistical methods. *International Journal of Computer-Supported Collaborative Learning, 3*(1), 69–84.

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. 275–301). New York: Springer.

Diziol, D., Rummel, N., Spada, H., & McLaren, B. (2007). Promoting learning in mathematics: Script support for collaborative problem solving with the Cognitive Tutor Algebra. In C. A. Chinn, G. Erkens & S. Puntambekar (Eds.), *Mice, minds and society. Proceedings of the Computer Supported Collaborative Learning (CSCL) Conference 2007, Vol 8, I* (pp. 39–41). International Society of the Learning Sciences.

Diziol, D., Rummel, N., & Spada, H. (2009). Procedural and conceptual knowledge acquisition in mathematics: Where is collaboration helpful? In C. O'Malley, D. Suthers, P. Reimann, & A. Dimitracopoulou (Eds.), *Computer Supported Collaborative Learning Practices—CSCL2009 Conference Proceedings, ISLS* (pp. 178–187). International Society of the Learning Sciences.

Diziol, D., Rummel, N., & Spada, H. (2010). *Why is collaboration not always effective? Insights from the analysis of students' learning trajectories.* Paper presented at the ISEE workshop at the 9th International Conference of the Learning Sciences. Chicago, IL, USA.

Diziol, D., Rummel, N., Spada, H., & Haug, S. (2010b). Learning in mathematics: Effects of procedural and conceptual instruction on the quality of student interaction. In K. Gomez, L. Lyons, & J.

Radinsky (Eds.), *Learning in the disciplines: Proceedings of the 9th International Conference of the Learning Sciences (ICLS 2010)* (Vol. 2, pp. 370–371). Chicago: International Society of the Learning Sciences.

Diziol, D., Walker, E., Rummel, N., & Koedinger, K. (2010c). Using intelligent tutor technology to implement adaptive support for student collaboration. *Educational Psychology Review, 22*(1), 89–102.

Ellis, S., Klahr, D., & Siegler, R. (1993). *Effects of feedback and collaboration on changes in children's use of mathematical rules.* Paper presented at the meeting of the Society for Research in Child Development, New Orleans.

Ericsson, K. A. (2003). Valid and non-reactive verbalization of thoughts during performance of tasks. *Journal of Consciousness Studies, 10*(9–10), 1–18.

Gadgil, S., & Nokes, T. (2009). Analogical scaffolding in collaborative learning. In N. A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31th Annual Conference of the Cognitive Science Society* (pp. 3115–3120). Austin: Cognitive Science Society.

Gweon, G., Rosé, C. P., Albright, E., & Cui, Y. (2007). Evaluating the effect of feedback from a CSCL problem solving environment on learning, interaction, and perceived interdependence. In C. A. Chinn, G. Erkens, & S. Puntambekar (Eds.), *Mice, minds and society. Proceedings of the Computer Supported Collaborative Learning (CSCL) Conference 2007* (pp. 231–240). International Society of the Learning Sciences.

Hausmann, R. G. M., Chi, M. T. H., & Roy, M. (2004). Learning from collaborative problem solving: An analysis of three hypothesized mechanisms. In K. D. Forbus, D. Gentner, & T. Regier (Eds.), *26nd Annual Conference of the Cognitive Science Society* (pp. 547–552). Mahwah: Lawrence.

Hiebert, J., & Wearne, D. (1996). Instruction, understanding, and skill in multidigit addition and subtraction. *Cognition and Instruction, 14*(3), 251–283.

Jonassen, D. H. (2003). Designing research-based instruction for story-problems. *Educational Psychology Review, 15*(3), 267–296.

Kenny, D. A., Kashy, D. A., & Bolger, N. (1998). Data analysis in social psychology. In D. Gilbert, S. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (4th ed., Vol. 1, pp. 233–265). Boston: McGraw-Hill.

KMK. (2004). *Bildungsstandards im Fach Mathematik für den Mittleren Schulabschluss (Jahrgangsstufe 10) [Standards for mathematics education for junior high school graduation, grade 10]*. Darmstadt: Betz Druck.

Koedinger, K. R., Anderson, J. R., Hadley, W. H., & Mark, M. A. (1997). Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education, 8*, 30–43.

Krause, U.-M., & Stark, R. (2004). Too much of a good thing? Unwanted side effects of successful instructional interventions. In P. Gerjets, P. A. Kirschner, J. Elen, & R. Joiner (Eds.), *Instructional design for effective and enjoyable computer-supported learning. Proceedings of the First Joint Meeting of the EARLI SIGs Instructional Design and Learning and Instruction with Computers (CD-ROM)*. Knowledge Media Research Center: Tuebingen.

Krauss, S., Brunner, M., Kunter, M., Baumert, J., Blum, W., Neubrand, M., et al. (2008). Pedagogical content knowledge and content knowledge of secondary mathematics teachers. *Journal of Educational Psychology, 100*, 716–725.

Lou, Y., Abrami, P. C., Spence, J. C., Poulsen, C., Chambers, B., & d'Apollonia, S. (1996). Within-class grouping: A meta-analysis. *Review of Educational Research, 66*(4), 423–458.

Lou, Y., Abrami, P. C., & d'Apollonia, S. (2001). Small group and individual learning with technology: A meta-analysis. *Review of Educational Research, 71*(3), 449–521.

Mevarech, Z. R., & Stern, E. (1997). Interaction between knowledge and contexts on understanding abstract mathematical concepts. *Journal of Experimental Child Psychology, 65*, 68–95.

Nathan, M. J., Kintsch, W., & Young, E. (1992). A theory of algebra-word-problem comprehension and its implications for the design of learning environments. *Cognition and Instruction, 9*(4), 329–389.

Nathan, M. J., Mertz, K., & Ryan, R. (1994). *Learning through self-explanation of mathematics examples: Effects of cognitive load.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

NCTM (2000). *Principals and standards for school mathematics.* Reston, VA.

Nokes, T. J., & Ross, B. H. (2007). Facilitating conceptual learning through analogy and explanation. In L. Hsu, C. Henderson, & L. McCullough (Eds.), *2007 Physics Education Conference* (pp. 7–10). American Institute of Physics Conference Proceedings.

O'Donnell, A. M. (1999). Structuring dyadic interaction through scripted cooperation. In A. M. O'Donnell & A. King (Eds.), *Cognitive perspectives on peer learning* (pp. 179–196). Lawrence Erlbaum Associates, Publishers.

Perry, M. (1991). Learning and transfer: Instructional conditions and conceptual change. *Cognitive Development, 6*, 449–468.

Reimann, P. (2007). Time is precious: Why process analysis is essential for CSCL (and can also help to bridge between experimental and descriptive methods. In C. A. Chinn, G. Erkens & S. Puntambekar

(Eds.), *Mice, minds and society. Proceedings of the Computer Supported Collaborative Learning (CSCL) Conference 2007, Vol 8,* II (pp. 590–607). International Society of the Learning Sciences.

Renkl, A. (2008). Kooperatives Lernen. In W. Schneider & M. Hasselhorn (Hrsg.). *Handbuch Psychologie, Bd. Pädagogische Psychologie*, S. 84–94. Göttingen, Hogrefe

Rittle-Johnson, B. (2006). Promoting transfer: Effects of self-explanation and direct instruction. *Child Development, 77*(1), 1–15.

Rittle-Johnson, B., & Alibali, M. W. (1999). Conceptual and procedural knowledge of mathematics: Does one lead to the other? *Journal of Educational Psychology, 91*(1), 175–189.

Rittle-Johnson, B., & Star, J. R. (2007). Does comparing solution methods facilitate conceptual and procedural knowledge? An experimental study on learning to solve equations. *Journal of Educational Psychology, 99*(3), 561–574.

Rittle-Johnson, B., Siegler, R. S., & Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: An interative process. *Journal of Educational Psychology, 93*(2), 346–362.

Slavin, R. E. (1996). Research on cooperative learning and achievement: What we know, what we need to know. *Contemporary Educational Psychology, 21*(1), 43–69.

Souvignier, E., & Kronenberger, J. (2007). Cooperative learning in third graders' jigsaw groups for mathematics and science with and without questioning training. *The British Journal of Educational Psychology, 77*, 755–771.

Staub, F. C., & Reusser, K. (1995). The role of presentational structures in understanding and solving mathematical word problems. In C. A. Weaver, S. Mannes, & C. R. Fletcher (Eds.), *Discourse comprehension. Essays in honor of Walter Kintsch* (pp. 285–305). Hillsdale: Lawrence.

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

Walker, E., Rummel, N., & Koedinger, K. R. (2011). *Adaptive support for CSCL: Is it feedback relevance or increased accountability that matters?* Paper to be presented at the 10th International Conference on Computer-Supported Collaborative Learning. Q1

Walker, E., Rummel, N., & Koedinger, K. R. (2009). Integrating collaboration and intelligent tutoring data in the evaluation of a reciprocal peer tutoring environment. *Research and Practice in Technology Enhanced Learning, 4*(3), 221–251.

Webb, N. M. (1989). Peer interaction and learning in small groups. *International Journal of Education Research, 13*, 21–39.