 JrnlID 11412_ArtID 9146_Proof# 1 - 28/04/2012

Computer-Supported Collaborative Learning
DOI 10.1007/s11412-012-9146-z

Scripted collaborative learning with the cognitive tutor algebra

Nikol Rummel · Dejana Mullins · Hans Spada

Received: 24 December 2010 / Accepted: 24 April 2012

Abstract With the aim to promote students' mathematics learning, we extended the Cognitive Tutor Algebra (CTA), a computer-based tutoring system for high school mathematics, to a collaborative setting. Furthermore we developed a collaboration script to support students' interactions. In an experimental classroom study, we compared three conditions: scripted collaborative learning, unscripted collaborative learning, and individual learning. After a 2-day learning phase, posttests assessed individual and collaborative reproduction of knowledge and skills, and future learning. First, with the collaboration script we aimed to improve students' interaction. Second, we assumed that due to an improved interaction students would benefit more from the learning opportunities during collaboration and, in consequence, their learning would increase as compared with the other conditions. To investigate the first assumption, we compared the interaction of a scripted dyad and an unscripted dyad. The in-depth process analyses revealed a positive impact of the script on student collaboration and problem solving during scripted interaction and in subsequent unscripted interaction. While this effect was mirrored in the learning gains of the two dyads, we could not establish a general learning effect in the quantitative between-condition comparison of student performance. Particularly for students with low prior knowledge, the removal of the script in the test phase initially entailed a decline in reproduction performance as students had to get used to the unscripted problem-solving situation. A notable finding was, however, that the collaborative conditions yielded the same outcomes as the individual condition in the individual reproduction test even though students had solved fewer problems during the learning phase and had only solved them collaboratively.

N. Rummel (✉)
Institute of Educational Research, RuhrUniversität Bochum, Universitätsstraße 150, 44801 Bochum, Germany
Q3
e-mail: nikol.rummel@rub.de

D. Mullins · H. Spada
Institute of Psychology, Albert-Ludwigs-Universität Freiburg, Freiburg im Breisgau, Germany

Keywords Computer-supported collaborative learning · Scripted collaboration · Mathematics learning · Experimental classroom study · Collaboration process analysis

Introduction

Interest in developing improved methods for mathematics instruction has increased since TIMSS (Third International Mathematics and Science Study) and PISA (Programme for International Student Assessment). There is broad agreement that the goal of instruction should go beyond improving students' solving of tasks where they can apply well-practiced procedures. Instead, school education should aim to equip students with competencies that prepare them for the challenges of their future life (Organisation for Economic Co-operation and Development [OECD] n.d.). According to the OECD, one of the most important competencies to be achieved in school is "Mathematical Literacy". In order to improve mathematics instruction and to support the development of students' mathematical literacy, different instructional approaches have been investigated (e.g., Dubinsky et al. 1997). One approach that is consistent with the curriculum recommendations from the National Council of Teachers of Mathematics (NCTM 2006), and that has proven effective for increasing students' learning of mathematics, is learning with cognitive tutors as, for example, developed by Anderson and colleagues (Anderson et al. 1995; Koedinger et al. 1997). Cognitive tutors present students with real-world tasks and adaptively support their problem-solving by providing just-in-time feedback and offering on-demand hints. Although cognitive tutors have repeatedly been shown to increase learning outcomes, they also have been criticized for facilitating shallow learning strategies (e.g. Aleven et al. 2004). For instance, students have been found to abuse hints given in the tutoring environment by merely copying the answers, instead of elaborating on the hints (Aleven et al. 2004). Also, students have been found to game the system, that is, they systematically exploit regularities in the software to perform well and to advance faster in the cognitive tutor curriculum (Baker et al. 2004). In consequence of such behaviors, a deeper understanding of underlying mathematical concepts and robust mathematical skills are not necessarily achieved. Against this background, we propose to extend cognitive tutors with scripted collaboration to promote students' elaborative sense-making activities, with the hope to yield better learning results and, ultimately, improved mathematical literacy. In the present study we evaluated collaborative extensions to an existing cognitive tutor, the Cognitive Tutor Algebra (© Carnegie Learning Inc.).

As research has shown, collaborative problem solving and learning have the potential to promote deeper elaboration of the learning content (Teasley 1995) and can yield improved conceptual understanding. In collaborative learning, the *process* is of central importance (e.g., Reimann 2007). According to the "interaction paradigm" (Dillenbourg et al. 1996), the interaction among students is the mediating variable that determines whether collaboration will yield effects on their learning outcome. Collaborative behaviors that account for the beneficial impact of collaboration are, for instance, giving and receiving explanations and joint knowledge construction (Hausmann et al. 2004; Rummel and Spada 2005; Meier et al. 2007). These mechanisms can lead to important opportunities for learning in collaborative settings, however only if they occur and if students take advantage of them. Unfortunately, students often do not show fruitful collaborative behaviors spontaneously, but need support (Rummel and Spada 2005). Two aspects that can be regarded as preconditions for a fruitful interaction are the flow of the collaboration and the motivation of the collaborating partners.

Collaboration flow refers to the degree to which students' actions and utterances build on each other and whether they maintain a joint focus on the task they are solving (Rummel et al. 2011). *Motivation* of the collaborating partners is indicated by students' attitude towards the collaboration and their commitment to the joint task (Meier et al. 2007). For students to benefit from the collaboration, it is crucial that they participate actively in the interaction—be it in a symmetrical relationship, or in complementary roles such as tutor and tutee (e.g., O'Donnell 1999; Slavin 1996). A related problem frequently reported is *unequal contribution* of the collaborating partners to the problem solving process as they do not feel mutually responsible for the collaborative outcome; a phenomenon that most often harms both learning partners (e.g., O'Donnell 1999; Slavin 1996): If the interaction is characterized by one student telling his or her partner what to do, and the other student is following the instructions without understanding why, the latter student will presumably fail to acquire a deeper understanding (Webb et al. 1995). At the same time, this eliminates any possibility for the learning partner to profit from the collaborative learning setting through giving or receiving help and joint knowledge construction.

One approach that has shown to be effective in fostering collaboration, also particularly in mathematics, is to provide guidance by means of a collaboration script (e.g., Berg 1993, 1994; King 2007; O'Donnell 1999; for an overview see Kollar et al. 2006). Collaboration scripts guide the learning partners through a sequence of interaction phases with designated activities and roles (O'Donnell 1999) and thus promote particular cognitive, metacognitive and social processes conducive to learning (King 2007). For instance, in a jigsaw script (Aronson et al. 1978; Dillenbourg and Jermann 2007) knowledge or materials relevant to solving the task at hand is distributed between the learning partners. Distributing expertise in this way has been shown to strengthen students' individual accountability for the collaborative task, thus leading to better, more engaged interactions, and promoting learning (Dillenbourg and Jermann 2007; Slavin 1992). Moreover, it has been demonstrated that scripts can serve as model for future collaborations (Rummel and Spada 2005, 2007).

In the current study, we therefore developed a collaboration script with two goals (cf. Dillenbourg and Jermann 2007; Rummel and Spada 2007): first, to support student interaction while working with the script and thus improve their learning (script as method; effects *of* the script); and, second, to improve students' collaboration skills, yielding fruitful collaborative behavior even when script support is no longer available (script as objective; effects *with* the script). The effects *with* the script should then help students to successfully tackle new tasks in a future collaborative learning situation (cf. Bransford and Schwartz 1999: preparation for future learning).

A potential pitfall of scripting collaboration is to "over-script" students that may already have enough collaboration skills (Dillenbourg 2002; Kollar et al. 2007). If the goal is for students to internalize the scripted behavior and to apply it even when script support is no longer available, then scripting could be ceased after some scripted collaboration (e.g., Rummel and Spada 2005, 2007) or faded out over time (Wecker et al. 2010). However, this is still no solution if script support was obsolete from the beginning. Also, it does not help in situations where students are "under-scripted" and would need more support than the script is providing. A promising idea is therefore to support students' collaboration in an adaptive fashion, tailored to their individual and changing needs for support. Intelligent tutoring technologies open a new horizon with regard to adaptive tutoring of collaboration. As Walker and colleagues (2009a, b, 2010, 2011; see also Diziol et al. 2010) have shown, the technology that is used by cognitive tutors to provide just-in-time adaptive support for domain learning can also be applied to provide just-in-time adaptive support for collaboration, that is, to prompt fruitful collaborative behaviors in relevant moments of the interaction. The work presented in the current paper is related to the work by Walker and colleagues as our collaboration script also built on the Cognitive Tutor Algebra and included some adaptive script elements.

Research questions and hypotheses

In the introduction, we described the risk that students might solve tasks within a cognitive tutoring system without acquiring deeper conceptual understanding. We discussed the potential of collaborative learning to increase students' elaboration of the learning material and yield improved learning outcomes. We argued that support is needed to ensure that students tap the potentials of a collaborative learning setting, and introduced collaboration scripts as a promising way to promote collaboration. Finally, we discussed the possibility of leveraging existing intelligent tutor technology to provide adaptive scripting of collaboration.

Against this background we developed collaborative extensions to the Cognitive Tutor Algebra (CTA), an established cognitive tutoring system for mathematics instruction at the high school level (e.g., Koedinger et al. 1997), and implemented a collaboration script to support students' collaborative learning with the system. To evaluate the effects of our collaborative script extensions to the CTA, we conducted an in vivo study, that is, a controlled classroom experiment. In the study we compared collaborative learning with script support *(scripted condition)* to collaborative learning without script support *(unscripted condition)* and individual learning *(individual condition)*. All three conditions were implemented within the CTA. After a 2 day learning phase we administered three posttests assessing individual and collaborative reproduction, and future learning.

Which effects did we expect from scripted collaborative learning? With the collaboration script we aimed to improve student interaction. As was argued above, it is through the interaction with their peers that students' understanding develops in a collaborative setting. Thus, we assumed that due to an improved interaction students would benefit more from the learning opportunities during collaboration and, in consequence, their learning would be increased. To investigate how the script influenced student interaction, we first conducted in-depth process analyses of two case studies (one dyad from the unscripted condition and one dyad from the scripted condition). More specifically, we looked at how the collaboration script influenced the quality of student interaction *during the learning phase*, that is, during scripted problem solving. Furthermore, we investigated how scripted practice during the learning phase related to the quality of student interaction *during subsequent, unscripted problem solving in the test phase*. And finally, we checked whether the interaction quality of the selected dyads was mirrored in their learning outcomes. In our process analyses we assessed the quality of the collaboration analogous to process analyses we had conducted in previous studies (Meier et al. 2007; Rummel et al. 2011). As the goal of the current study was to promote learning in mathematics, we additionally evaluated students' problem-solving during particularly challenging problem-solving steps.

In a second step we statistically compared the *learning outcomes* across all three experimental conditions in order to evaluate how collaboration, and especially scripted collaboration, affected learning. We expected to find the following effects: The mechanisms of collaborative learning were expected to lead to deeper learning particularly in the scripted condition, and thus to yield improved mathematical skill fluency as measured by our reproduction posttests. We furthermore hypothesized that the learning effect would carry over from collaborative to individual performance; that is, we were also expecting better performance of the collaborative conditions, and particularly the scripted condition, on the individual reproduction posttest. This would be an important effect, taking into account that school assessment is primarily based on the evaluation of individual performance. Finally, we assumed that scripted students would have learned to take advantage of the collaborative learning setting, and that this ability would help them to tackle new learning content consecutively; thus they should perform better than the other conditions on a future learning posttest assessing their performance on new learning content.

Method

Before we describe the study design and procedure in more detail, we briefly introduce the cognitive tutoring system that we employed in our study and the curriculum unit that we used as learning material, and we describe the collaboration script we developed.

Learning environment and material

The Cognitive Tutor Algebra (CTA) is a tutoring software for high school instruction used in over 2000 schools across the USA. As several studies have shown, learning with the CTA improves student performance by about one standard deviation compared to traditional classroom instruction on measures of algebra understanding (Koedinger et al. 1997, 2000). The CTA comprises 32 different curriculum units that cover the learning content of algebra I. It consists of several tools, and depending on the unit, some or all of them are

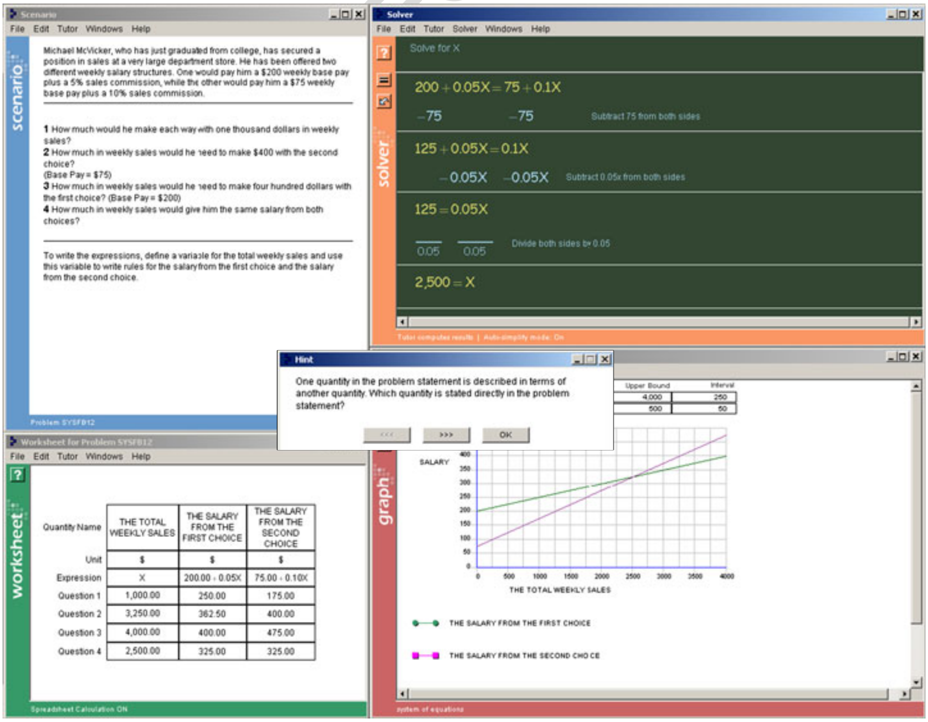

Fig. 1 Screenshot of the Cognitive Tutor Algebra, unit system of equations

displayed. Figure 1 shows a screenshot of the CTA from the unit system of equations (unit 13). This was the unit we used for the learning material in our study. Our participants had not yet been introduced to the system-of-equations concept in their classroom instruction.

In unit 13 of the CTA, the *Problem Scenario* (top left corner) shows a story problem with several questions. The story problems use concrete, real-world scenarios (for instance, in the example shown in Fig. 1, students have to compare two salary structures that were offered to Michael McVicker). Students are requested to find the *y*-values for a given *x*-value or the *x*-value corresponding to a given *y*-value, respectively. For instance, in question 1 of the example task, the weekly sales are given, and students have to find the resulting income for the two salary structures; in questions 2 and 3, students are told about McVicker's income and have to find the weekly sales he must have made. These types of questions are structurally similar to the questions in the unit linear equations (unit 7 of the CTA), which our participants were already familiar with (in the following, we will therefore refer to these questions using the term *simple questions*). One question is new in unit 13 and was thus particularly challenging for students participating in the study: the question of how to find the *intersection point* (i.e. question 4 in Fig. 1). Prior to answering this question students are additionally required to construct a graph of the problem situation.

In summary, when solving a system-of-equations problem such as the one in Fig. 1 with the CTA, students are required to perform the following steps (see Table 1): First, students label the columns of the *Worksheet* (see Fig. 1 bottom left) according to the entities described in the problem, enter the appropriate units and derive the algebraic expressions (step deriving expressions). Then they work on solving the questions of the story problem (step solving simple questions, step graphing, and step finding intersection point) making use of the help facilities of the CTA. The *Solver* window (see Fig. 1 top right) enables students to solve equations. To construct the graph of the problem situation in the *Grapher* window (see Fig. 1 bottom right), students first have to label the axes, set the appropriate bounds and intervals so that all points of the Worksheet can be plotted, and finally graph the lines (step graphing). The *Hint* window in the middle of the screen in Fig. 1 on top of the other windows gives an example for the hint messages the CTA provides on demand and when students make errors. In the hint window students can click on the arrow button to receive more detailed hints. The final hint tells them the answer to the current problem-solving step.

Table 1 Problem-solving steps (system-of-equation problems)

Steps	Students' tasks
Deriving expressions	label columns of Worksheet
	enter units
	derive algebraic expressions from the story problem and enter in Worksheet
Solving simple questions	solve questions 1 through 3 with help of the Solver tool (note: questions are structurally equivalent to questions from a previous CTA unit on linear equations (unit 7))
	enter solutions in Worksheet
Graphing	label axes
	set bounds and intervals
	graph the lines
Finding intersection point	equate the two expressions in the Solver tool
	solve the resulting equation for x
	enter solution in Worksheet

In addition to the hints, the CTA provides just-in-time feedback by marking student errors red. Students insert the answers to the questions of the story problem in the corresponding cells of the worksheet.

The school that participated in our study uses the CTA curriculum in their regular mathematics instruction. In classroom courses following the CTA curriculum, three of five course periods a week are classroom lessons; during the remaining two periods, students work on the CTA in the computer lab (Koedinger 1998, Koedinger et al. 1997). Therefore our study participants were well-acquainted with the CTA functionality and were used to learning with this software. This is important to note as often initial positive or negative effects of computer-based learning environments have to be ascribed to the novelty of the environment to students.

A collaboration script for solving problems on the cognitive tutor algebra

We developed a collaboration script that supported students as they collaboratively learned to solve system-of-equations problems using the CTA. The script (see Fig. 2) employed a jigsaw schema (Aronson et al. 1978; Dillenbourg and Jermann 2007) as general framework; in other words, it distributed the responsibility for the story problem between the learning partners: During an *individual phase*, each student solved questions containing one linear equation in the CTA; during the following *collaborative phase*, students joined on a single computer to solve questions combining the two linear equations into a system-of-equations problem. For the system-of-equations problems, students were prompted to take responsibility for problem steps relating to their individual expertise (e.g., they explained to their partner how to derive the equation corresponding to their part of the story problem and were responsible to answer the simple questions corresponding to their problem part). Then they were asked to jointly solve the step pertaining to the new problem type: finding the intersection point. The individual and collaborative phases were repeated for each story problem students solved while working on the CTA. The script was directly implemented in the CTA software.

The jigsaw framework already provided a setup that has been shown to promote fruitful collaboration by increasing learners' individual accountability. In order to further support students' individual accountability, the interaction was additionally supported by *fixed script elements* that prompted particular collaborative behaviors and allocated roles. Based on the task structure, the collaborative problem solving process was divided into several steps. A short instruction preceded each step, prompting students to engage in particular collaborative behaviors. For instance, at steps where students had to contribute their individual expertise the responsibilities were marked by color coding and students were told to alternate between

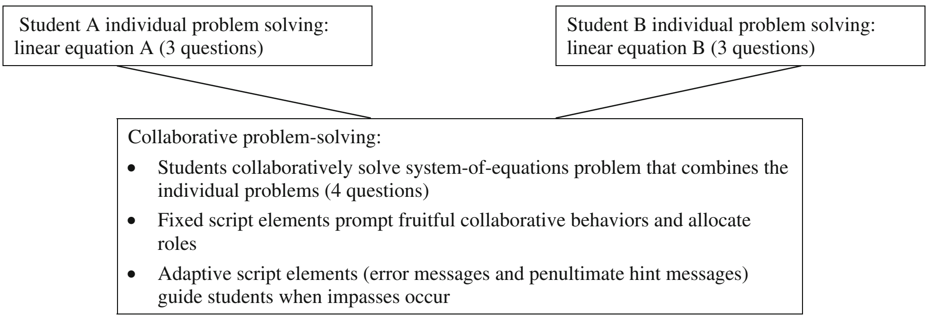

Fig. 2 Design of the collaboration script: Jigsaw schema with integrated additional script elements

the roles of explainer and listener. The explainer was prompted to give elaborated explanations while the listener was prompted to ask for further explanation when having problems in understanding.

In the introduction, we have discussed adaptive scripting as one possible solution to avoid providing too little support or over-scripting collaboration. Following this argumentation we additionally implemented *adaptive elements* in our collaboration script in order to counteract problematic student behaviors reported in the literature on learning with cognitive tutors (e.g. trial and error, hint abuse, gaming behavior): An *error message* popped up when dyads made an error. It prompted students to learn from the error by mutually reflecting on their problem solving process or by requesting a hint from the CTA. The error message aimed at reducing gaming behavior and at increasing the amount of expedient help requests. Second, when students engaged in hint abuse, that is, when they clicked on the hint widget repeatedly in order to receive the bottom-out hint, a *penultimate hint message* appeared (see Fig. 3). It prompted students to mutually elaborate on the hints received so far and to try to find the answer on their own and thus learn for future problem solving.

Study design and procedure

The study took place during five class periods over the course of a week: a single period on day 1, and two block periods on days 2 and 3 (see Table 2). The first minutes at the beginning of each period were used for organizational purposes: on day 1, students received a short introduction to their condition; on day 2 and 3, teachers rearranged dyads if one partner was missing (see explanation in the participants section).

On days 1 and 2 (*learning phase*), students solved a system-of-equations problem according to their condition working at their own pace. In the *scripted condition* the dyads' interaction was structured by our collaboration script. As described, the script guided students to alternate between individual and collaborative work phases while solving

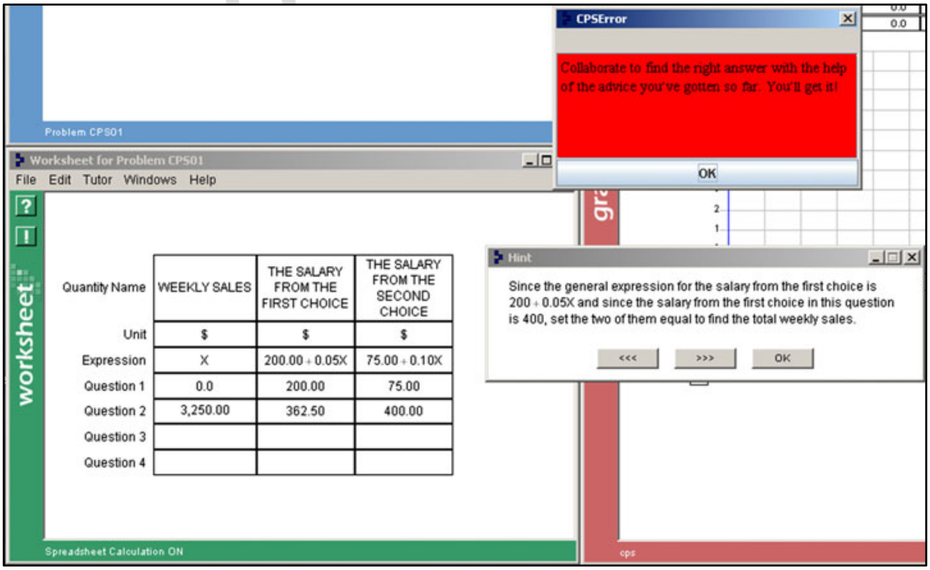

Fig. 3 Screenshot of adaptive hint prompt

Table 2 Study design and procedure

	Scripted collaboration	Unscripted collaboration	Individual learning
Learning phase	Day 1 (single period)		
		Short introduction	
	Scripted collaborative problem-solving on the CTA[a]	Collaborative problem-solving on the CTA	Individual problem-solving on the CTA
	Day 2 (block period)		
	Regrouping of dyads with missing learning partner		
	Scripted collaborative problem-solving on the CTA (students continue with problem from previous day)	Collaborative problem-solving on the CTA (students continue with problem from previous day)	Individual problem-solving on the CTA (students continue with problem from previous day)
Test phase	Day 3 (block period)		
	Regrouping of dyads with missing learning partner		
	Condition-specific reproduction test: collaborative problem-solving (CTA)		Condition-specific reproduction test: individual problem-solving (CTA)
	Future learning test: collaborative problem-solving (CTA)		Future learning test: individual problem-solving (CTA)
	Individual reproduction test: individual problem-solving (CTA)		

[a] *CTA* cognitive tutor algebra

problems with the CTA and adaptively supported them during their collaboration. During the individual work phases students worked on separate computers; for the collaborative phases they joined on one computer. In the *unscripted condition*, two students joined on one computer to collaboratively solve problems with the CTA, but did not receive specific support for their collaboration. This condition corresponds to the way collaborative learning is often implemented in classrooms: students are simply put together in small groups to work on certain tasks; however, without support, they might fail to take advantage of the collaborative setting. The *individual condition* served as an ecological control condition corresponding to current practice in the CTA curriculum: students individually solved problems with the CTA.

In all conditions, the problems that students solved consisted of seven questions: six introductory linear equations questions (corresponding to the simple questions in Fig. 1 and Table 1), followed by one question targeting the system-of-equations concept. This seventh and last question asked students to compute the intersection point. Students in the scripted condition answered three of the linear equations questions during the individual phase and the remaining four questions, including the intersection point question, during the collaborative phase. Learning time was kept constant across conditions. Students worked at their own pace, solving problems until time was up. Students in all conditions worked on the same problems. Their problem-solving was supported by the CTA, which provided immediate feedback and hints in its regular fashion, as described.

On day 3 (*test phase*), three posttests were administered to evaluate the effects of the experimental conditions on the learning outcomes. Students first solved a *condition-specific reproduction* test and a test assessing *future learning*. These tests were solved collaboratively in the collaborative conditions and individually in the individual condition. Next, all participants solved an *individual reproduction* test; this test was solved individually in all conditions. All three tests took place on the computer within the CTA.

Participants

The current study was conducted as an in vivo *experiment* at one of the LearnLab research facilities of the Pittsburgh Science of Learning Center (PSLC, http://learnlab.org). Five teachers agreed to host the study in their algebra classes (eight classes and 139 students in total). Parents were asked to give informed consent for their children's participation. To guarantee student anonymity, each student received a fictitious name that was used to identify the student throughout the study. These names were used as logins for the CTA.

To prevent internal validity threats such as treatment diffusion, the study was conducted in a *between-classroom design*. The participating eight classes were randomly assigned to conditions, taking into account the following preconditions: classes taught by the same teacher were assigned to different conditions, and each condition was supposed to consist of a comparable number of students or dyads respectively. In both collaborative conditions, teachers assigned students to homogenous dyads based on their math grade, making sure to pair students that got along well. In our statistical analyses we took care to control for differences in prior knowledge that may have resulted from the between-classroom design.

The school that participated in our study is a vocational high school: for half of the day, students attend regular classes in different grades at their home schools; the other half of the day, they attend the vocational school to take part in instructional program courses (e.g., carpentry and culinary arts) and "basics" courses, such as mathematics. In a pre study conducted at the same school, we realized that—due to the specific school format—the rate of student absenteeism was quite high (Diziol et al. 2007) In order to decrease the loss of data that would result from excluding both learning partners if one student was missing, students were regrouped at the beginning of each day when necessary. Regrouping rules guided teachers' decisions when forming new dyads, ensuring that all teachers dealt with this issue in a similar way. Conditions did not differ in the rate of student attrition (χ^2=.75, p=.69). To ensure a high ecological validity, we included as many students as possible in our data analyses: we included students that remained in the same condition throughout the study, that participated in at least 1 day of the learning phase, and that were present on the test day. These conditions were met by about three quarters of the sample. The sample of students included in final data analyses consisted of 106 students, 74 boys and 31 girls. Information about the gender of one student was missing. The average age of students was 15.86 (SD=.74), their average school grade was 9.88 (SD=.43). Due to technical difficulties during the test day, test data was lost for a differing number of students. The resulting sample sizes for the different posttests can be found in Table 3.

Analysis of the collaboration process and the learning outcome

We analyzed the effects of scripted collaborative learning with the CTA in two steps: First, we conducted analyses of the collaboration process of two dyads (one from the scripted and one from the unscripted condition). The analyses were done using two rating schemes and a narrative approach. Second, we statistically compared the learning outcomes of all

Table 3 Number of participants included in data analysis of the three post tests

	Scripted	Unscripted	Individual
Condition-specific reproduction	18 dyads	19 dyads	16 individuals
Future learning	23 dyads	19 dyads	16 individuals
Individual reproduction	38 individuals	39 individuals	17 individuals

participants across the three conditions based on the posttest data. Table 4 gives an overview of the dependent variables that are explained in more detail in the following two sections.

Analysis of the collaboration process

We recorded student interaction during the learning phase and during the collaborative reproduction posttest. A screen capture tool launched automatically when students started the CTA and stopped when students quit the software. The tool recorded students' verbal interaction and their actions on the computer screen. For the analysis, we integrated the screen recordings (audio-video data) with log data from the CTA using ActivityLens, a software program for the collaboration process analysis developed by Avouris and colleagues (Avouris et al. 2007). The integration of the different data sources enabled us to segment the interaction based on the task structure and to navigate to particularly interesting collaboration sequences (e.g., interaction after hint requests or errors) based on the log data. We used ActivityLens both for the rating analyses and for the narrative analysis approach.

We developed two *rating schemes* that assessed the quality of student interaction from two perspectives. Table 5 provides an overview of all rating dimensions with examples for high and low ratings. Ratings were done on a five-point rating scale ranging from 0 (*very bad*) to 4 (*very good*). In addition, the second rating scheme included a variable evaluating the dyad's overall problem-solving strategy according to five distinct categories; this variable is shown in the last row of Table 5.

The first rating scheme focused on the quality of the collaborative behavior in more general terms; here we assessed the interaction process throughout the solving of entire problems (i.e. across all problem-solving steps, see Table 1). The dimensions for analyzing the *quality of collaboration* were adapted from a rating scheme that we had developed and evaluated in earlier research (Meier et al. 2007; Rummel et al. 2011). The dimension

Table 4 Overview of the dependent variables

Collaboration process	Rating schemes	Quality of the collaboration
		Quality of the problem-solving process
	Narrative approach	Actions and interactions
Learning outcomes (Posttests)	Condition-specific reproduction	Error rate
		Assistance score
	Future learning	Error rate
		Assistance score
	Individual reproduction	Error rate
		Assistance score

Table 5 Examples for low and high ratings of interaction quality

Dimension	Examples *very bad* (0 points)	Examples *very good* (4 points)
Quality of the collaboration (rating scheme 1)		
Collaboration flow	there is little or no talk	partners communicate coherently and monitor each other's attention and understanding
Collaborative motivation	partners show a negative attitude towards the interaction/the task; there is an unequal contribution to the problem-solving process	partners show a highly positive attitude towards the joint problem-solving; both partners are actively involved in the interaction
Elaboration on content	there is little or no talk; partners talk about irrelevant topics (off-topic conversation)	partners give explanations of their actions/ proposals and make references to mathematical concepts
Elaboration on hint	partners do not read the hints, but immediately ask for the next hint	partners mutually discuss the hints in order to learn from them
Quality of the problem-solving process (rating scheme 2)		
Mathematical understanding	partners need a lot of CTA assistance to solve the steps and show no understanding of the correction	partners solve step correctly on first attempt, revealing a deep understanding of the underlying principles
Capitalization on social resource	partners ignore each other's presence and joint potential to find a solution	partners make proposals how to derive the correct solution and discuss them together
Capitalization on system resource	partners engage in trial-and-error or hint abuse	partners mutually reflect on errors and hints
Dyad's strategy	(0) trial and error	
	(1) hint abuse	
	(2) immediate correction	
	(3) (proposal-) correct input	
	(4) elaborating with partner	

collaboration flow assessed whether students were responsive to each other's actions and utterances, and whether they maintained a joint focus. Students received low ratings if there was only little talk and high ratings if they were responsive to each other's comments and monitored their partner's attention. *Collaborative motivation* assessed students' attitudes toward the joint problem-solving activity. Low ratings in this dimension were given if students showed a negative attitude toward the interaction with their partner and toward the joint problem-solving activity, while high ratings were only given if both learning partners were actively involved in the problem-solving process. The dimensions *elaboration on content* and *elaboration on hint* evaluated the extent and quality of students' elaborations of the learning content more generally and, specifically, in response to tutor hints. For instance, students received low ratings in the dimension elaboration on hint if they did not read the hints but immediately asked for the next hint until they reached the bottom-out hint that gave them the correct answer; in contrast, they received high ratings if they jointly discussed the CTA hints. To analyze students' interactions concerning the quality of collaboration, we segmented the recordings based on the problem-solving steps described above (see Table 1). Each segment was rated separately; ratings then were averaged across segments of each problem or posttest, respectively.

The second rating scheme evaluated the *quality of the problem-solving process* during particularly challenging problem-solving steps. With this rating scheme we assessed whether students took advantage of the help resources in the learning environment. Based on the literature on learning in mathematics and based on the task structure, we chose two particularly difficult steps of the system-of-equations problems for analysis: deriving the expressions corresponding to the linear equations, and finding the intersection point (see Table 1). During these selected problem-solving sequences we evaluated students' interactions concerning the following aspects: *Mathematical understanding* assessed the dyad's comprehension of the problem steps, taking into account both the amount of CTA help they needed for solving the steps and the level of understanding they expressed when reading hints or correcting errors. We gave low ratings if the dyad needed a lot of CTA assistance to solve a step and if they engaged in trial and error and hint abuse until they found the correct solution; we gave medium ratings if they needed CTA assistance, but revealed some understanding of the correction in their following interaction, for instance, by referring to the underlying mathematical principles; and finally, we gave the highest ratings if the dyad immediately solved a problem step correctly and if their interaction revealed that their correct solution was not due to chance but to a deeper understanding of the underlying mathematical principles. The dimensions *capitalization on social resource* and *capitalization on system resource* assessed whether students took advantage of the support offered in the learning environment by the CTA and by the learning partner to improve their collaborative learning process. For instance, students received low ratings with regard to social resource if they ignored each other's potential for finding the solution and if they did not pay attention to each other's suggestions. High ratings were given if students explained their problem-solving actions to their partner or discussed how to proceed in solving the problem. For system resource, students received low ratings if they engaged in trial-and-error behavior or hint abuse. High ratings were given if they used the help offered by the CTA effectively to increase their learning; for instance, if they discussed and resolved errors flagged by the CTA. The categorical dimension *dyad's strategy* assessed the dominant problem-solving strategy that students showed according to five distinct categories. The first two strategies, trial and error and hint abuse, denote strategies ineffective for learning. In contrast, the strategies immediate error correction, correct input, and elaboration with the learning partner prior to entering the correct solution are regarded as effective problem-solving strategies that potentially yield learning. In the presentation of the results, we summarize the dimension dyad's strategy by indicating the percentage of effective problem-solving strategies employed by the students. In a final step, the ratings of the two problem-solving steps were averaged for each of the assessed dimensions.

The two rating schemes were applied to the interaction data from the 2 days of the learning phase and from the collaborative reproduction posttest on day 3. All problems solved during those days were rated. The results of the rating analyses thus provide a good overview of the development of the collaboration processes within the two dyads over the 3 days of the study. In order to guide the raters' assessment, we developed a rating handbook that described the dimensions in more detail and gave examples for high and low ratings similar to the way done in Table 5. Two raters independently assessed the quality of the interaction, and analysis of the inter-rater reliability showed good results (between $r=.66$ and $r=1.00$).

In addition to the ratings, we took a *narrative approach* in order to closely follow student interaction during one particular problem-solving step: finding the intersection point. The rating analysis revealed huge differences in interaction quality concerning this particular problem-solving step, therefore it seemed interesting for further analysis. Also from a theoretical point of view, this step seemed a good choice for analysis: While most other parts of the problems required problem-solving steps that were already known to the students participating in the study, this step was totally new to them. To investigate how students learned to tackle this problem-solving step, we prepared transcripts of the respective interaction sequences of the two dyads. The analysis then involved multiple cycles of reviewing the students' interaction in ActivityLens and carefully studying the transcripts. When replaying and studying the interaction we took notes on the actions in the CTA environment, the interaction with the learning partner, and the reactions to script instructions. Furthermore, we noted whether actions or interactions that should have occurred did not take place; for instance, if students missed the opportunity to discuss a CTA hint. Although our observations were also guided by those theoretical considerations that formed the basis for the rating schemes, the detailed analysis allowed us to pay attention to additional aspects emerging bottom-up from the data.

Analysis of the learning outcomes

In the test phase, we assessed the impact of the experimental conditions on learning with two reproduction posttests and a future learning posttest (see Table 4). All three tests took place on the computer with the CTA. During the test phase, script support was no longer available in the scripted condition; neither were any of the other two conditions scripted.

Reproduction was assessed by having students solve problems isomorphic to those during instruction. Depending on the condition, the first reproduction test was solved either individually or collaboratively (*condition-specific reproduction*). The second reproduction test was solved individually in all conditions (*individual reproduction*). In both reproduction tests, a maximum of two problems could be solved. Second, students' *future learning* was evaluated with a test that asked students to solve problems of a future CTA unit on inequalities. The test comprised four inequality problems that instructed students to calculate two points and graph the inequality in a coordinate plane. The future learning test was solved either individually or collaboratively according to the condition. However, no script support was available in the scripted condition.

For all tests, two variables were extracted from the CTA log data: The *error rate* measures the relative number of steps that were not solved correctly on the first attempt, as indicated by the student making an error or requesting a hint. An error rate of 0 means that the student solved each step correctly on the first attempt; an error rate of 1 indicates that the student needed CTA assistance (error feedback or hint) for each step of the problem. If a student's first attempt on a step was not correct, he often needed multiple attempts (i.e., made multiple errors or requested several hints) to solve this step correctly. Therefore, we additionally calculated an *assistance score*. The assistance score is the average number of incorrect attempts and hints requested across all steps, thus assessing the assistance a student needed to correctly solve the problems.

Prior knowledge as covariate

Students' prior knowledge in algebra can be expected to have a substantial impact on the acquisition of new learning material. For instance, students need basic knowledge of equation solving and plotting points. In order to statistically control for individual differences, we collected data on students' prior knowledge to include it as covariate in the statistical model. Prior knowledge was operationalized by students' current level of performance in algebra (0–100 %) as reported by their teachers.

Results

We analyzed the effects of scripted collaborative learning with the CTA in two steps: First, we conducted analyses of the collaboration process of two dyads (one from the scripted and one from the unscripted condition). The analyses were done using two rating schemes and a narrative approach. Results from the ratings are summarized in Table 6 for the learning phase and in Table 7 for the condition-specific reproduction posttest. The outcome data of the two dyads are provided in Table 8. Second, we statistically compared the learning outcomes of the three conditions based on the posttest data. The results of the two reproduction posttests and the future learning posttest are presented in Table 10.

Results of the rating analysis

As described above, we had aimed to record student interaction during the learning phase and during the collaborative reproduction posttest. However, the screen capture tool failed to start recording several times leaving us with only a few complete process recordings. In addition, in a number of recordings, the audio quality was not sufficient to allow for an analysis of students' utterances. Thus the choice for our in-depth process analysis was severely limited. We chose two dyads for which we had complete or almost complete recordings of acceptable quality: The dyad *Aristotle* (scripted condition) and the dyad *Telemann* (unscripted condition).

As shown in Table 6, the scripted dyad Aristotle only solved two problems *during the learning phase*. After having completed the individual phase, students started the collaborative phase of problem 1 on the first day (deriving expressions, and solving questions 1 and 2, see Table 1) and finished it at the beginning of the second day (solving question 3: graphing, and finding intersection point; see Table 1). The collaborative phase of problem 2 was solved on the second day of the learning phase. In contrast, the unscripted dyad Telemann solved four problems during the learning phase. Problem 1 was solved on the first day, and problems 2 to 4 were solved on the second day of the learning phase. Unfortunately the video of the first problem was incomplete. The recording stopped when students started to graph the lines in the Grapher; thus, for the following problem-solving process, only log data are available. Therefore, we were not able to rate the last two steps of this problem (i.e. graphing the equation and calculating the intersection point, see Table 1). The smaller number of problems that were solved in the scripted dyad as compared to the unscripted dyad is concordant with the ratio of solved problems in the whole study sample (unscripted condition M=3.50, SD=1.83; scripted condition M=1.79, SD=.80) and can be explained by the script instructions that directed students in their collaborative activities—and that asked for more than they would probably have engaged in when collaborating without script support.

When comparing the dyads Aristotle and Telemann with regard to the *quality of the collaboration process* during the learning phase, we can see huge differences. The interaction of the dyad Aristotle is characterized by a constantly good *collaboration flow* and a high *collaborative motivation* during the learning phase. At the beginning of their interaction, the dyad Telemann also shows a good collaboration flow and a high collaborative motivation for the joint problem-solving (see Table 6). However, for both dimensions, ratings decreased during the course of the second and third problem solved by Telemann. The slight improvement in the collaboration flow and the collaborative motivation for the fourth problem can be explained by an interaction sequence at the end of the third problem: During the second and t6.1 **Table 6** Ratings of interaction quality during the learning phase

t6.2		1st problem				2nd problem				3rd problem				4th problem			
t6.3		Quality of the collaboration															
t6.4		CF	CM	EC	EH	CF	CM	EC	EH	CF	CM	EC	EH	CF	CM	EC	EH
t6.5	Aristotle	3.7	3.7	3.0	3.0	3.8	3.3	1.8	2.0	–[a]	–	–	–	–	–	–	–
t6.6	Telemann	3.0[b]	3.0	2.0	1.5	2.0	1.8	1.3	0.0	1.3	1.0	0.8	0.0	2.5	2.0	1.5	0.0
t6.7		Quality of the problem-solving process															
t6.8		MU	SOR	SYR	DS[c]	MU	SOR	SYR	DS	MU	SOR	SYR	DS	MU	SOR	SYR	DS
t6.9	Aristotle	3.0	4.0	3.0	100 %	4.0	2.0	–[d]	100 %	–[a]	–	–	–	–	–	–	–
t6.10	Telemann	2.0[b]	2.0	1.0	0 %	1.5	0.5	0.5	0 %	1.5	0.0	0.5	0 %	3.0	1.0	1.0	50 %

Quality of the collaboration: *CF* collaboration flow; *CM* collaborative motivation; *EC* elaboration on content; *EH* elaboration on hint;

Quality of the problem-solving process: *MU* mathematical understanding; *SOR* capitalization on social resources; *SYR* capitalization on system resources; *DS* dyad's strategy

[a] Dyad Aristotle solved only 2 problems during the learning phase

[b] For the 1st problem of Telemann, rating is only available for the first two sections *preparation* and *simple questions*; the rest of the video was missing due to technical difficulties

[c] The dyad's strategy is summarized as the percentage of effective strategies (i.e. immediate correction, correct input, elaborating with partner vs. trial and error and hint abuse) employed during interaction

[d] The dyad neither asked for hints nor made errors when solving the most challenging problem-solving steps of this problem

Table 7 Ratings of interaction quality in the condition-specific reproduction posttest

	Quality of the collaboration			
	CF	CM	EC	EH
Aristotle	4.0	4.0	2.8	2.0
Telemann	1.5	1.8	0.8	0.0
	Quality of the problem-solving process			
	MU	SOR	SYR	DS[a]
Aristotle	2.5	3.5	3.0	100 %
Telemann	1.0	1.0	1.0	0 %

Quality of the collaboration: *CF* collaboration flow; *CM* collaborative motivation; *EC* elaboration on content; *EH* elaboration on hint;

Quality of the problem-solving process: *MU* mathematical understanding; *SOR* capitalization on social resources; *SYR* capitalization on system resources; *DS* dyad's strategy

[a] The dyad's strategy is summarized as the percentage of effective strategies (i.e. immediate correction, correct input, elaborating with partner vs. trial and error and hint abuse) employed during interaction

third problem, Telemann B shows little interest in interacting, ignoring his partner's utterances and solving the problem on his own; this causes Telemann A to complain about his partner's attitude, and he asks him to engage in the interaction as well, which leads to improvement in their collaboration on the fourth problem. More detail on this instance will be provided in the results of the narrative analysis. As discussed in the theoretical background and indicated by the results in Table 6, the two dimensions collaboration flow and the collaborative motivation are important prerequisites for the overall collaboration quality. It is likely that if these dimensions are rated as low, a dyad also shows low ratings on the other dimensions (e.g., Telemann, third problem). But a high collaboration quality concerning collaboration flow and collaborative motivation is not sufficient, as a high amount of interaction does not guarantee deeper elaboration. For instance, despite the high collaboration flow during the first problem, Telemann shows only a medium *elaboration on the content* and a low *elaboration on the hints* they receive. In fact, their elaboration on both dimensions is low throughout their interaction during the learning phase, whereas Aristotle shows high elaboration particularly during the first problem they solve, that is, when they encounter the system-of-equations task type for the first time.

We see even higher differences between the dyads' interactions during the learning phase when comparing their ratings concerning the *quality of their problem-solving process* during the particularly challenging problem-solving steps: deriving the expressions and finding the

Table 8 Descriptive variables and posttest results of Aristotle (scripted) and Telemann (unscripted)

Gender	Aristotle		Telemann	
	Male	Male	Male	Male
Prior knowledge: unit in CTA	8	7	10	10
Condition-specific reproduction				
Error rate	0.38		0.31	
Assistance score	1.03		1.31	
Future learning				
Error rate	0.54		0.53	
Assistance score	3.17		3.11	

intersection point. The dyad Aristotle makes effective use of the opportunities provided by the collaborative learning environment: They discuss their solution approach and work together on solving the difficult problem-solving steps (*capitalization on social resource*). They reflect on the hints they have requested and capitalize on the errors they have made during the first problem (*capitalization on system resource*). Thus, they manage to solve the difficult problem-solving steps of the second problem without the need for CTA assistance. Furthermore, they exclusively engage in effective problem-solving *strategies*. As a consequence, the dyad Aristotle shows a high mathematical understanding during the first problem; during the second problem they even receive the highest possible ratings on this dimension. The narrative analysis further illustrates how the collaboration script supported the interaction of the students in this dyad.

In contrast, the Telemann partners barely take advantage of the collaborative learning environment, that is, of the *social* and *system resources*. While the dyad still receives medium ratings on these dimensions during the first problem, the ratings are close to zero for the second and third problem they solve. Furthermore, with the exception of the final problem, they solely engage in ineffective problem-solving strategies (*dyad's strategy*), frequently showing trial and error and hint abuse behaviors. As a consequence, Telemann barely shows any progress in their *mathematical understanding* during the learning phase: They need a large amount of CTA assistance to solve the problems, but show only a low understanding of the corrections and the hints they receive. The improved rating for the fourth problem does not indicate an improved understanding of the system-of-equations concept (i.e., the target concept in our study): decomposing the ratings of the two analyzed problem-solving steps reveals that only the step "deriving expressions" was rated higher (with 4), whereas the step "finding the intersection point" still only received a rating of 2. This also explains why the dyad Telemann did not succeed in finding the intersection point in the condition-specific reproduction test.

Interestingly, the scripted dyad Aristotle shows a higher *quality of collaboration* not only during the learning phase, but also during the *condition-specific reproduction posttest* (see Table 7). The interaction of the dyad Aristotle shows a better *collaboration flow* and a higher *collaborative motivation* than the interaction of the dyad Telemann. In the dyad Aristotle, both learning partners are engaged in the interaction, while the learning partners of the dyad Telemann do not establish a joint focus on the problem and do not contribute equally to the problem-solving process. Moreover, Aristotle receives good ratings for the two dimensions *elaboration on the content* and *elaboration on the hints*. Telemann on the other hand shows a low level of elaboration on both dimensions.

Also the *quality of the dyads' problem-solving process* differs during the condition-specific reproduction test. The dyad Aristotle shows a medium level of *mathematical understanding*. Compared to the final problem during the learning phase (see Table 6) the dyad thus receives a slightly lower rating on this dimension. Decomposing the two averaged ratings reveals that this is mainly due to difficulties with deriving the expressions from the story problem and not due to difficulties with the new and central question type finding the intersection point: for the interaction sequence "deriving the expressions" Aristotle receives the rating 2; the sequence "finding the intersection point" is rated with 3. As was the case during the learning phase, the dyad capitalizes effectively on the *social* and *system* resources and engages in effective problem-solving *strategies* to solve the most difficult problem-solving steps. In contrast, Telemann again barely capitalizes on the *social* and the *system resources* and engages in trial and error and hint abuse (ineffective *dyad's strategy*). Furthermore, the two students show a low level of *mathematical understanding*.

Results of the narrative approach

In the previous section, we compared the ratings of the quality of the interaction process of the dyads Aristotle and Telemann. The analysis showed how the students' interaction evolved over the course of the learning phase and how it was rated in the condition-specific (i.e., collaborative) reproduction posttest. In the following sections, we analyze in detail the interaction during the new and most challenging step of the system-of-equations problems: finding the intersection point. The narrative analysis was based on transcripts and video data. We reviewed the interaction multiple times and took notes on the actions and interactions to describe the problem-solving process in detail. The results from the rating analysis already indicated substantial differences in the interaction quality during this particular problem step and we attempt to further illuminate these differences here. Moreover, the in-depth analysis enables us to investigate the effects of the collaboration script on student interaction and learning, answering questions like: Does the script promote equal contribution to the problem-solving process? And is the adaptive support successful in fostering student elaboration?

Analyzing the dyads' collaboration during the learning phase

When solving the intersection point question of the first problem, the dyad *Aristotle* starts by reading the question out loud together: Aristotle A reads the first part "How much in weekly sales would give him the same salary for both choices?", and Aristotle B the second part "Find the answer algebraically". Thus, they start out with a joint focus of attention on the task. Next, Aristotle A articulates his confusion about the question several times and proposes to guess the answer; meanwhile, Aristotle B attempts to understand the problem posed by elaborating on the problem statement. He reads the question once again, accentuating the significant information: "How much in weekly sales would give him the SAME salary for both choices? Find the answer algebraically". Furthermore, he gives an example to describe the situation they are looking for: "… he's gonna make 600$ in (.) you know first choice and then 600$ in the second choice" (note: "first choice" and "second choice" refer to two job offers to be compared in this system-of-equations problem). This elaboration leads his partner Aristotle A to conclude that "(t)here has to be a pattern" that should allow them to find the answer. When he realizes that the salaries for the first and the second job offer resulting from the previous question they have solved were quite similar (total weekly sales $400; salary for first choice $400, salary for second choice $475), he simply enters a value for the weekly sales ($500) that is close to the one given in the previous question. The answer is wrong, and an adaptive script message comes up, reminding the students to consult with their partner or ask for a hint if they do not know how to find the solution. Following this advice, Aristotle B suggests asking for a CTA hint. Even though the hint already tells them quite clearly how to proceed ("Given that the expression for the salary from the first choice and the salary from the second choice are equal, write an equation and solve it to find the total weekly sales"), they click through the hints until—before the bottom-out hint—a second adaptive script message (penultimate hint message) pops up, prompting them to collaboratively make use of the hints received so far. The following episode is characterized by productive co-construction. The two students work hand in hand proposing problem-solving steps; they complete each other's sentences and build on each other's comments. For example, when Aristotle B says "Now, just—", Aristotle A states at the same time "And (do that) in there?"; then Aristotle B takes up and answers: "Yeah, 75 plus point—or 0 or whatever point". This collaborative contributing to the problem-solving process indicates that both students are learning together how to find the intersection point. Aristotle A takes over the responsibility for typing in the CTA as they solve the equation for x. Yet, both students are actively involved and pay attention to the problem-solving steps: They always discuss the necessary steps before entering them in the CTA. Despite their good collaboration, however, they are not able to completely solve the equation on their own. They have difficulties with the transformation step that requires combining both variable terms on one side. After two unsuccessful attempts, the CTA automatically launches a hint message; however, the hint message unfortunately is erroneous and does not propose a suitable next step, thus the dyad asks the teacher how to proceed. The teacher helps them to solve the problem step, and the dyad finishes solving the equation for x.

During the second problem, the dyad Aristotle successfully applies the knowledge gained from the first problem in order to find the intersection point. Again, Aristotle B reads out the question. Immediately, both students agree on how to approach the question: to go to the Solver and equate the two expressions of the problem. Aristotle A says: “We have to do that thing again”, and Aristotle B agrees: “Yeah, Solver, that’s easy, new equation, all right, you start typing in”. The almost simultaneous start of their talking indicates that both students are actively involved in problem-solving and that they have both gained an understanding of what to do. The motivation to be equally engaged in problem-solving is also expressed in the following sequence, in which they explicitly distribute the workload: When Aristotle B suggests that his partner enters the equation: “All right, you start typing in”, Aristotle A agrees and suggests that Aristotle B tells him the equation to write down:“Ok, tell me what to type in”. Aristotle A’s request does not imply that he would not be able to derive the equation on his own. In fact, at one point he writes down an arithmetic operator before Aristotle B tells him to. He pays attention to the problem solving and does not have to rely on his partner to find the solution. As during the first problem, Aristotle A takes responsibility for mouse and keyboard as they solve the equation for x; however, in contrast to Telemann B in the unscripted dyad (see below), he begins each problem-solving step by proposing what to do next and then makes a short pause, allowing his partner to agree or disagree. The dyad successfully solves the equation and enters their answer in the Worksheet.

In the following paragraphs, we elaborate on the difficulties of the unscripted dyad *Telemann* in learning how to find the intersection point. When solving the intersection point question of the first problem,[1] the two students enter the correct answer in the Worksheet immediately after finishing the graphing (after about 57 s) and without using the Solver tool. This indicates that the dyad does not find the intersection point algebraically, but they employ a graphical strategy: they identify the point’s coordinates in the Grapher window. If the coordinates of the intersection point are integers, as was the case in the first problem, this is a successful strategy that demonstrates students’ understanding of the relationship between the graphical and the tabular representation. However, the strategy fails if the point’s coordinates are decimal numbers, as was the case in the subsequent problems.

During the second problem, the dyad again tries the graphical strategy to find the intersection point: At the end of the graphing step, Telemann A states that the intersection point must be approximately at 7.2 min. He proposes entering 7 in the Worksheet, stating that “it [the CTA] should correct it”. This statement is a typical example of relying on the CTA support functionalities and gaming the system. Even though the CTA marks their

[1] For the first problem of the learning phase, video data of Telemann’s interaction during this sequence were not available; therefore, the analysis is based on log data.

answer wrong, the dyad sticks to their strategy: They enter further numbers close to 7 until the CTA automatically launches a hint message after the third incorrect attempt (trial and error). They click on the “next” button in the hint dialogue until the bottom-out hint is displayed. It instructs them to equate the two expressions in order to find the answer, but the dyad simply copies the equation given in the hint into the Solver window; a typical case of hint abuse as described in the introduction. During the subsequent equation solving, Telemann B takes over the responsibility, entering actions and transforming the equation in the CTA. However, he barely ever comments on what he is doing. Meanwhile, Telemann A reads out loud some of his partners’ actions and the error messages presented by the CTA. The actions and verbal utterances of the two students often do not refer to each other, indicating that they are not really paying attention to what their partner is doing. For instance, at one point Telemann A proposes a transformation step without realizing that his partner has already tried out exactly the same step without success a couple of seconds ago. Telemann B, on the other hand, shows little interest in interaction in general: He neither explains his own actions nor does he react to the solution proposals of his partner. Telemann A reacts to this behavior with off-topic talk and plays around with his microphone. The dyad struggles most with transforming the equation $-8\ M=-6\ M - 100$ to $-2\ M=-100$. To perform this step, students have to put all terms referring to the variable to one side (here, by adding 6 M). After several unsuccessful attempts to transform this equation, Telemann B follows his partner’s proposal to ask for a hint. He clicks on the “next” button in the hint window as quickly as possible until he reaches the bottom-out hint that tells them the next problem-solving step. In fact, the time interval between receiving one hint and clicking ahead to the next hint is too short to even read the hints. In other words, the dyad does not try to elaborate on the help they receive, but deliberately abuses the hints. When performing the step suggested in the bottom-out hint, Telemann B makes a typo, entering 6 instead of 6 M. Although the reaction of Telemann A clearly expresses his confusion: “What the beef. It’s like, er, what is it like, er”, Telemann B does not attempt to explain his actions when correcting the error. In the end, Telemann A no longer insists on receiving an explanation, but merely comments: “Ok, you figured it out”.

When solving the third problem, the dyad again initially tries to find the intersection point by employing a graphical solution approach. After the first attempt is marked as wrong by the CTA, Telemann A remarks that they might have to use the Solver again: “…(oh) we’ll have to do this on the solv-thingee”. Telemann B does not follow his advice, but tries out two more values until the CTA automatically launches a hint message telling them to approach the problem by writing an equation. Even though the dyad has just solved a similar problem, they do not capitalize on their previous experience and the information given in the hint; instead, Telemann B again immediately clicks to the bottom-out hint and copies the equation given there. As in the previous problems, he takes control of the CTA. His obvious lack of interest in collaboration also reduces the efforts by his partner: Although Telemann A still makes a few proposals on problem-solving steps, he mainly engages in off-topic talk. As in the previous problem, Telemann B does not follow his partner’s proposals, but solves the question on his own. When Telemann A suggests an erroneous problem-solving step (adding 9 instead of 9D), Telemann B does not correct him, but merely enters the correct step. The lack of interest in collaborating finally leads Telemann A to complain: When Telemann B again enters a problem step while Telemann A is still trying to figure out what to do next, he verbally expresses his frustration: “Hey, why aren’t you speaking at all? This is supposed to be a group effort here!”. At first, Telemann B does not take the complaint seriously, but rather plays it down, responding that “(s)omebody has to push buttons”. Telemann A insists: “but you are (also) supposed to explain how this is DONE!”. In consequence, the collaboration slightly improves during the solving of the fourth problem.

Even though using a graphical solution approach to find the intersection point had proven unsuccessful in the previous three problems, the dyad Telemann again tries this strategy on the fourth problem. In contrast to the previous problems, they do not even wait for the CTA hint message to automatically launch after several errors, but ask for a hint immediately after their second unsuccessful attempt. As before, they click through the hint dialogue and copy the equation provided in the bottom-out hint (hint abuse). While the dyad's problem-solving is still of low quality, their motivation to collaborate with each other has slightly increased compared to the previous problems, and they pay attention to each other's utterances and actions. For instance, when Telemann A proposes problem-solving steps, Telemann B follows his proposals until they find the correct answer. The improved collaboration is also reflected in the ratings of the dyad's interaction during the fourth problem (see Table 6 and related result presentation above).

Short overview of dyads' collaboration during the reproduction test

Although none of the dyads was scripted during the condition-specific reproduction test, the two dyads still differ in their interaction. The dyad *Aristotle* solved only two problems during the learning phase and thus had rather little opportunity to practice the new question type intersection point. Nevertheless they successfully solve the posttest problem with little assistance by the CTA. The problem-solving process of the dyad Aristotle is again characterized by mutual contributions and knowledge co-construction. For example, when Aristotle B wonders: "Equals what, what has to be equal?", Aristotle A explains what they need to do and tries to help his partner by referring to their earlier experiences: "Yeap, cause that's what we did yesterday". Finally Aristotle B gets it: "Ok, remember. So. Solver", and enters the equation in the solver window. Furthermore, the dyad takes advantage of the CTA learning environment and employs the strategy they were instructed to use by the script during the learning phase: When they are stuck in their problem-solving or when the CTA marks one of their actions as error, they do not engage in trial and error, but ask for a hint, which they then discuss and try to use to proceed. For instance, when a CTA hint tells them to "subtract 0.35 M from both sides", the two students initially agree that this is what they have just done and wonder. All of a sudden Aristotle A notices: "Oh, I forgot for M", and Aristotle B concurs: "Oh yeah". Now they are able to proceed without clicking any further through the hint hierarchy.

In contrast, although they solved four analogous problems during the learning phase and although they receive ample support by the CTA (error flagging and hint messages), the dyad *Telemann* does not succeed in finding the intersection point when collaboratively solving the system-of-equations problem in the posttest. The inferior performance of the dyad Telemann in finding the intersection point in the reproduction test can be attributed both to their suboptimal problem-solving behavior during the learning phase and to their unfruitful interaction during the test phase: As they did during the learning phase, they do not effectively capitalize on the collaborative learning environment at hand. When Telemann A tries to gain an understanding of the task and attempts to discuss it with his partner at the beginning, Telemann B simply ignores him. Furthermore, when Telemann A tries to understand what his partner is doing later in the process, he does not receive appropriate answers. For instance, at some point during the problem-solving process Telemann A requests an explanation: "Now what are you doing for this?", but Telemann B merely responds: "Praying". At another point when Telemann A asks Telemann B how he found a certain value: "How did you find the bottom one?", Telemann B answers: "Very carefully". Telemann A insists: "And you did that how, other than carefully?", but receives no further answer. Even after several unsuccessful attempts, Telemann B is not willing to start interacting with his partner, but further engages in trial and error and hint abuse until time is up. He does not leverage the competencies of his partner and in the end they fail to solve the test problem.

Learning outcome of the two dyads

If the hypothesized connection between collaboration quality and learning outcome holds true, the interaction patterns of the two analyzed dyads should link to their posttest results. Thus, in this section we descriptively relate the interaction quality with prior knowledge and the learning outcome as assessed by the two posttest variables error rate and assistance score. The two dyads entered the study with very different levels of prior knowledge: Of the dyad Aristotle one student had gotten as far as unit 8 of the CTA, while the second student was still working on unit 7, the unit that introduced linear equations, which was a prerequisite for solving the system-of-equations problems during the study. In contrast, both students of the dyad Telemann had already reached unit 10 of the CTA prior to the study. Yet, in the collaborative posttests "condition-specific reproduction" and "future learning" the two contrasting dyads show equally good performance (see Table 8): In the collaborative reproduction test, Telemann has a slightly lower error rate, but needs more CTA assistance to correct their errors and to find the right solution. In the future learning test, the dyads' performance is approximately the same. Thus, the students of the dyad Aristotle learned more: they both had entered with lower levels of prior knowledge, but reached comparable learning outcomes as the two Telemann partners. This result is in line with the findings from the process analyses and provides some initial support for the assumption that better collaboration is likely to lead to better learning.

Learning outcome of the whole sample: Between-condition comparison

Can the differences in the learning gains we observed for the two case dyads also be found in the between-condition comparison of the whole sample?

As we had expected *prior knowledge* to have a substantial impact on the acquisition of new learning material and because we have seen differences in the prior knowledge of the two analyzed dyads, we first compared the three study conditions concerning their prior knowledge, assessed as students' current level of performance in algebra (0–100 %). Descriptively, prior knowledge was highest in the unscripted condition and lowest in the scripted condition (see Table 9), indicating a similar pattern as the one seen in the analyzed dyads. The differences were, however, not statistically significant, $F(2,103)=1.77$, $p=.18$.

Next, we tested the influence of prior knowledge on the learning outcomes. The theoretically assumed correlation between prior knowledge and outcome measures was confirmed by the empirical data: Prior knowledge had a significant impact on all outcome measures ($r=.32-.54$, $p<.05$). Therefore, it was included as covariate in the data analyses. For the collaborative posttests that were analyzed on the dyadic level, we used the dyad's average

Table 9 Prior knowledge

	Scripted	Unscripted	Individual
Prior knowledge M (SD)	81.90 (10.37)	85.25 (6.03)	84.47 (8.35)

prior knowledge as a covariate. To balance the descriptive differences between conditions we report the adjusted means for the following analyses (cf., Huitema 1980). These are the values that would be predicted if the covariate means of conditions were the same as the grand covariate mean.

To analyze the effect of the study conditions, we computed a MANCOVA analysis for each of the three posttests. Two independent a priori contrasts tested our hypotheses: First, we compared the individual condition with the collaborative conditions to assess the impact of collaboration; second, we contrasted the two collaborative conditions with each other to evaluate the script's effect. As described above, the outcome variables of interest were the error rate and the assistance score. The error rate measures students' ability to solve a step correctly on the first attempt, while the assistance score evaluates the average amount of assistance (errors and hint requests) needed to solve the problems. In those cases where we found indications of an interaction between prior knowledge and condition (aptitude treatment interaction), the interaction term was included in the GLM model as the exclusion of the interaction term would violate the assumption of homogenous regression slopes (Field 2005).

Adjusted means and standard errors for the three posttests are presented in Table 10. For the *condition-specific reproduction* test, the MANCOVA analysis revealed a significant aptitude treatment interaction of prior knowledge and condition, $F(4,94)=3.30$, $p=.01$, $\eta^2=.12$, thus the model including the interaction term was used in the following analyses. As expected, prior knowledge had a strong influence on both outcome measures, $F(2,46)=13.66$, $p=.00$, $\eta^2=.37$. Furthermore, conditions differed significantly with regard to the measures of condition-specific reproduction, $F(4,94)=3.34$, $p=.01$, $\eta^2=.12$. The subsequent ANCOVA analysis of the *error rate* revealed a significant influence of the covariate prior knowledge, $F(1,47)=24.96$, $p=.00$, $\eta^2=.35$. However, we did not find a significant interaction of prior knowledge and condition, $F(2,47)=.14$, $p=.87$, nor did we find a significant effect of condition on the error rate, $F(2,47)=.09$, $p=.92$. In the ANCOVA analysis of the *assistance score*, we found a marginally significant interaction of prior knowledge and condition, $F(2,47)=2.55$, $p=.09$, $\eta^2=.10$. Again, prior knowledge had a significant effect, $F(1,47)=6.15$, $p=.02$, $\eta^2=.12$. Furthermore, data analysis revealed a marginally significant difference between conditions, $F(2,47)=2.81$, $p=.07$, $\eta^2=.11$, with most assistance needed

Table 10 Posttest results

	Scripted M (SE)	Unscripted M (SE)	Individual M (SE)
Condition-specific reproduction			
Error rate	.41 (.03)	.38 (.03)	.36 (.03)
Assistance score	1.10 (.12)	.86 (.12)	.93 (.13)
Individual reproduction			
Error rate	.35 (.02)	.36 (.02)	.36 (.04)
Assistance score	1.01 (.11)	.98 (.11)	.97 (.17)
Future learning			
Error rate	.36 (.03)	.44 (.03)	.30 (.03)
Assistance score	2.01 (.34)	2.73 (.37)	1.85 (.40)

For error rate and assistance score, smaller numbers indicate better performance

**p<.01; \ast p<.05; + p<.10; – = not assessed by dyads of the scripted condition (see Table 10). The predefined contrasts did not reveal significant results. To analyze the significant aptitude treatment interaction effect in more detail, we calculated regression analyses with prior knowledge as the predictor and assistance score as the criterion separately for each of the three conditions.

As indicated by the regression slopes in Fig. 4, the influence of prior knowledge on the assistance score was highest in the scripted condition (regression coefficients: individual condition $b=-.01$, unscripted condition $b=-.02$, scripted condition $b=-.04$), thus the slight disadvantage of the scripted condition regarding the assistance score could at least partly be ascribed to the high amount of assistance needed by students with low prior knowledge.

Prior to analyzing the data of the *individual reproduction* test, we had to attend to a methodological issue: The analysis of individual posttest data in a study on collaborative learning and problem solving raises the question if the observations of two dyad partners can be considered independently (e.g., Cress 2008). Following the methodological approach suggested by Kenny and colleagues (1998), we therefore analyzed the intraclass correlations between individual posttest scores of dyad partners in the individual reproduction test. Neither the analysis of the variable error rate nor the analysis of the variable assistance score revealed a consequential nonindependence (i.e. an intraclass correlation between dyad partners that is higher than $r=.45$ and significant at an alpha level of .20, cf. Kenny et al. 1998). Thus, we were able to include both dyad partners in the analysis individually.

For the individual reproduction test, results of the MANCOVA revealed a significant effect of prior knowledge on student performance, $F(2,89)=17.63$, $p=.00$, $\eta^2=.28$. Condition did not show an effect, $F(4,180)=.15$, $p=.96$. Result of the subsequent ANCOVAs were concordant with the MANCOVA analysis: Prior knowledge significantly influenced the *error rate*, $F(1,90)=35.37$, $p=.00$, $\eta^2=.28$; however, condition did not impact the amount of errors on the first attempt, $F(2,90)=.02$, $p=.98$. Also the ANCOVA analysis

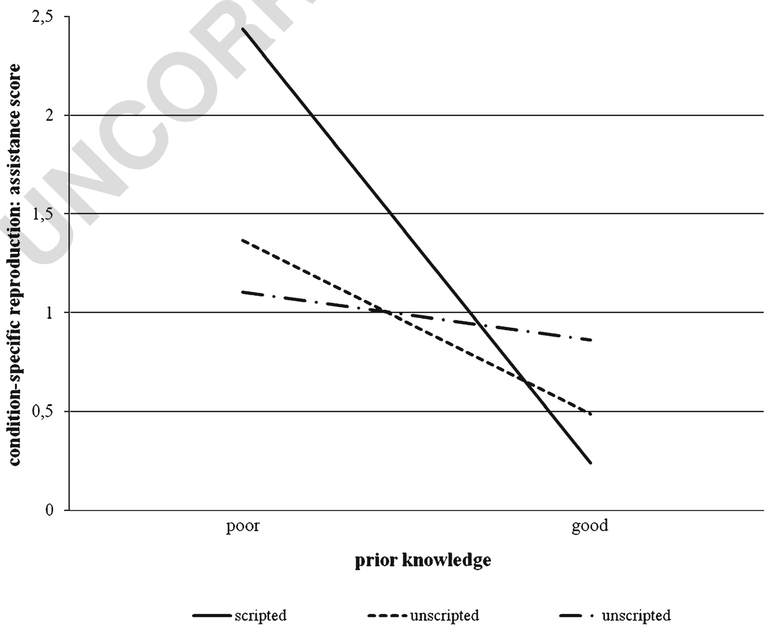

Fig. 4 Influence of prior knowledge on the assistance score in the condition-specific reproduction test (regression slopes)

of the *assistance score* showed a significant influence of prior knowledge, $F(1,90)=26.83$, $p=.00$, $\eta^2=.23$, while study conditions did not differ in the amount of assistance needed to solve problems, $F(2,90)=.03$, $p=.97$.

Results of the MANCOVA analysis of the *future learning* test showed, once more, that prior knowledge influenced students' performance, $F(2,53)=11.03$, $p=.00$, $\eta^2=.29$. Furthermore, we found a significant effect of condition $F(4,108)=2.74$, $p=.03$, $\eta^2=.09$. The separate ANCOVAs for the two outcome measures revealed that the significant result of the multivariate analysis could be ascribed to the variable *error rate*: Conditions differed with regard to the average number of errors on their first attempt, $F(2,54)=5.46$, $p=.01$, $\eta^2=.17$. Furthermore, both planned contrasts yielded significant results: The individual condition showed a lower error rate than the two collaborative conditions, $t(54)=2.67$, $p=.01$, and dyads from the scripted condition had a lower error rate than dyads from the unscripted condition, $t(54)=2.11$, $p=.04$. Prior knowledge had a significant influence on error rate, $F(1,54)=21.86$, $p=.00$, $\eta^2=.29$. Although the pattern was similar with regard to the *assistance score*, neither the overall difference between conditions, $F(2,54)=1.54$, $p=.22$, nor the planned contrasts reached statistical significance (for the first contrast $t(54)=.1.09$, $p=.28$, for the second contrast $t(54)=1.43$, $p=.16$). Again, prior knowledge had a significant influence on students' achievement, $F(1,54)=14.53$, $p=.00$, $\eta^2=.21$.

Discussion and conclusions

Summary of results

In the present study we tested collaboration extensions to the Cognitive Tutor Algebra (CTA, © Carnegie Learning Inc.), a tutoring system for high-school mathematics, with the goal to promote student learning. As we argued in the introduction, research has demonstrated that fruitful collaboration does not automatically result from having two students work together. Therefore, we developed a collaboration script to support the interaction. In an experimental classroom study we compared scripted collaboration to unscripted collaboration and individual learning. In our analyses we tested two assumptions: First we compared the collaboration process of one dyad from the scripted condition and one dyad from the unscripted condition, in order to test the assumption that the collaboration script would increase fruitful interaction and thus promote the collaborative learning process. We analyzed the interaction of the two dyads with two rating schemes: one rating scheme evaluated collaboration quality from a rather general point of view, and the other rating scheme looked at the quality of the problem-solving process in the specific setting (i.e. collaborative learning with the CTA). In addition, we conducted an in-depth narrative analysis of one particularly difficult step in the system-of-equations tasks that students encountered in our study: calculating the intersection point. Both types of process analyses were carried out for the collaboration during the learning phase and during the condition-specific reproduction posttest, where dyads collaborated without script. We also related the process analyses to the learning outcomes of the two dyads. Second, we tested the assumption that collaboration—and especially scripted collaboration—would lead to improved learning by statistically comparing the learning outcomes across conditions for the whole sample.

In summary, in the *process analyses* we found clear differences between the interaction patterns of the two analyzed dyads. The results of the *rating analysis* showed that the interaction of the scripted dyad Aristotle during the learning phase was of higher quality than the interaction of the unscripted dyad Telemann. The scripted dyad Aristotle collaborated in a productive way, particularly after some adaptive support had been provided by our collaboration script. On the other hand the unscripted dyad Telemann did not take advantage of learning opportunities provided by the collaborative setting, but mainly abused the CTA hints to solve problems faster. Moreover, the scripted dyad Aristotle continued to show a higher quality in their collaboration and in their problem-solving during the condition-specific (i.e. collaborative) reproduction posttest than the unscripted dyad Telemann. In other words, the two Aristotle students were rather successful in transferring their good collaborative behavior from the scripted interaction during the learning phase to the test phase, where script support was no longer available.

The in-depth *narrative analysis* of the intersection point problem-solving step supported the results revealed by the ratings: The analysis of the relevant sequences in the problem solving of the dyad Aristotle during the learning phase clearly showed that both students learned how to find the intersection point algebraically. During the first problem, the two students were initially unsure how to approach the question and had difficulties when solving the equation. At this point we could see how the adaptive script element influenced the interaction. An adaptive script message encouraged students to ask for a hint, in other words, the script instructed them on a strategy fruitful for learning: asking for help. Next, a penultimate hint message prevented students from abusing the hint hierarchy to get the right answer. Surprisingly, merely mentioning that they might be able to solve the problem step on their own was sufficient to keep these two students from requesting the final hint that would have given them the answer, and stimulated them to collaboratively solve the step on their own. In the second problem, Aristotle did not need CTA assistance (error flagging or hint messages) anymore either to derive the equation or to solve it and compute the intersection point. During the condition-specific reproduction test, the problem solving of the dyad Aristotle was again characterized by mutual contributions and knowledge co-construction. They succeeded in solving the intersection point question with only little assistance by the CTA.

In contrast, the analysis of the collaborative problem solving of the dyad Telemann during the learning phase revealed that they did not achieve an understanding of how to find the intersection point algebraically. In none of the four problems did they derive the equations for calculating the intersection point on their own. During the whole learning phase, they abused the hints given by the CTA to copy the solution from the bottom-out hint. In fact, they even moved the hint window closer to the Solver tool in order to facilitate the copying. They only collaboratively engaged in the problem-solving process after Telemann A expressed his frustration. Cleary, a more elaborative way of using the learning resources available (system resources and social resources) would have been desirable. Unfortunately, also during the collaborative reproduction posttest, the dyad Telemann failed to collaborate fruitfully and did not find the intersection point even though they received ample support by the CTA (error flagging and hint messages).

The differences that we saw in the interaction patterns of the *two dyads* were also confirmed to some extent when descriptively comparing their learning gains: the dyad Aristotle started at a much lower level of prior knowledge than the dyad Telemann, but performed as well as Telemann in the collaborative reproduction test and in the future learning test.

We could not clearly establish benefits of the scripted collaboration condition in the *between-condition comparison* of the *learning outcomes* of the whole sample (for an overview of the results, see Table 10). While the analysis of the *condition-specific reproduction* test revealed no difference in the error rate, we found differences in the assistance students needed to solve problems. As the aptitude treatment interaction effect and the subsequent regression analyses revealed, a high need of assistance was particularly found in those dyads of the scripted condition who had entered the collaboration with poor prior knowledge. On average, these dyads made more errors and asked for a higher amount of hints per problem-solving step compared to students with a comparable prior knowledge level that learned in the individual or in the unscripted condition. In the *individual reproduction* test, however, the disadvantage of students of the scripted condition who had entered with low prior knowledge no longer persisted: There was no statistical difference between conditions concerning the number of errors made and the amount of assistance needed to solve the problems. In the *future learning* test, we found significant differences for the variable error rate, favoring individual learning over collaborative learning, and scripted collaboration over unscripted collaboration. The assistance score showed the same pattern, but the differences did not reach significance.

Discussion of results

Why did the collaborative learning conditions not yield improved learning outcomes in the *reproduction tests*? First, it is possible that during the learning phase collaborative students, and particularly those in the unscripted condition, did not engage in the types of elaborative collaborative behaviors considered beneficial for learning. This interpretation is in line with the results of process analyses of the dyads Aristotle and Telemann: The analyses revealed elaborative discussions, particularly after hints, in the scripted dyad Aristotle, while the unscripted dyad Telemann frequently engaged in ineffective learning behaviors. This problem became obvious in the rating analysis (see dimensions *elaboration on the content* and *elaboration on hints*) and was further corroborated by the narrative analysis. Furthermore, Aristotle showed a better *collaboration flow* and higher *collaborative motivation*, which are important prerequisites for an overall high collaboration quality as was discussed. Also, these dimensions can be regarded as indicators of increased accountability, a goal we had intended to achieve by the jigsaw design of our collaboration script. This interpretation is further supported by the ratings of the mathematical problem-solving process: Aristotle made good use of the social resources and the system resources and overall showed a good problem-solving strategy. On a critical note we have to concede, however, that the results revealed by the case analyses are promising, but we do not know if they would hold for the entire sample. This is a general problem of case methodology: case analyses permit much more fine-grained evaluation of learning processes than could be gained by quantitative cross-conditions comparisons. On the other hand, the generalizability of the results is limited. For instance, the question must be asked how cases were selected. As described above, our selection was dictated by practicality: Due to technical problems, only a few process recordings were complete and of a quality that enabled analysis of students' utterances.

Furthermore, it is possible that students' efforts were not enough to make up for the "collaboration forfeit", that is, the loss of practice opportunities during the learning phase due to the time expenditure of the collaboration. Collaboration often takes more time than individual problem solving and thus can reduce the amount of practice (e.g., Lou et al. 2001; Walker et al. 2008). This problem might have affected particularly the scripted condition as the script directed students in their collaborative activities and asked for more than they would naturally have engaged in when collaborating without script support. Statistical analyses confirm that the number of problems solved during the learning phase differed between conditions, $F(2,40)=8.32$, $p<.01$[2]. More specifically, dyads in the scripted condition solved significantly fewer problems than dyads in the unscripted condition, $t(40)=2.42$, $p=.02$, and taken together dyads in the two collaborative conditions on average solved significantly fewer problems than students in the individual condition, $t(40)=3.31$, $p=.00$ (means and standard deviations of solved problems: scripted condition $M=1.79$, $SD=.80$; unscripted condition $M=3.50$, $SD=1.83$; individual condition $M=4.60$, $SD=2.50$). This finding is also mirrored in the number of problems solved by the two dyads whose learning processes we analyzed: The scripted dyad Aristotle solved only two problems during the learning phase; the unscripted dyad Telemann solved four problems. In other words, students in the collaborative conditions had fewer opportunities to practice the mathematical skills necessary to solve the problems of the reproduction tests than students learning individually, and students in the scripted condition had the fewest opportunities. Furthermore, in related work (Mullins et al. 2011) we found that collaborative settings can encourage students to divide the work, particularly when learning with task types that target procedural skill fluency, and that this type of task distribution negatively affects procedural learning in mathematics. To conclude, although we were not able to show that collaboration and in particular scripted collaboration yielded improved reproduction at posttest, the results show that collaboration is at least as effective as individual learning even when the learning time is held constant. This is true even though the amount of practice in the collaborative conditions was significantly less than the amount of practice in the individual condition; it appears, thus, that the interaction with the learning partner was able to compensate for the loss in practice.

Third, the higher need for assistance in the scripted condition particularly in the collaborative reproduction test could be explained by the increased demands on these students in the test phase: For students in the individual and in the unscripted condition, the problem-solving situation was exactly the same as during the learning phase, but students in the scripted condition were now required, for the first time, to solve system-of-equations problems without script support. As illustrated by the results, the loss of support was particularly severe for students with low prior knowledge, while students with high prior knowledge were able to tackle the problems even though script support was no longer available. Along similar lines, the process analyses of the scripted dyad Aristotle indicate that requesting (and consequently receiving) CTA help just-in-time, when impasses occur, can be a useful learning strategy for students with low prior knowledge. Generally speaking, it could be promising to support students in an adaptive fashion, tailored to their individual and changing needs for help. This hypothesis is supported by related studies in which we were able to demonstrate that intelligent tutoring technologies can be leveraged to provide adaptive tutoring of collaboration, that is, to prompt fruitful collaborative behaviors in relevant moments of the interaction and thus increase student learning (Walker et al. 2009a, b, 2010, 2011; Diziol et al. 2010). The assumption that the higher amount of assistance needed by weaker students in the scripted condition was temporary, due to the new, unscripted problem-solving situation, and not due to inferior learning gains, is supported by the results of the individual reproduction test (which was administered last).

Which conclusions can be drawn regarding the conditions' impact on *future learning*? Students of the individual condition made fewer errors when solving the new problem type (inequality problems) than students of the collaborative conditions; so apparently they were better able to handle the new learning tasks in the CTA learning environment. In fact, this result is not too surprising and is consistent with a phenomenon often reported in the learning sciences: When confronted with a new learning strategy or a new learning environment, students' learning outcome is often reduced initially as they have to abandon previous habits and accustom to the new situation; however, over time and with sufficient training, the advantages can become evident (e.g., Artelt 2000). In the present study, students in the collaborative conditions had to learn how to take advantage of the collaborative learning setting while at the same time being confronted with a new problem type. However, all students had already gained a lot of experience in tackling new problem types with the help of the CTA during regular classroom sessions, which worked in favor of the individual condition. Interestingly, analysis of the future learning test showed that, compared to unscripted collaboration, scripted collaboration helped students to get accustomed to the new collaborative learning situation: The amount of errors made in the future learning test was lower for dyads of the scripted condition than for dyads of the unscripted condition even though script support was no longer available. This gives at least some indication that the guidance of the collaboration script prepared students for the future collaborative learning situation (cf., script as objective, Dillenbourg and Jermann 2007) and that dyads had learned to take advantage of the resources available.

Along these lines, it could be hypothesized that benefits of collaborative learning would increase in future learning situations if collaboration was practiced over longer periods of time, and that this increase would be accelerated if script support was provided to students initially. In other words, in the present study the learning time might have been insufficient to establish differences between conditions large enough to be detected by the statistical analysis. Indications supporting this hypothesis can be found in the study conducted by Berg (1993). She compared scripted collaboration with individual learning in a traditional teacher-dominated classroom structure. The treatment lasted for 30 days in total. Scripted collaboration did not only improve students' learning of the material that was taught during the learning phase, but also their achievement in future chapters that were taught in traditional fashion in both conditions. Moreover, results from another, recent study support this hypothesis: In a collaborative learning study using a similar script approach as the present study, Westermann and Rummel (2012) found significant differences between a collaborative learning condition and a non-collaborative control condition from the second week onwards. The advantage of the collaborative condition continuously increased after the second week until the end of the study in the fourth week.

Outlook

Finally, we would like to note that the present study cannot give final answers regarding the impact of collaboration and in particular of scripted collaboration on student learning. In future research it would be desirable to study the effects of collaborative learning with research designs that span a longer term and more instructional sessions. However, implementing the script over a longer period of time might still result in problems due to overscripting. Thus, adaptive support not only concerning the problem-solving process, but also concerning the collaborative support would still be a desirable goal of future research. Just recently, Walker and colleagues (2011) were able to establish learning benefits of adaptive collaboration support in a peer tutoring setting with the CTA.

The current study was conducted as an in vivo experiment at one of the LearnLab research facilities of the Pittsburgh Science of Learning Center (PSLC, http://learnlab.org): That is, the study was conducted in classrooms, by teachers, during school time. We tried to address criticism brought forward against classic classroom research by trying to execute our study with the same methodological rigorousness we would have used in the lab, and a cautious awareness towards aspects of the situation we could not control in the same way. As reported, during data collection we struggled with "in vivo problems", such as student attrition and a server breakdown during the test day. We addressed these issues in our data analysis and controlled them as much as possible a posteriori. Yet, they may still limit the generalizability of our study results. Furthermore, we might have been unable to establish existing differences between conditions due to the data loss. Our study thus clearly has some limitations. Nevertheless we would like to advocate this type of research in order to achieve the goals Levin (2004, p. 182) formulated for educational research: scientific credibility, contextual "accretability", and educational credibility.

Acknowledgements This research was supported by the Pittsburgh Science of Learning Center, NSF Grant # 0354420, by the Landesstiftung Baden-Württemberg, Germany, and by the Virtual PhD Programm, VGK (DFG). We thank Bruce McLaren for his valuable contributions during initial stages of the project. We are grateful to Jonathan Steinhart, Erin Walker, Dale Walters and Sung-Joo Lim for their support concerning the technical implementations of the Tutor environment; and to Kathy Dickensheets and Lars Holzäpfel for their support in "getting the math right". Further we would like to express our gratitude to the teachers from CWCTC for their motivated involvement in the project. Also, we would like to thank our student research assistants Martina Rau and Katharina Westermann, and Michael Wiedmann, for their help on data coding and data analysis. Special thanks go to Katharina Westermann for her help in preparing this manuscript.

References

Aleven, V., McLaren, B., Roll, I., & Koedinger, K. R. (2004). Toward tutoring help seeking: Applying cognitive modelling to meta-cognitive skills. In J. C. Lester, R. M. Vicari, & F. Paraguaçu (Eds.), *Proceedings of Seventh International Conference on Intelligent Tutoring Systems, ITS 2004* (pp. 227–239). Berlin: Springer.

Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *Journal of the Learning Sciences, 4*(2), 167–207.

Aronson, E., Blaney, N., Sikes, J., Stephan, C., & Snapp, M. (1978). *The jigsaw classroom*. Beverly Hills: Sage.

Artelt, C. (2000). *Strategisches Lernen [Strategic learning]*. Münster: Waxmann.

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, 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. (1993). *Structured cooperative learning and achievement in a high school mathematics class.* Paper presented at the Annual Meeting of the American Educational Research Association, Atlanta.

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.

Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple implications. *Review of Educational Research, 24*, 61–100.

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. (2002). Over-scripting CSCL: The risks of blending collaborative learning with instructional design. In P. A. Kirschner (Ed.), *Three worlds of CSCL. Can we support CSCL* (pp. 61–91). Heerlen: Open Univeriteit Nederland.

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.

Dillenbourg, P., Baker, M., Blaye, A., & O'Malley, C. (1996). The evolution of research on collaborative learning. In P. Reimann & H. Spada (Eds.), *Learning in humans and machines: Towards an interdisciplinary learning science* (pp. 189–211). Oxford: Elsevier/Pergamon.

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, Inc. ISSN 1819-0146

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

Dubinsky, E., Mathews, D., & Reynolds, B. E. (Eds.). (1997). *Readings in cooperative learning for undergraduate mathematics*. Washington: Mathematical Association of America.

Field, A. P. (2005). *Discovering statistics using SPSS* (2nd ed.). London: Sage.

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: Erlbaum.

Huitema, B. E. (1980). *The analysis of covariance and alternatives*. New York: Wiley.

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 (Vol. 1)* (4th ed., pp. 233–265). Boston: McGraw-Hill.

King, A. (2007). Scripting collaborative learning processes: A cognitive perspective. In F. Fischer, I. Kollar, H. Mandl, & J. Haake (Eds.), *Scripting computer-supported collaborative learning. Cognitive, computational, and educational perspectives* (pp. 18–19). New York: Springer.

Koedinger, K. R. (1998, June 5–6, 1998). *Intelligent cognitive tutors as modeling tool and instructional model*. Paper presented at the NCTM Standards 2000 Technology Conference.

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.

Koedinger, K. R., Corbett, A. T., Ritter, S., & Shapiro, L. J. (2000). *Carnegie Learning's Cognitive Tutor ™: Summary Research Results*. Retrieved January 16, 2006, from http://www.carnegielearning.com/approach_research_reports.cfm

Kollar, I., Fischer, F., & Hesse, F. W. (2006). Computer-supported collaboration scripts—a conceptual analysis. *Educational Review, 18*(2), 159–185.

Kollar, I., Fischer, F., & Slotta, J. D. (2007). Internal and external scripts in computer-supported collaborative inquiry learning. *Learning & Instruction, 17*(6), 708–721.

Levin, J. R. (2004). Random thoughts on the (in)credibility of educational-psychological intervention research. *Educational Psychologist, 39*(3), 173–184.

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.

Meier, A., Spada, H., & Rummel, N. (2007). A rating scheme for assessing the quality of computer-supported collaboration processes. *International Journal of Computer-Supported Collaborative Learning, 2*, 63–86.

Mullins, D., Rummel, N., & Spada, H. (2011). Are two heads always better than one? Differential effects of collaboration on students' computer-supported learning in mathematics. *International Journal of Computer Supported Collaborative Learning, 6*(3), 421–443. doi:10.1007/s11412-011-9122-z.

National Council of Teachers of Mathematics. (2006). *Overview of principles and standards for school mathematics.* Retrieved June 5, 2006, from http://www.nctm.org/standards/overview.htm

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). Erlbaum.

Organisation for Economic Co-operation and Development [OECD] (n.d.). *Programme for International Student Assessment.* Retrieved May 25, 2005, from http://www.pisa.oecd.org/pages/0,2966,en_32252351_32235968_1_1_1_1_1,00.html

PSLC. (2006). *Learnlab. Pittsburgh Science of Learning Center.* Retrieved January 16, 2006, from http://learnlab.org Q2

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.

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.

Rummel, N., & Spada, H. (2007). Can people learn computer-mediated collaboration by following a script? In F. Fischer, H. Mandl, J. M. Haake, & I. Kollar (Eds.), *Scripting computer-supported communication of knowledge Cognitive, computational, and educational perspectives* (pp. 39–55). New York: Springer.

Rummel, N., Deiglmayr, A., Spada, H., Karimanis, G., & Avouris, N. (2011). Analyzing collaborative interactions across domains and settings: An adaptable rating scheme. In S. Puntambekar, C. Hmelo-Silver, & G. Erkens (Eds.), *Analyzing interactions in CSCL: Methods, approaches and issues* (pp. 367–390). Berlin: Springer.

Slavin, R. E. (1992). When and why does cooperative learning increase achievement? Theoretical and empirical perspectives. In R. Hertz-Lazarowitz & N. Miller (Eds.), *Interaction in cooperative groups. The theoretical anatomy of group learning* (pp. 145–173). New York: Cambridge University Press.

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.

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. (2008). To tutor the tutor: Adaptive domain support for peer tutoring. In B. P. Woolf, E. Aïmeur, R. Nkambou, & S. P. Lajoie (Eds.), *Proceedings of the Ninth International Conference on Intelligent Tutoring Systems (ITS 2008)*, *Lecture Notes in Computer Science, Vol. 5091* (pp. 626–635). Springer, ISBN 978-3-540-69130-3

Walker, E., Rummel, N., & Koedinger, K. (2009a). CTRL: A research framework for providing adaptive collaborative learning support. *User Modeling and User-Adapted Interaction: The Journal of Personalization Research (UMUAI), 19*(5), 387–431.

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

Walker, E., Rummel, N., & Koedinger, K. (2010). Automated adaptive support for peer tutoring in high-school mathematics. 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. 151–153). International Society of the Learning Sciences, Inc.

Walker, E., Rummel, N., & Koedinger, K. (2011). Designing automated adaptive support to improve student helping behaviors in a peer tutoring activity. *International Journal of Computer Supported Collaborative Learning, 6*(2), 279–306.

Webb, N. M., Troper, J. D., & Fall, R. (1995). Constructive activity and learning in collaborative small groups. *Journal of Educational Psychology, 87*(3), 406–423.

Wecker, C., Kollar, I., Fischer, F., & Prechtl, H. (2010). Fostering online search competence and domain-specific knowledge in inquiry classrooms: Effects of continuous and fading collaboration scripts. 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. 1 (pp. 810–817). International Society of the Learning Sciences, Inc.

Westermann, K., & Rummel, N. (2012). Delaying instruction—Evidence from a study in a university relearning setting. *Instructional Science*. doi:0.1007/s11251-012-9207-8.