Educ Psychol Rev (2018) 30:531–558
DOI 10.1007/s10648-017-9424-4

INTERVENTION STUDY

The Relative Merits of Explicit and Implicit Learning of Contrasted Algebra Principles

Esther Ziegler[1,2] · **Peter A. Edelsbrunner**[1] · **Elsbeth Stern**[1]

Published online: 17 September 2017

Abstract Knowledge representations that result from practicing problem solving can be expected to differ from knowledge representations that emerge from explicit verbalizing of principles and rules. We examined the degree to which the two types of learning improve problem-solving knowledge and verbal explanation knowledge in classroom instruction. We presented algebraic addition and multiplication problems to 153 sixth graders randomly assigned to two conditions. Students in the explicit learning condition had to verbally compare contrasted algebra problems. Students in the implicit learning condition had to generate and solve new problems. On three follow-up tests over 10 weeks, students in the explicit learning condition exhibited better problem-solving knowledge than students in the implicit learning condition, as well as some advantages in verbal concept knowledge. Implicit learning showed some advantages on not directly taught but incidentally learned aspects. Overall, this outcome favors the explicit learning of concepts. Explicit comparison fostered student performance on non-verbal and verbal measures, indicating that verbalization facilitates effective comparison.

Keywords Explicit comparison · Problem solving · Explicit and implicit learning · Verbal explanation · Mathematics learning

The proverb "practice makes perfect" has long been a mantra in mathematics education, and it is understood that knowledge develops by continuously working on mathematical problems. However, less agreement exists concerning the instructions that students should receive when

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10648-017-9424-4) contains supplementary material, which is available to authorized users.

✉ Esther Ziegler
esther.ziegler@gess.ethz.ch

1 Institute for Research on Learning and Instruction, ETH Zurich, Zurich, Switzerland

2 Institute for Learning Sciences and Higher Education, ETH Zurich, Zurich, Switzerland

working on problems. Should students be explicitly directed to verbalize the underlying rules and principles during problem solving, or should they have the opportunity to implicitly acquire the rules as a consequence of extensive problem solving (Aleven & Koedinger, 2002; Matthews & Rittle-Johnson, 2009)?

Knowledge representations that result from practicing problem solving can be expected to differ from knowledge representations that emerge from the instruction to verbalize the underlying principles and rules. In the present study on early algebra learning, we examine the effects of the two types of learning opportunities, one type promoting implicit and the other promoting explicit learning. A broad set of outcome measures is included, allowing a detailed assessment of the scenarios in which both learning opportunities have better pay off. For both learning opportunities, we arrange a contrasted presentation of addition and multiplication algebra problems because doing so has proven successful in prior studies, as described below. Subsequently, we discuss in detail the potential benefits and shortcomings of implicit and explicit learning.

The Potential of Contrasted Algebra Concept Learning

Mathematical concepts can encompass multiple features, rules, principles, and procedures. In mathematical settings, different concepts are traditionally introduced in succession with sequenced or blocked processing of concepts in the form "$a1a2a3b1b2b3c1c2c3$." This blocked learning, however, can lead to considerable confusion around the concepts introduced (Rohrer, 2012; Ziegler & Stern, 2014, 2016). This problem is especially pronounced if concepts are similar or highly related, as is often the case in mathematics. For example, the continuity of addends in algebraic addition "$ab + a + ab + a = a + a + ab + ab = 2a + 2ab$" needs to be clearly distinguished from the splitting up of multiplicands in multiplication "$ab \cdot a \cdot ab \cdot a = a \cdot b \cdot a \cdot a \cdot b \cdot a = a^4b^2$".

Two reasons for confusion caused by blocked concept learning might be considered. The first reason is the emphasis on superficial automation. When students repeatedly work on problems related to a single concept, they might successfully induce the presented computing procedures without necessarily grasping the underlying principles. In that situation, successful induction allows efficient automated execution, albeit without having developed a deep understanding of the concepts. This lack of understanding can hinder the successful transfer of the newly gained knowledge to novel problems (Kamii & Dominick, 1997; Sweller, 1994). The second reason for confusion is the problem of interferences. When students learn similar concepts in succession, they might confuse those concepts because they have established associations to common elements (Anderson, 1996; McCloskey & Cohen, 1989). This problem indicates that blocked learning might be adequate mostly for simple, unconnected concepts. When attempting to establish broadly connected mathematical concept knowledge, blocked learning might be less adequate.

One approach to overcoming superficial automation and interferences is mixing, in which problems based on different concepts are presented at the same time. Mixing can be implemented in different ways. In interleaved processing ("$a1b1a2b2a3b3$"), problem type a is alternated with problem type b. In simultaneous or contrasted processing ("$a1b1$, $a2b2$, $a3b3$"), problem types a and b are juxtaposed. Interleaved processing has been found to be beneficial for learning mathematical formulas and rules (Rohrer & Taylor, 2007; Taylor & Rohrer, 2010). Following these findings indicating advantages of alternating problem types, Kang and Pashler (2012) tested a condition with simultaneous processing, implementing the juxtaposed presentation of different painting styles. They supposed that offering the opportunity for spatial comparison would generate additional learning gains. However, this condition revealed no advantage in comparison with interleaved processing. Presenting concepts simultaneously and interleaved thus appears to lead to similar results, both offering the opportunity for implicit comparison of the concepts, which is considered to facilitate discrimination between the concepts (Birnbaum et al., 2013; Carvalho & Goldstone, 2014). These results indicate that in simultaneous presentation, the principles may also be processed in an interleaved way, without actively connecting the juxtaposed concepts. Thus, mere juxtaposition does not necessarily and automatically involve explicit thinking about the concepts' structure, but it can be viewed as an implicit process of learning different concepts (DeKeyser, 2003; Koedinger & Aleven, 2007). Simultaneous processing might primarily promote fluent and accurate problem solving, with an incidental registration of features and rules.

In the juxtaposition of two concepts, the opportunity to compare is offered, and comparisons may occur implicitly (Mitchell et al., 2008). However, solely offering the opportunity for comparisons may not be sufficient for eliciting full comparison benefits. Rather, comparison of juxtaposed problems should be triggered, for example, by prompting students to describe the similarities between concepts or to answer comparison questions (Kurtz et al., 2001; Schwartz & Bransford, 1998). In explicit comparison, juxtaposed concepts are then actively linked by verbally expressing their similarities and differences. Thus, explicit comparison can be viewed as a conscious, verbal process of learning different concepts that is stimulated when students are instructed to direct their attention to underlying principles (Renkl, 2015). Consequently, verbal representations of the concepts that primarily promote the knowledge needed to describe and explain how relevant problems are solved may be constructed, and there can be positive influence on transferring the acquired knowledge to novel problems (Aleven & Koedinger, 2002; Rittle-Johnson, 2006).

Learning with contrasted material has frequently proven to be beneficial (Alfieri et al., 2013; Gentner, 2010; Rittle-Johnson & Star, 2011). Whether the learning of contrasted concepts occurs more effectively through explicit verbalization or the degree to which it can be stimulated through extensive problem-solving practice remains uncertain.

Implicit and Explicit Approaches to Learning

The implicit and explicit approaches to learning differ in terms of whether or not students' attention is directed to the underlying structure of the learning materials. Implicit learning is non-intentional learning, in which learners' attention is not directed to rules and principles. In mathematics instruction, this type of learning occurs when receiving an instruction to complete a task or to solve a problem that does not emphasize the relevant structure or explicitly direct one's attention to principles. Thus, knowledge develops through repeated application, with little deliberate awareness of principles and rules, and requiring little mental effort (Aleven & Koedinger, 2002; Koedinger et al., 2012). Such problem-solving skills gradually improve with practice and are honed through success/failure feedback each time that they are employed (Anderson, 1995; DeKeyser, 2003). In their knowledge-learning-instruction (KLI) framework, Koedinger et al. (2012) describe this type of learning as a non-verbal induction and refinement process that modifies the conditions that control the retrieval and application of knowledge. Several learning mechanisms are involved, such as perception, generalization, discrimination, classification, schema induction, and causal induction. These processes refine knowledge, for example, by adding relevant and removing irrelevant features of representations; as such, they render knowledge more accurate, general, and discriminating. The resulting knowledge is that of cues and conditions for choosing steps or the actions needed to solve a problem or to complete a task (DeKeyser, 2003). Thus, it is important to provide instances that include all of the relevant and less relevant features to gain a complete representation of the concepts. Implicit learning leads to the automation and fluency of skills, thus supporting successful short-term performance (Koedinger et al., 2012). During implicit learning, however, students might also become aware of rules and explicitly learn things without being prompted to focus their attention on these aspects. The initially implicitly acquired knowledge can continuously develop toward more explicitly available knowledge (Karmiloff-Smith, 1994; Siegler & Stern, 1998; Sun et al., 2005). The question arises to which degree students become aware of rules and principles if they are not explicated and to which degree they are able to explicate such knowledge.

Explicit learning is intentional, deliberate learning that aims at making learners aware of rules and regularities and therefore requires more mental effort than implicit learning. Explicit learning can be triggered by directing student attention toward the structure of concepts. This type of learning can be achieved by asking students to verbally explain the learning materials. For example, in mathematics learning through explanation, students are prompted to search for verbal rules or analog examples and to interpret these elements (Aleven & Koedinger, 2002; Gentner et al., 2009). Such verbalizations, with the attempt to explicitly map rules onto given instances, can help learners focus on critical and relevant features. Positive effects on learning through verbal explanations have been demonstrated regarding different types of materials and among students of different ages (Atkinson et al., 2003; Chi, 2000; Renkl, 1997; Rittle-Johnson, 2006; Wong et al., 2002). Koedinger et al. (2012) describe this type of learning as an understanding and sense-making process, verbally mediated learning and thinking with the aim of comprehending or reasoning. Several learning mechanisms are likely involved such as comprehension strategies, verbal explanation learning, discovery learning, and verbal rule-mediated deduction. Explicit learning has shown broad benefits for concept learning (Aleven & Koedinger, 2002). In general, explicit learning occurs relatively slowly due to the explication process but leads to lasting results (Baroody et al., 2007; Koedinger et al., 2012). Through its application, explicitly acquired knowledge can continuously develop toward more automated and implicitly available knowledge.

Different word pairs are used to label implicitly and explicitly acquired knowledge: implicit and explicit knowledge refer to the level of consciousness during knowledge acquisition (DeKeyser, 2003; Sun et al., 2005); procedural and conceptual (or declarative) refer to the representation of knowledge as skills and strategies or as facts and principles (Aleven & Koedinger, 2002; Anderson, 1996; Rittle-Johnson et al., 2015); and non-verbal and verbal refer to the degree of explication required while working on problems (Koedinger et al., 2012). Sometimes these three approaches to labeling knowledge are used almost synonymously; they do however denote different characteristics of knowledge.

The acquisition of one type of knowledge is not limited to the corresponding learning opportunity, a fact that may be overlooked when using dichotomies to describe learning and knowledge. For example, learning opportunities that are supposed to improve procedural knowledge can also benefit conceptual knowledge, and vice versa (Aleven & Koedinger, 2002; Rittle-Johnson et al., 2015; Schneider & Stern, 2010). Also, asymmetries are reported, such that explicit learning opportunities benefit both types of knowledge, but implicit learning solely benefits procedural knowledge (Aleven & Koedinger, 2002; Koedinger et al., 2012; Rittle-Johnson, 2006). Non-verbal problem-solving practice is indeed effective for improving problem-solving knowledge. Students typically perform less well on verbal knowledge measures following implicit problem solving learning than they do following explicit verbal explanation learning (Aleven & Koedinger, 2002). Intense verbal explanation training can, however, bring gains on verbal explanation measures, and also on problem-solving measures (Rittle-Johnson et al., 2015; Schneider & Stern, 2010).

To further disentangle how both types of learning influence both types of knowledge, an experimental design in which the processing time is kept constant between instructional conditions that emphasize either problem-solving practice or verbal explanation training would be necessary (Aleven & Koedinger, 2002; Matthews & Rittle-Johnson, 2009). Whether implicit or explicit learning occurs and which kind and amount of knowledge is gained are not directly observable (Koedinger et al., 2012). This effect can be indirectly assessed by providing non-verbal tasks (e.g., problem solving) or verbal tasks (e.g., explaining concepts and procedures). In the present study, we used several measures of problem-solving knowledge and verbal explanation knowledge to assess student learning.

Implicit and explicit learning might also differ in their impact on learning outcomes in ways that are not always taken into account. Classroom instruction mainly focuses on guiding learner attention toward the principles and rules that are mandatory for learning a concept. In addition to these directly taught features of the concept, students might incidentally learn side aspects of the material (e.g., alphabetical ordering in algebra to keep an overview of the solution processes). Such aspects are not mandatory for solving problems, but they can be helpful for making neat arrangements. Implicit and explicit learning might differ with regard to how well they support the learning of features that were not the direct focus of the learner's attention, and consequently, implicit learning might be favored (Ziegler & Stern, 2014; Williams & Lombrozo, 2010). Thus, it remains unclear how directly instructed principles and incidentally learned aspects of material are influenced by whether students learn implicitly via problem solving or explicitly via verbal explanations.

The Present Study

In the present study, we examine whether the introduction of contrasted algebra material reveals its potential when it is combined with explicit learning via verbalizing the differences between problems or when it is combined with implicit learning in the form of extensive problem solving. We examine the differential effects of these two conditions on a broad set of verbal explanation and non-verbal problem-solving measures.

For the explicit learning condition, we considered a former dataset with students who were trained via explicit verbal explanations from Ziegler and Stern (2014). In this condition, the students were prompted to find and explicitly extract principles and rules by studying examples and generating verbal explanations. In the implicit learning condition, we used the original contrasted material regarding algebraic addition and multiplication and designed a new problem-solving activity. In this condition, instead of studying examples and generating verbal explanations, the students had to generate and solve their own new problems. Thus, the students were expected to automate the underlying rules by applying them without being explicitly directed to the concepts' structures.

Our study contributes to the literature in various ways. First, only few controlled classroom studies have been conducted (e.g., Aleven & Koedinger, 2002), and these were with students using a cognitive tutor in addition to other classroom activities. In the present study, we implemented an intense, well-controlled classroom instruction that extended over 4 days in which students processed the full instruction material either explicitly or implicitly. Second, among classroom studies many examine the influence of one of the two types of learning on different types of knowledge (e.g., Rittle-Johnson & Star, 2007; Star & Rittle-Johnson, 2009). However, only a few studies have combined both explicit and implicit learning approaches with both verbal and non-verbal knowledge measures (e.g., Aleven & Koedinger, 2002). The present study includes well-balanced explicit and implicit learning materials with a broad assessment of verbal explanation and problem-solving knowledge. Third, differences in learning gains when learning is juxtaposed with concept material have been found for delays including immediately after learning, 1 h after, 1 week after, and up to 4 weeks after (e.g., Rau et al., 2013; Rittle-Johnson et al., 2012). The present study extends these testing periods by including three measurement points up to 10 weeks after instruction, with the aim of increasing the meaningfulness for school learning.

For non-verbal problem-solving knowledge measures, we expected an outperformance of the explicit learning condition for problem-solving knowledge in which the students had to solve contrasted addition and multiplication problems. Algebraic term transformation requires distinguishing between addition and multiplication principles. Thus, although our implicit learners received extensive practice in problem solving of juxtaposed materials, we expected the deliberate verbalization of solution procedures to be more beneficial (Aleven & Koedinger, 2002; Atkinson et al., 2003; Rittle-Johnson, 2006). However, we expected advantages for the implicit learning condition for the application of conventional aspects that were not directly taught but could be incidentally learned. Indirectly taught conventions encompassed those of sorting letters in alphabetical order, e.g., 3ab, not 3ba, and writing an unnecessary number 1, e.g., b, not 1b. Implicit learning is relatively inaccurate but performs well in the short term (Koedinger et al., 2012; Sun et al., 2007). Therefore, we expected an advantage of the implicit learning condition in the short term and a greater decline in the long term.

For all of the verbal explanation knowledge measures, we expected an outperformance of the explicit learning condition in which the students repeatedly verbally explained the contrasted problems. Explicit learning leads to an improved verbal concept representation that occurs in a delayed manner (Koedinger et al., 2012). Therefore, we expected the explicit learning condition to show long-term advantages for verbal concept knowledge. We also expected an advantage of the explicit learning condition for the conventions not directly taught, even though these conventions are implicitly acquired. We assumed that it was more likely that students in the explicit condition would learn these features because they verbalized the concepts, making their awareness and processing of these features more likely.

Methods

Participants

Sixth graders with no prior formal instruction in algebra were chosen as study participants. In the Swiss mathematics curriculum, algebra is not introduced until secondary school, which corresponds to grade 6 or 7 in the USA. The participants were recruited from six urban and suburban public schools in middle class neighborhoods in the canton of Zurich/Switzerland. To acknowledge the focus on self-learning and because of time restrictions in this controlled experimental study, the teachers were asked not to choose students (a) with insufficient German language comprehension, (b) with special needs, or (c) who were unable to fulfill the minimum standards of school performance. All of the students were volunteers, and their parents provided written consent. Every class was rewarded with 200 Swiss francs (approximately US$200), and each student received a small gift.

A total of 155 students from eight classes participated. The students were randomly assigned to one of two conditions within their class cohorts. Thus, the students were instructed in mixed groups, although the different materials appeared highly similar to them. The explicit learning condition was composed of 79 students ($M = 12.4$ years, $SD = 0.5$; 44 females), and the implicit learning condition was composed of 74 students ($M = 12.3$ years, $SD = 0.5$; 41 females). Two students who did not finish the intervention were excluded. In addition, one student was absent at the second and another was absent at the third follow-up assessment.

Design and Procedure

In a 2 (condition: explicit learning, implicit learning) × 3 (time: 1 day, 1 week, 10 weeks after instruction) mixed-factorial design with repeated measures, we investigated the effects of explicit (verbal explanation) and implicit (problem solving) learning of contrasted algebra material on sixth graders' learning using problem-solving and verbal knowledge measures.

Each student participated in four intervention sessions and three follow-up sessions, as listed in Table 1. In both conditions, the students participated in 90-min intervention sessions on four consecutive days, during which they worked through a self-study program with nine worksheets.

Table 1 Overview of the activities of each session

	Session	Duration	Activities
Intervention sessions	1st day (Mon)	2 lessons	Pretest—prior algebra knowledge (5 min) Introduction—short slide presentation (5 min) Session 1: work sheets and immediate learning tests 1–3
	2nd day (Tue)	2 lessons	Repetition test (5 min) Session 2: work sheets and immediate learning tests 4 + 5
	3rd day (Wed)	2 lessons	Repetition test (5 min) Session 3: work sheets and immediate learning tests 6 + 7
	4th day (Thu)	2 lessons	Repetition test (5 min) Session 4: Work sheets and immediate learning tests 8 + 9
Follow-up sessions	1 day later	2 lessons	Follow-up session 1: "problem solving test" and "verbal concept test" (45 min) Survey part: personal data, reasoning test, and arithmetic test (45 min)
	1 week later	1 lesson	Follow-up session 2: "problem solving test" and "verbal concept test" (45 min)
	10 weeks later	1 lesson	Follow-up session 3: "problem solving test" and "verbal concept test" (45 min)

The intervention occurred in groups of 10–15 students in rooms at the school. The students worked individually on their learning programs and sat far enough from their classmates that they could not look at one another's worksheets. They were instructed to work independently and directly ask the instructors if they had questions about or problems with the material. All of the groups were trained by the first author, who was educated as a primary school teacher and was continuously present, along with a research assistant, to guide the intervention and testing. To control for a fair intervention for both conditions, we included an implementation check (see page 16).

Learning Materials

The instructional material consisted of nine worksheets with worked examples composed of algebraic expressions that required simplification using algebraic transformation strategies (for an excerpt, see Fig. 1). The students worked on these worksheets within four learning sessions. In the initial session, because the material required self-learning, a short presentation consisting of three slides was shown to demonstrate how to read and write terms with letters and how to use the mathematical expression "raise to the power of." No other whole-group instruction was provided. Rather, the students in both conditions were asked to independently derive the principles from the worked examples and extend their knowledge using the self-study activities. Worked examples are acknowledged as effective learning material (Renkl, 2005;

A3	$xy + xy + xy = 3 \cdot xy = 3xy$ $2b + 2b + 2b + 2b + 2b = 5 \cdot 2b = 10b$ $3cx + 3cx = 2 \cdot 3cx = 6cx$	M3	$xy \cdot xy \cdot xy = x \cdot y \cdot x \cdot y \cdot x \cdot y = x \cdot x \cdot x \cdot y \cdot y \cdot y = x^3 \cdot y^3 = x^3y^3$ $2b \cdot 2b \cdot 2b \cdot 2b \cdot 2b = 2 \cdot b \cdot 2 \cdot b \cdot 2 \cdot b \cdot 2 \cdot b \cdot 2 \cdot b = 2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 \cdot b \cdot b \cdot b \cdot b \cdot b = 32 \cdot b^5 = 32b^5$ $3cx \cdot 3cx = 3 \cdot c \cdot x \cdot 3 \cdot c \cdot x = 3 \cdot 3 \cdot c \cdot c \cdot x \cdot x = 9 \cdot c^2 \cdot x^2 = 9c^2x^2$
A4	$c^2 + c^2 + c^2 + c^2 = 4 \cdot c^2 = 4c^2$ $a^4 + a^4 = 2 \cdot a^4 = 2a^4$ $x^3 + x^3 + x^3 = 3 \cdot x^3 = 3x^3$	M4	$c^2 \cdot c^2 \cdot c^2 \cdot c^2 = c \cdot c \cdot c \cdot c \cdot c \cdot c \cdot c \cdot c = c^8$ $a^4 \cdot a^4 = a \cdot a \cdot a \cdot a \cdot a \cdot a \cdot a \cdot a = a^8$ $x^3 \cdot x^3 \cdot x^3 = x \cdot x \cdot x \cdot x \cdot x \cdot x \cdot x \cdot x \cdot x = x^9$
A5	$2x + 5x + 2x = 9x$ $3bc + bc + 6bc = 3bc + 1bc + 6bc = 10bc$ $y^3 + 4y^3 = 1y^3 + 4y^3 = 5y^3$	M5	$2x \cdot 5x \cdot 2x = 2 \cdot x \cdot 5 \cdot x \cdot 2 \cdot x = 2 \cdot 2 \cdot 5 \cdot x \cdot x \cdot x = 20 \cdot x^3 = 20x^3$ $3bc \cdot bc \cdot 6bc = 3 \cdot b \cdot c \cdot b \cdot c \cdot 6 \cdot b \cdot c = 3 \cdot 6 \cdot b \cdot b \cdot b \cdot c \cdot c \cdot c = 18 \cdot b^3 \cdot c^3 = 18b^3c^3$ $y^3 \cdot 4y^3 = y \cdot y \cdot y \cdot 4 \cdot y \cdot y \cdot y = 4 \cdot y \cdot y \cdot y \cdot y \cdot y \cdot y = 4 \cdot y^6 = 4y^6$
A6	$m + m + a + m + a + m = a + a + m + m + m + m = 2 \cdot a + 4 \cdot m = 2a + 4m$ $4 + x + z + x + 4 + x = x + x + x + z + 4 + 4 = 3 \cdot x + 1 \cdot z + 8 = 3x + z + 8$	M6	$m \cdot m \cdot a \cdot m \cdot a \cdot m = a \cdot a \cdot m \cdot m \cdot m \cdot m = a^2 \cdot m^4 = a^2m^4$ $4 \cdot x \cdot z \cdot x \cdot 4 \cdot x = 4 \cdot 4 \cdot x \cdot x \cdot x \cdot z = 16 \cdot x^3 \cdot z^1 = 16x^3z$

Fig. 1 Worked examples: addition learning steps (*A3–A6*) and multiplication learning steps (*M3–M6*). *Gray cursive* the intermediate steps to the solution that were marked red in the original version; *black bold* the problems and the results

Sweller, 2006). They provide problems with their solution steps, and in doing so, they offer a good basis for students to derive new principles on their own. Worked examples serve equally well as a basis both for implicit learning by imitating the examples and practicing similar problems and for explicit learning by verbally explaining the examples and their solutions.

Worksheets

The worksheets consisted of worked examples, a self-study section, and a practice section. The worked examples and the practice section were equal; only the self-study section differed between the two conditions.

Worked Examples The students in both conditions received worksheets with the same worked examples (for the development of the material, see Ziegler & Stern, 2014). At the top of each worksheet, the students were presented with two blocks of two to three worked algebra examples in a contrasted format. The block on the left contained worked addition problems, and the block on the right contained the same problems presented as multiplication problems (Fig. 2). The worked examples provided students with examples of intermediate steps for arriving at a solution and therefore they included the problem, the solution steps, and the final solution. The addition problems were identical to the multiplication problems and differed only in their operation signs, which naturally impacted the solution steps and results. The students' task was then to learn the underlying principles and rules that the worked examples contained by processing the self-study section. An example of an addition core principle is the continuity of summands such as "x" or "xy" ($xy + x + xy + x = x + x + xy + xy = 2x + 2xy$); an example of a multiplication core principle is the splitting of factors ($xy \cdot x \cdot xy \cdot x = x \cdot y \cdot x \cdot x \cdot y \cdot x = x^4y^2$; also see the Additional file 1).

Self-Study Section This section on each worksheet differed between the two conditions. For the *explicit learning condition*, we used the contrasted learning materials from a previous investigation (Ziegler & Stern, 2014). Students in the explicit learning condition received instructions to explicitly direct their attention to the concepts' structures. They

Addition **A3**	**Multiplication** **M3**
xy + xy + xy $= 3 \cdot xy$ **= 3xy**	**xy · xy · xy** $= x \cdot y \cdot x \cdot y \cdot x \cdot y$ $= x \cdot x \cdot x \cdot y \cdot y \cdot y$ $= x^3 \cdot y^3$ **= x^3y^3**
2b + 2b + 2b + 2b + 2b $= 5 \cdot 2b$ **= 10b**	**2b · 2b · 2b · 2b · 2b** $= 2 \cdot b \cdot 2 \cdot b \cdot 2 \cdot b \cdot 2 \cdot b \cdot 2 \cdot b$ $= 2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 \cdot b \cdot b \cdot b \cdot b \cdot b$ $= 32 \cdot b^5$ **= $32b^5$**
3cx + 3cx $= 2 \cdot 3cx$ **= 6cx**	**3cx · 3cx** $= 3 \cdot c \cdot x \cdot 3 \cdot c \cdot x$ $= 3 \cdot 3 \cdot c \cdot c \cdot x \cdot x$ $= 9 \cdot c^2 \cdot x^2$ **= $9c^2x^2$**

Fig. 2 Worked examples in a contrasted format. *Gray italic* the intermediate steps to the solution that were marked red in the original version; *black bold* the problems and the results

were prompted to compare and explain the contrasted addition and multiplication problems and record how to solve these problems. To elicit explicit comparisons and obtain verbal representations of the principles, the students were prompted to verbalize differences and explicitly describe the underlying rules. The students' self-explanations were checked by the research leader. Because the students worked on similar materials for several days, we provided minimal feedback to prevent them from both exerting only minimal effort and incorrectly learning the concepts. We attempted to keep the feedback as similar as possible between the two conditions by limiting it as follows: We provided a standardized correct/incorrect and sufficient/insufficient feedback. If the explanations were incorrect, then the students were required to correct them. If the explanations were too short, then the students were asked to complete them in greater detail before proceeding. No other feedback was given. Because the material consisted of worked examples, all of the students were able to find correct and more detailed explanations in a second attempt. Further, the contents were resumed at the beginning of the next unit so that the students received repeated opportunities to learn the material over the entire intervention. Students in the *implicit learning condition* processed the same contrasted worked examples as in the other condition (Figs. 1 and 2), but they were not prompted to explicitly compare the examples or to verbalize the differences between the addition and multiplication examples (i.e., their attention was not directed to the concepts' structures). We still wanted to have the generating component in the tasks of both conditions. Therefore, while the students in the explicit condition studied examples and generated verbal explanations, the students in the implicit conditions generated and solved their own problems. Thus, the focus in the implicit condition shifted from careful analysis and explanation of principles to their application. The goal was to induce a non-verbal representation of the principles. To keep the time between both conditions constant, we examined how many examples students were able to solve and generate (on average) in the time that the verbal explanation students used to write down their explanations via pilot testing. Thus, the students were provided two to three sentences introducing the new worked examples and were then prompted to generate five to six of their own examples and solve them. The newly generated examples should be based on the model of the presented examples so that the students would experience and re-apply the underlying rules. The students were also asked to record the intermediate steps to the solution and to invent varied and notable examples using other numbers and letters. Similar to the explicit learning condition, the implicit condition involved standardized feedback. The generated examples were checked by the research leader; if they were incorrect, then the students were required to correct them before continuing to the subsequent section. If the examples provided by the students were too short or were poorly generated, then the students were asked to supplement the examples. However, as with the explicit condition, no other feedback was given.

Practice Problems On the back of the worksheets, the students were provided practice problems to deepen their understanding of the learned principles. Students in both conditions were provided the same four to six problems per sheet with solution instructions to record the intermediate steps of the solution. Before the research leader corrected the practice problems, the students independently reviewed the problems using the examples on the front page. Then, the students received feedback on whether they were correct/incorrect and had to correct the incorrect problems. The internal consistency of the practice problems in our sample was moderate, Cronbach's $\alpha = .76$.

Immediate Learning Tests

After each worksheet, the students completed a learning test sheet with three to eight problems without the use of the instruction materials in order to examine their learning progress. In total, there were 58 items. These problems were similar to those presented on the worksheet that the students had completed. The students did not receive any feedback on these problems, which were corrected only later for analysis. The internal consistency of the immediate learning test problems in our sample was high, Cronbach's $\alpha = .86$.

Implementation Check

To ensure fair instructions for both conditions, the students' accuracy in solving the *practice problems* and the *immediate learning tests* served as an implementation check. The percentages of correctly solved practice and immediate learning test problems were assessed and used as an indicator that students could learn the algebra principles under both self-study conditions. The analysis showed no difference between the conditions with regard to the accuracy of solving the *practice problems* (explicit learning condition: M = 84.6%, SD = 10.3; implicit learning condition: M = 83.2%, SD = 10.8), $t(1, 151) = 0.86$, $p = .389$, and the accuracy of completing the *immediate learning tests* (explicit learning condition: M = 81.8%, SD = 13.0; implicit learning condition: M = 79.1%, SD = 12.1), $t(1, 151) = 1.37$, $p = .173$. The similar achievement indicates that there were equal challenges and comparable student engagement in the intervention activities for the two conditions.

Assessment of Students' Preconditions

Prior Algebra Knowledge

Although Swiss sixth graders have not yet received any formal algebra instruction, it is possible that some could spontaneously solve algebraic transformation problems by referring to their arithmetic knowledge. Therefore, the students were required to take a test assessing prior algebra knowledge at the beginning of the intervention. The test was composed of eight algebra problems requiring transformations: "$a + a + a + a =$," "$5 + a + a + 5 + a =$," "$c \cdot c \cdot c =$," "$2 \cdot 2 \cdot z \cdot 2 \cdot z =$," "$7b + 7b =$," "$7b \cdot 7b =$," "$ab \cdot 4ab =$," and "$xy + xy + xy + xy =$."

Control Variables

To ensure random assignment to the two conditions, the students' individual characteristics, which are considered to affect algebra learning and testing, were assessed with four measures. The students' mathematical school achievement was measured in the form of their *grades in mathematics* and *German*. Mathematics was chosen because of the mathematical material and German because the verbal concept test required verbally explaining the algebraic principles. The grades were reported by the students' teachers. *Arithmetic knowledge* was included to assess the automation of basic mathematical knowledge. It was assessed with two speed tests consisting of two sheets, each with 28 arithmetic items. The students were required to solve each sheet within 90 s and were allowed to write only the results. The first sheet contained two-digit additions, the second sheet two-digit multiplications. The arithmetic knowledge score was determined by the number of correct answers. *Reasoning ability* is a teacher-independent measure of information processing capacity. Because the programs were self-learning material and the algebra rules had to be derived from worked examples in both conditions, reasoning was considered to be an important factor. It was assessed with a figural and numerical subtest of the German intelligence test (subtests 3 and 4 of the LPS by Horn, 1983), which is based on Thurstone's primary mental abilities test.

Problem-Solving Test

The *problem-solving* test was a measure of problem-solving knowledge that encompassed 58 items to assess students' abilities to correctly apply the transformations practiced in intervention sessions 1–4 (e.g., "$a^2 \cdot a \cdot ay \cdot 4a =$," "$5ab + b + 3b + 2ab + 2b =$," "$y \cdot y^3 \cdot y^2 \cdot y =$," "$2 + 5x + 4 + 2x + 3 =$"). In the test, the example problems from the worksheets were represented, but with other variables and numbers. Each item was an algebraic term that needed to be transformed into the shortest version by applying the correct rules. The items were ordered by increasing difficulty. Because we did not expect the students to remember the items, the identical problem-solving test was used for all three of the follow-up assessments. Based on students' test answers, we assessed two kinds of knowledge. First, covering knowledge that was directly taught in the self-learning sessions, we assessed students' *problem-solving knowledge*. Second, we assessed how often the students stuck to two types of *conventions* (i.e., sorting letters in alphabetical order (e.g., *3ab*, not *3ba*) and writing an unnecessary number one (e.g., *b*, not *1b*)). These conventions were not directly taught in the self-learning materials, but they were obvious from the worked examples and could thus be learned incidentally.

Problem-Solving Knowledge

This knowledge was determined by the number of correct answers on the 58 algebra items, with a maximum of 58 points. For an incorrectly solved item, one point was deducted from the total score. The internal consistency of the problem-solving test in our sample was high for all measurement points, Cronbach's α = .93 at T1, .93 at T2, and .94 at T3. In addition, the problem-solving test was analyzed for *careless errors*. These errors should be independent of the intervention and are therefore seen as a measure to examine for equality of conditions. Careless errors were mistakes such as miscounting the number of letters, for example, "$n \cdot n \cdot n \cdot n \cdot n = n^4$," or arithmetic errors, for example, "$b \cdot a \cdot 4 \cdot a \cdot 4 \cdot a = 18a^3b$."

Convention Knowledge

This knowledge was assessed by how often the students stuck to the alphabetical-order and number-one conventions in the problem-solving test. All of the problems were analyzed for the use of these conventions. These conventions were not directly taught, and students never received feedback concerning their use of conventions; thus, they were not counted as errors of problem-solving knowledge but as two separate convention knowledge measures. These conventions are also not mandatory to learning and distinguishing between algebraic addition and multiplication. Nevertheless, they help provide an overview when processing complex algebraic expressions and therefore, they are practical for learning. While processing the worksheets, the students could incidentally learn the conventions not directly taught. Thus, students' usage of the conventions indicates whether they also processed superficial characteristics while focusing on structures.

The *alphabetical-order convention* is an agreement to alphabetically sort letters, which provides a better overview when there are many variables—for example, "$u^2 \cdot ax \cdot u^2 \cdot u \cdot ax = a^2u^5x^2$" or "$n + b + n + x + b + n = 2b + 3n + x$." Student use of the conventions not directly taught was determined by the number of answers with correct alphabetical ordering. Twenty-one (of 58) problems requested alphabetical ordering so that a maximum of 21 points was possible. The internal consistency in our sample was adequate for all measurement points, Cronbach's α = .78 at T1, .80 at T2, and .85 at T3.

The *number-one convention* is an agreement that it is not necessary to write the number "1" if there is a single letter (e.g., "$z + n + n = 2n + z$" rather than "$2n + 1z$," "$b \cdot a \cdot 4 \cdot a \cdot 4 \cdot a = 64a^3b$" rather than "$64a^3b^1$"). Student use of the convention not taught directly was determined by the number of omissions of a superfluous number "1". However, if a single letter such as x was not counted as "$1x$" in the problem "$x + 2x \neq 2x$," then it was assessed as an error in problem-solving knowledge. Thirteen (of 58) problems requested attention to a number "1", that is, a maximum of 13 points was possible. The internal consistency in our sample was adequate for all measurement points, Cronbach's α = .83 at T1, .83 at T2, and .78 at T3.

Verbal Concept Test

The *verbal concept test* was applied to assess verbal concept knowledge and verbal convention knowledge. This test assessed the students' ability to explain how to apply algebraic addition and multiplication. The students were asked to record two separate descriptions that explained how to solve each type of problem. For each explanation, the students were prompted with four hints, which were designed to activate their knowledge: "Describe in detailed steps how problems with letters are solved," "Mention what one has to pay attention to," "You can explain it by means of examples," and "Imagine you would like to explain the principles to classmates." The students' verbal concept explanations of algebraic addition and multiplication were scored based on a coding scheme (Table 2). Two trained raters independently scored and coded the students' answers for accuracy and completeness. The raters resolved disagreements in their scores by discussion. Inter-rater reliability was .89 at the first measurement point and .88 at the second measurement point (Cohen's kappa). Because of this high reliability, the test taken at the third point of measurement was scored by only one of the two raters. Based on the verbal concept test, similar to the problem-solving test, we assessed students' learning gains on directly taught verbal concepts and on not directly taught but incidentally learnable verbal conventions.

Verbal Concept Knowledge

This knowledge was judged by the amount of correctly reported algebraic concept features. These features were the principles that were considered central and essential for distinguishing algebraic addition from algebraic multiplication (Table 2). There were different scores for addition and multiplication that were summed up to an overall score, with a maximum of 11 points possible. As an additional measure, students' *misconceptions* were assessed. They were determined by the number of errors written down in the verbal concept explanations. Such

Table 2 Coding scheme for verbal concept knowledge

For every 1 of the following concept features mentioned in the verbal concept test, 1 point was scored for verbal concept knowledge (coding scheme adapted from former studies, e.g., author & coauthor, blinded 1). The listed features were the principles that were considered central and essential for distinguishing algebraic addition from algebraic multiplication, that is, in multiplications, the different kinds of splitting up of letters with exponents x^2, double letters ab, and letters with coefficients $2z$, and in addition, the distinction of different summands that cannot be merged as $x + x^2$, or $a + ab$, or $x + 4$.[a]

Points were equally given if (a) a key element was explicitly mentioned or (b) a key element became visible in the solution and the intermediate steps of the example. For "example statements," we only gave the point if students showed the rule in an explicit way, similarly to a verbal statement. If a point was described incorrectly, it was counted as a mistake.

Sub-analysis of addition: Concept features (total 6 points):

1 Sorting by summands (summands are the letter endings (sometimes with exponents: a, ab, a^2, …)
→ only ½ point: if only single letters were mentioned or used (a, b, c, …)
→ 2 points: if sufficiently detailed in a way that examples became redundant or dispensable

2 Summands are not split

3 Summands do not change in the result, exponents remain ($a^3 + a^3 = 2a^3$), no point if only with single letters

4 Letters with different exponents are not summarized, merged, or unified ($x + x^2$)

5 Single letters are not summarized, merged, or unified with double letters ($a + ab$)

6 Letters and numbers are not summarized, merged, or unified ($x + 4$)

Sub-analysis of multiplication: Concept features (total 5 points):

1 All factors are split
→ only ½ point: if only single letters are mentioned or used ($2x \cdot 3y \cdot 3x$)
→ 2 points: if sufficiently detailed, full description (separate units in the single components or mention important rules twice)

2 Letters with exponents are separated into single letters ($x^2 = x \cdot x$)

3 Double letters are separated ($ab = a \cdot b$)

4 Coefficients and letters are separated ($2z = 2 \cdot z$)

5 Exponents are added ($c^2 \cdot c^2 \cdot c^2 = c^6$)

[a] For an additional measure of secondary verbal concept knowledge, see Additional file 1

errors could be incorrect verbal statements or incorrect solution steps in examples used in the explanations. The errors in both the addition and multiplication explanations were added for a total score of misconceptions.

Verbal Convention Knowledge

This knowledge was determined by the number of statements mentioning the need to pay attention to the *alphabetical-order* or *number-one conventions* in separate scores for the addition and multiplication explanations. The students received points if they provided a textual description, e.g., "letters in the problem solution are sorted in alphabetical order," or if they demonstrated the correct application of a convention in an example, e.g., "$n + b + n + x + b + n = b + b + n + n + n + x = 2b + 3n + x$." Points from text or example descriptions in the addition and multiplication explanations were added to the total scores. For students' verbal alphabetical-order knowledge and verbal number-one knowledge, a maximum of 4 points each was possible. Two trained raters independently scored and coded the students' answers for convention knowledge. Inter-rater reliability was .99 at the first measurement point (Cohen's kappa). This almost perfect agreement can be ascribed to the high objectivity for the scoring of conventions. Thus, the data for the second and third points of measurement were scored by only one of the two raters.

Results

The results are presented in three sections. First, we analyzed whether the students' preconditions were similar in both conditions. Second, we report the effects of condition on the problem-solving and verbal explanation outcome measures. Neither gender differences nor gender-by-condition interactions were observed for any of the study measures. For each of the follow-up measures, we conducted separate, mixed-factorial ANOVAs with condition as a between-subject factor (explicit verbal explanation versus implicit problem solving) and time as a within-subject factor (1 day (T1), 1 week (T2), 10 weeks (T3)). Third, we report three-way ANOVA probing interaction effects between condition and type of measure. In these analyses, in order to examine whether the two conditions differentially impacted the two types of measures, we z-transformed the scores on the non-verbal and verbal knowledge measures. The three factors in these ANOVAs then were condition and measure (non-verbal vs. verbal measure) as between-subject factors and time as within-subject factor. The two-way interaction of condition × measure and the three-way interaction condition × measure × time were in focus in these ANOVAs. In the first three-way ANOVA, the interaction of implicit and explicit learning with problem-solving knowledge and verbal concept knowledge was assessed, the two main measures of algebra knowledge. For problem solving knowledge, we used the overall *z*-score on this measure, and for verbal concept knowledge, we aggregated the *z*-scores of concept knowledge and misconceptions. The focal interaction tests thus were condition × measure (problem solving knowledge vs. verbal concept knowledge) and condition × measure × time. In the second three-way ANOVA, we *z*-transformed the non-verbal alphabetical-order convention score of the problem solving test and the verbal alphabetical-order convention score of the verbal concept test. The focal tests in this ANOVA were the two-way interaction of condition × measure (non-verbal vs. verbal alphabetical-order knowledge), and the three-way interaction condition × measure × time. Finally, for a three-way ANOVA probing interactions on number-one knowledge, we *z*-transformed the non-verbal number-one convention score of the problem solving test and the verbal number-one convention score of the verbal concept test. The focal tests in this ANOVA were the two-way interaction condition × measure (non-verbal number-one knowledge vs. verbal number-one knowledge), and the three-way interaction condition × measure × time.

For analyses with violated sphericity assumption, we applied Greenhouse-Geisser correction with adjusted F ratios (according to Field, 2009) though doing so did not lead to different significance levels in any measure. When there was a main effect of the condition, post hoc t tests were performed for the three measurement points. After accounting for the main effect of time, Bonferroni-corrected comparisons with the first measurement point were conducted to determine how stable the effects were over time. Table 3 provides an overview of descriptive statistics and the effects of condition on the follow up measures at each time point.

Students' Preconditions

Prior Algebra Knowledge

No differences were observed between the conditions on students' prior algebra knowledge, t (1, 151) = 0.26, $p = .799$, $d = .05$. A floor effect was found, indicating that the students lacked

Table 3 Means and standard deviations of students' performance by condition on the problem solving test and the verbal concept test. Independent *t*-tests for the three measurement points

	T1			T2			T3		
	Explicit learning $N = 79$	Implicit learning $N = 74$		Explicit learning $N = 79$	Implicit learning N = 74		Explicit learning N = 79	Implicit learning N = 74	
	M (SD)	M (SD)	*d*	M (SD)	M (SD)	*d*	M (SD)	M (SD)	*d*
Problem solving test									
Problem solving knowledge (max 58 pts)	48.4 (8.6)	45.3 (10.3)	.33	48.1 (9.5)	45.1 (10.5)	.30	44.2 (11.4)	38.1 (11.7)	.53
Convention knowledge									
Alphabetical-order knowledge (max 21 pts)	9.2 (4.2)	9.9 (4.8)	.16	8.5 (4.2)	9.7 (5.2)	.26	10.0 (5.1)	12.1 (5.7)	.39
Number-one knowledge (max 13 pts)	8.4 (4.0)	9.2 (3.8)	.21	8.3 (4.1)	9.5 (3.6)	.29	9.1 (3.5)	9.4 (3.2)	.09
Verbal concept test									
Verbal concept knowledge (max 11 pts)	4.6 (2.3)	4.8 (2.9)	.08	4.2 (2.3)	4.8 (3.2)	.22	4.1 (2.4)	3.8 (3.0)	.11
Misconceptions[a]	0.2 (0.5)	0.6 (1.3)	.41	0.1 (0.4)	0.8 (1.3)	.74	0.3 (0.8)	1.1 (1.4)	.71
Verbal convention knowledge:									
Verbal alphabetical-order knowledge (max 4 pts)	0.06 (0.3)	0.01 (0.1)	.22	0.05 (0.2)	0.08 (0.3)	.12	0.04 (0.2)	0.00 (0.0)	.28
Verbal number-1 knowledge (max 4 pts)	0.83 (0.8)	0.81 (0.7)	.03	0.61 (0.9)	0.72 (0.7)	.14	0.52 (0.7)	0.57 (0.7)	.07

T1 1 day later for all the other measures, *T2* 1 week later, *T3* 10 weeks later

$*p < .05$, $** \ p < .01$, $*** \ p < .001$

[a] The misconceptions score is coded in reverse order, with a low value representing a good score

prior algebra knowledge and that they had no intuition for how to solve algebra problems (M = 0.90 of 8 problems, SD = 1.10).

Control Variables

A multivariate analysis of variance (MANOVA) did not indicate differences among the conditions with regard to the control variables, grades in *mathematics* (explicit: M = 4.82, SD = 0.67; implicit: M = 4.82, SD = 0.66), grades in *German* (explicit: M = 4.83, SD = 0.51; implicit: M = 4.81, SD = 0.59), *arithmetic knowledge* (explicit: M = 8.90, SD = 3.75; implicit: M = 9.99, SD = 4.07), and students' *reasoning ability* (explicit: M = 27.96, SD = 3.60; implicit: M = 27.50, SD = 3.66), F (4, 148) = 1.27, p = .283, η^2_p = .03. This result indicates a random assignment to the two conditions.

Effect of Condition on the Problem-Solving Test

For the problem-solving test, there were three measures: problem-solving knowledge and two not directly taught convention measures. In the explicit learning condition, students were expected to perform better on problem-solving knowledge with better development over the long term, whereas in the implicit condition, students were expected to perform better on the two convention measures compared with the explicit condition, especially in the short-term. Figure 3 depicts the course of the problem-solving measures.

Problem-Solving Knowledge

As expected, a main effect of condition was observed in favor of the explicit learning condition, F (1149) = 7.42, p = .004, η^2_p = .05 (Table 3). Post hoc t tests revealed differences at all measurement points with weak to moderate effects, at T1, p = .021, d = .33, T2, p = .032, d = .30, T3, p = .007, d = .53. There was also a main effect of time, F (1.48, 219.81) = 84.10, p < .001, η^2_p = .36. Bonferroni-corrected comparisons showed a change from T1 to T3, p < .001, but not from T1 to T2, p = 1.000. And there was a condition × time interaction, F (1.48, 219.81) = 4.41, p = .023, η^2_p = .03, with simple contrasts showing a difference from T1 to T3, p < .001 but not from T1 to T2, p = .803. The graph revealed a more pronounced performance decrease in the implicit learning condition (Fig. 3). Thus, these results demonstrate a long-term advantage of using verbal explanations when learning algebra concepts. As expected, for careless errors, there was neither a condition effect, F (1, 149) = 0.36, p = .552, η^2_p = .00, nor an interaction, F (2, 298) = 0.44, p = .957, η^2_p = .00, indicating that careless errors were independent of condition or intervention. Notably, there was a time effect for careless errors, F (2, 298) = 4.09, p = .018, η^2 = .03, with a decrease in errors at T3, which suggested a more relaxed processing of the test 10 weeks later in both conditions.

Alphabetical-Order Knowledge

As expected, students in the implicit learning condition outperformed those in the explicit learning condition regarding the alphabetical-order conventions not directly taught, F (1, 149) = 3.92, p = .025, η^2_p = .03 (Table 3). As expected, there was also a time effect, F (1.72, 255.73) = 23.54, p < .001, η^2_p = .14, but no interaction, F (1.72, 255.73) = 2.13,

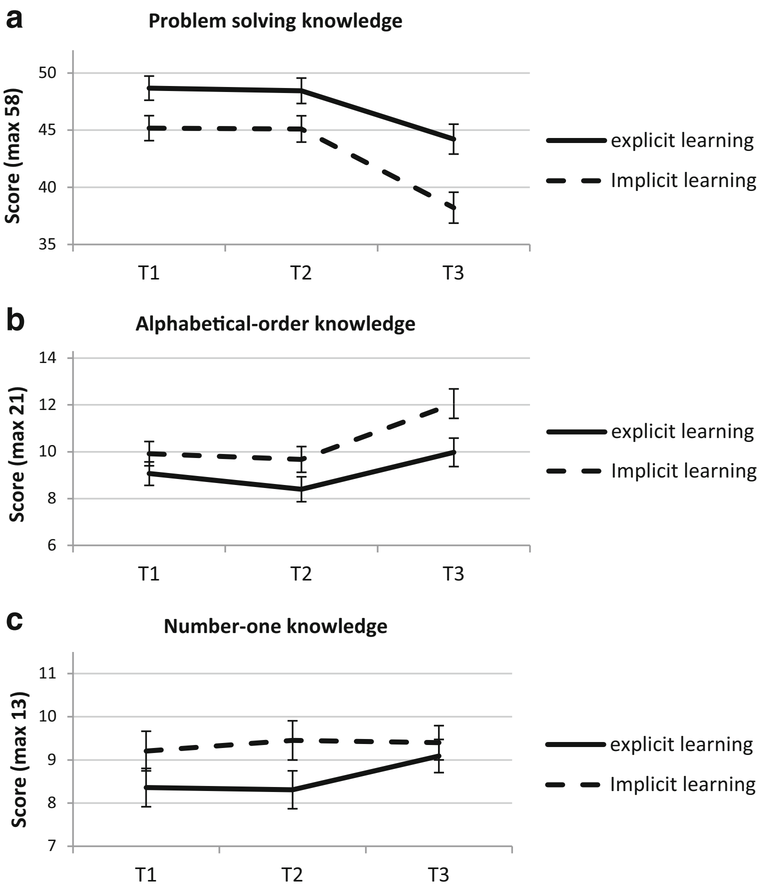

Fig. 3 **a**–**c** Performance on the three problem-solving test measures. *T1* 1 day later, *T2* 1 week later, *T3* 10 weeks later. *Error bars* represent standard errors

$p = .129$, $\eta^2_p = .01$. Bonferroni-corrected comparisons showed an increase in the score for both conditions from T1 to T3, $p < .001$, but not from T1 to T2, $p = .142$.

Number-One Knowledge

The implicit learning condition showed higher means across all measurement points on measures of the number-one conventions not directly taught; however, no evidence of a difference was found, F (1, 149) = 1.88, $p = .086$, $\eta^2_p = .01$ (Table 3). There was no effect of time, F (1.85, 275.04) = 2.68, $p = .075$, $\eta^2_p = .02$, and no interaction, F (1.85, 275.04) = 2.04, $p = .136$, $\eta^2_p = .01$.

Effect of Condition on the Verbal Concept Test

For the verbal concept test, there were four measures: two verbal concept measures, and two not directly taught convention measures. In the explicit learning condition, students were expected to perform (1) better on all verbal explanation measures and (2) with better development over the long term. Figure 4 depicts the course of the verbal explanation measures.

Verbal Concept Knowledge

Unexpectedly, no evidence of a condition difference was found on the main score of verbal concept knowledge, F (1, 149) = 0.24, p = .319, η^2_p = .00 (Table 3). However,

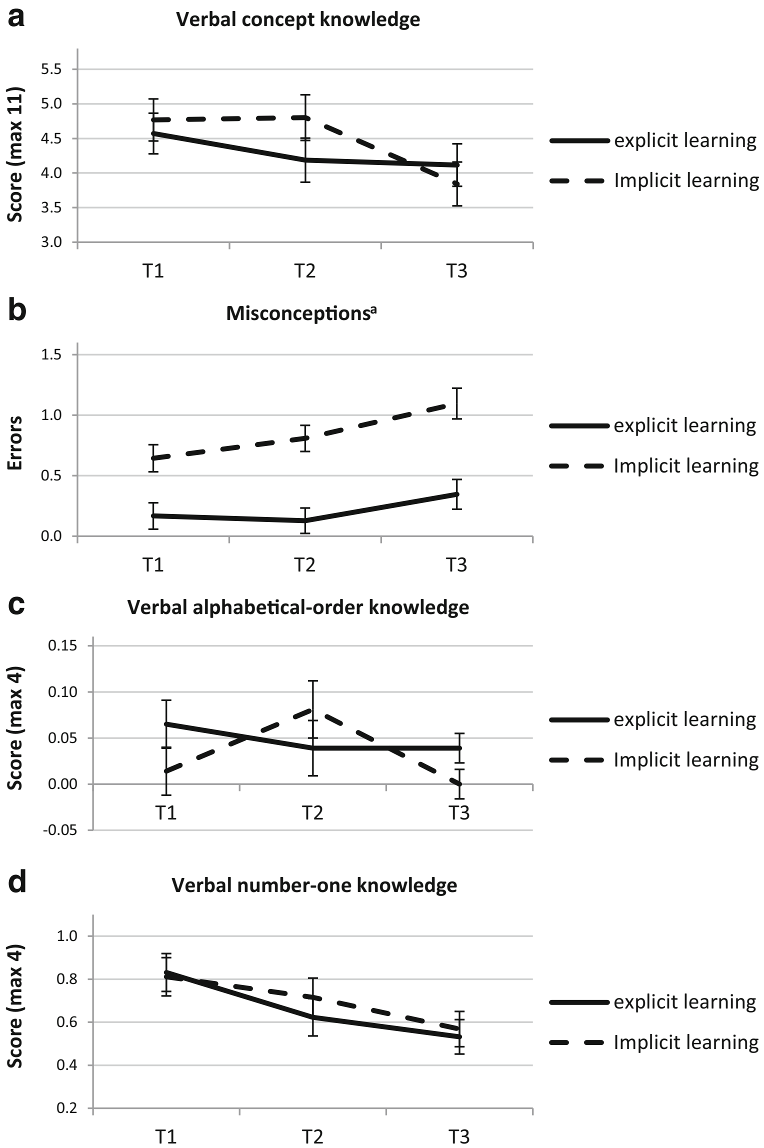

Fig. 4 **a**–**d** Performance on the four verbal concept test measures. *T1* 1 day later, *T2* 1 week later, *T3* 10 weeks later. *Error bars* represent standard errors. *a* For misconceptions, the number of errors was reported, i.e., the score is coded in reverse order, with a low value representing a good score

there was a time effect, F (2, 298) = 6.83, p = .001, η^2_p = .04, with Bonferroni-corrected comparisons showing a decrease from T1 to T3, p = .003, but not from T1 to T2, p = .989. There was no condition × time interaction, F (2, 298) = 2.62, p = .074, η^2_p = .02. The graphs revealed a cross-over at T3 caused by stable scores in the explicit learning condition and a drop of the implicit learning condition; however, this result did not achieve significance (Fig. 4). Thus, we did not find a difference between the two conditions regarding verbal concept knowledge.

We conducted additional analyses to examine where the unexpected lack of difference between the implicit learning and the explicit learning conditions on concept features stems from. In the verbal concept test, the students could describe rules by applying them in examples. The implicit learners engaged in this strategy much more often (M = 3.30, SD = 2.15) than the explicit learners (M = 1.13, SD = 1.16), t (1, 151) = −7.84, p < .001. Crucially, the amount of applied examples was highly correlated to students' verbal concept knowledge score in the explicit learning condition, at T1, r (79) = .50, p < .001, at T2, r (79) = .46, p < .001, and at T3, r (78) = .56, p < .001, and in the implicit learning condition, at T1, r (74) = .50, p < .001, at T2, r (73) = .45, p < .001, and at T3, r (74) = .65, p < .001. Thus, the students in the implicit learning condition might have performed well on this measure due to their frequent practice of creating examples during the intervention.

Misconceptions

As expected, there was a main effect of condition on misconceptions of the verbal concept test in favor of the explicit learning condition, F (1, 149) = 25.47, p < .001, η^2_p = .15 (Table 3). There was an effect of time, F (1, 149) = 7.21, p = .001, η^2_p = .05, with Bonferroni-corrected comparisons and the graph showing a decrease from T1 to T3, p = .003, but not yet from T1 to T2, p = 1.000, but no interaction, F (1, 149) = 1.29, p = .276, η^2_p = .01 (Fig. 4). This result shows that explicit explanation learning led to fewer misconceptions and this change remained stable up to 1 week.

Verbal Alphabetical-Order Knowledge

In both conditions, only a few students verbally reported alphabetical-order knowledge, with average scores only between 0.00 to 0.06 out of 4 points. There were no effects of condition, F (1, 149) = 0.66, p = .209, η^2_p = .00, time, F (1.85, 276.23) = 1.29, p = .276, η^2_p = .01, or their interaction, F (1.85, 276.23) = 2.02, p = .138, η^2_p = .01. Hence, in the verbal assessment, implicit (verbal) learning had no effect on the alphabetical-order convention not directly taught.

Verbal Number-One Knowledge

There was no effect of condition, F (1, 149) = 0.69, p = .344, η^2_p = .00, but there was a time effect, F (2, 298) = 7.18, p = .001, η^2_p = .05, with Bonferroni-corrected comparisons showing a decrease from T1 to T3, p = .001, but not from T1 to T2, p = .118, but no interaction, F (2, 298) = 0.312, p = .732, η^2_p = .00. This result showed no advantage for the explicit (verbal) learning of the number-one convention not taught directly when the knowledge is verbally assessed.

Interaction between Conditions and Knowledge Measures

Interactions between condition and knowledge measures were examined on the main measures of algebra knowledge, on alphabetical-ordering knowledge, and on number-one knowledge.

On alphabetical-order knowledge, there was an interaction between condition and measure, $F\,(1, 296) = 4.41$, $p = .036$, $\eta^2_p = .02$, but no three-way interaction condition × measure × time, $F\,(2, 592) = 2.16$, $p = .116$, $\eta^2_p = .01$. There was also no main effect of condition, $F\,(1, 296) = 1.52$, $p = .219$, $\eta^2_p = .01$. The interaction between condition and measure qualifies the advantage of implicit learning over explicit learning on the non-verbal alphabetical-order measure portrayed in Fig. 3b to be significantly stronger than the non-significant condition effect on the verbal alphabetical-order measure visible in Fig. 4c. Thus, the effect of the two conditions differed between the two alphabetical-order measures; it was stronger on the non-verbal measure than on the verbal measure. On number-one knowledge, there was no interaction between condition and measure, $F\,(1, 296) = 0.62$, $p = .433$, $\eta^2_p = .00$, no three-way interactions condition × measure × time, $F\,(2, 592) = 0.34$, $p = .706$, $\eta^2_p = .00$, and also no main effect of condition, $F\,(1, 296) = 1.83$, $p = .177$, $\eta^2_p = .01$. Thus, no difference in the effects of the two conditions was found between the two types of number-one knowledge measures.

On the main measure algebra knowledge, there was no interaction between condition and measure, $F\,(1, 298) = 1.54$, $p = .215$, $\eta^2_p = .01$, and no three-way interaction condition × measure × time, $F\,(2, 596) = 0.07$, $p = .924$, $\eta^2_p = .00$. However, there was a main effect of condition, $F\,(1, 296) = 7.05$, $p = .008$, $\eta^2_p = .02$, highlighting the advantage of explicit learning across both the verbal and non-verbal algebra knowledge measures compared with implicit learning.

Discussion

We investigated whether the introduction of contrasted algebra material across 4 days is more effective when students engage in explicit verbal explanation of concepts or when they engage in implicit problem solving. We assessed the degree to which the two types of learning influenced different measures of these kinds of knowledge. We were particularly interested in disentangling the effect of prompting explicit comparison from the effect of prompting extensive problem solving of contrasted material for both the intentional and incidental processing of material.

Overall, as expected, explicit comparison outperformed implicit problem solving on the main measures assessing algebraic term transformation. The explicit learning condition showed higher problem-solving knowledge and fewer misconceptions than the implicit learning condition, but unexpectedly, no difference in verbal concept knowledge was observed. For the implicit learning condition, an advantage was found for measures that assessed students' use of not-directly-taught conventions on the problem-solving test. These results confirm that verbal explanations are a powerful method of learning not only the same type of verbal explanation knowledge but also, in particular, the other type of non-verbal problem-solving knowledge. Moreover, an influence of implicit learning on verbal explanation measures was not found.

An interaction analysis showed that the advantage of explicit learning was stable across the main algebra knowledge measures, with a main effect of condition but no interaction effect with the type of measure. Thus, on our main algebra knowledge measures, our results confirm an asymmetry of knowledge acquisition (Aleven & Koedinger, 2002). Both on verbal and non-verbal measures, explicit learning showed advantages in comparison with implicit learning.

A further interaction analysis showed a different pattern of effects on incidentally learned aspects. On one incidentally learned aspect, that is, alphabetical order knowledge, implicit learning showed an advantage on the non-verbal measure that was not present on the verbal measure.

The Strengths of Explicit Comparison

As expected, explicitly learning juxtaposed addition and multiplication algebra problems by verbal comparison better supported students' problem-solving performance than by extensive problem solving. As an explanation, we assume that the challenging distinction of the two similar concepts (algebraic addition and multiplication) requires an explication of the differences (Gentner et al., 2003; Koedinger et al., 2012). Algebraic addition and multiplication are not complex mathematical concepts per se compared with most other mathematical concepts because both consist of many principles and rules that, in themselves, are easy to understand. However, what makes basic algebra challenging is that the many principles that belong to either addition or multiplication must be properly distinguished. Materials with low processing demands benefit from instructional conditions that increase the information and processing load (Wulf & Shea, 2002). We did not measure student effort experienced during instruction. However, the advantage of the implicit learning condition in the learning of not directly taught algebraic conventions might indicate that the students had free capacity available (see also Ziegler & Stern, 2014). Explicit comparison might have channeled this free capacity, creating a more adequately demanding environment than problem solving. That is, the comparisons may have triggered understanding and sense-making processes that improved capturing the differences and improved the distinction between addition and multiplication (Koedinger et al., 2012). We also found that problem-solving knowledge in the explicit learning condition remained consistent over time, compared with a pronounced performance decrease in the implicit learning condition. This result emphasizes the longer term effect of explicit comparison. In interleaved and comparison learning research, immediate learning or delays of a few hours or days, and sometimes delays of a week or a month, are often examined (Rau et al., 2013; Rittle-Johnson et al., 2012; Rittle-Johnson & Star, 2009; Rohrer & Taylor, 2007). In accordance with our assumption, extending the measurement period to 10 weeks revealed increasing outcome differences between the two conditions over time. This result confirms that implicit learning leads to more automated but short-term knowledge (Koedinger et al., 2012). Thus, we suppose that the similarity between the two concepts might have led to interference caused by induction learning with little deliberate awareness, which made retrieval after 10 weeks more difficult (Anderson, 1996).

On the verbal concept knowledge measure, the explicit learning condition showed fewer misconceptions than the implicit learning condition. We consider misconceptions as an important indicator of students' algebra knowledge. Misconceptions can originate from insufficient knowledge, and the more knowledge learners have, the fewer misconceptions they have (Körner, 2005). In our intervention, the explication of principles and relevant features that were requested in the explicit learning condition appear to have led to a better representation of the correct algebraic procedures, preventing students from reporting and applying misconceptions. In particular, we believe that explicated features and rules help distinguish addition from multiplication, for example, the continuity of summands such has "x" or "xy" in addition ($xy + x + xy + x = x + x + xy + xy = 2\times + 2xy$) that must be clearly distinguished from the splitting of factors in multiplication ($xy \cdot x \cdot xy \cdot x = x \cdot y \cdot x \cdot x \cdot y \cdot x = x^4y^2$). Verbalizing this difference explicitly contrasts the splitting in multiplication from the summands in addition (i.e., a repeated verbal assignment regarding the central features distinguishing addition from multiplication). This presumably better representation gained through the explicit learning of contrasted material might be explained by the potential of explanations to detect and fill gaps as well as the differences in material to be learned (VanLehn et al., 1992) or the need to connect and integrate visual perceptual and verbal components (Aleven & Koedinger, 2002).

The Strengths of Extensive Problem Solving

As expected, the implicit learning condition scored higher on the problem-solving test than the explicit learning condition in terms of the application of the alphabetical-ordering convention that were not directly taught. We assume that the implicit learners had greater capacities to notice features not directly taught because the problem generation and solving was less challenging than the explicit comparisons under the explicit learning condition. Thus, making instructions more demanding by directing learners' attention to central features of concepts might support the learning of verbal concept knowledge (e.g., Sweller, 1994); simultaneously, however, at the same time it might reduce attention to incidental aspects (Ziegler & Stern, 2014). Although alphabetical ordering of variables is a minor component of learning algebraic term transformation, it is not unimportant. Thus, to support the development of verbal concept knowledge and the learning of conventions alike, future settings could investigate the sequencing or combining of explicit verbalization and implicit problem solving.

We did not find differences between the two learning conditions with respect to the concept knowledge assessed with a verbal measure. It was unexpected that the students who learned by explaining differences in the presented worked examples did not acquire better skills in expressing their knowledge, especially given that they have acquired better problem solving knowledge. In additional analyses, we found a potential explanation for these results. The students in the implicit condition generated and solved problems throughout the instruction, making it more likely for these students to include many examples in their written concept explanations. On average, students from the implicit learning condition implemented approximately three times as many examples in their explanations as students from the explicit learning condition. The number of applied examples was also highly correlated with the verbal concept knowledge scores. Thus, the verbal concept test used in our study apparently reflects differences in students' answer behavior that do not necessarily reflect differences in algebra knowledge. More likely, the groups may have differently responded to the verbal concept measure, a phenomenon that is well known from research investigating the disadvantages of specific groups related to algebra measures (Holland & Wainer, 2012; O'Neill et al., 1993). Support for this assumption comes from the clear outperformance of explicit learners on the verbal misconceptions score, a good indicator of learning benefits. We assume that explaining a concept by using an example might be easier for students than expressing concept knowledge with words (e.g., von Aufschnaiter & Rogge, 2010). Thus, it is unclear to which degree the lack of difference between the groups can be attributed to differences in verbal concept knowledge, and to differences in answer behavior. In future studies, such interactions between intervention and the functioning of measures should be considered when planning studies. Specific intervention conditions should not have advantages caused by characteristics of measures or scoring procedures, which cannot be ruled out for our verbal concept measure.

Interactions between Conditions and Knowledge Measures

Analyses of the interaction between the two conditions and the two types of knowledge measures indicated that the conditions differed in the amount to which they supported students' alphabetical-order knowledge, while no significant interactions were found for number-one knowledge and the main measures of algebra knowledge. The significant interaction indicates that the advantage of implicit learning on non-verbal alphabetical order knowledge was stronger than the difference between the two conditions on verbal alphabetical-order knowledge. The lack of interaction on the main measures of algebra knowledge together with the clear main effect of condition supports the asymmetry hypothesis, as explicit learning outperformed implicit learning on both problem solving and verbal concept knowledge measures. As a limitation of this analysis, it should however be noted that the measures to assess the two types of knowledge differed in the type of items. Such differing characteristics between measures cannot be eradicated by standardizing scores obtained with different measures, because effects of assessment methods can substantially influence unstandardized and standardized effect estimates (Fiedler, 2011; Podsakoff et al., 2003). The present study therefore provided a first but limited opportunity to compare the effects of both types of learning across a broad set of knowledge measures. For similar comparisons in future studies, we suggest to use more than one type of measure for each of the two types of knowledge. Applying latent variable models on data obtained with different types of measures, for example, would then allow more reliable comparisons of effects across measures, by disentangling effects of assessment methods from real effects of intervention (Eid et al., 2003).

Verbal and Non-verbal Knowledge Measures

The problem-solving and verbal concept knowledge measures used in our study are related to the frequently used pairs of implicit and explicit, procedural and conceptual (declarative), or non-verbal and verbal knowledge (Aleven & Koedinger, 2002; Anderson & Lebiere, 1998; Koedinger et al., 2012; Rittle-Johnson et al., 2015; Sun et al., 2005). These pairs signify different characteristics of knowledge acquisition, such as the level of consciousness during knowledge acquisition, the representation of knowledge, or the degree of explication while working on problems. The different characteristics are however seldom thoroughly considered and distinguished as knowledge dimensions. For example, for Aleven and Koedinger (2002), procedural and declarative knowledge correspond to implicit and explicit knowledge. Because it is difficult to decide which type of knowledge is involved in a measurement and to validly differentiate the types of knowledge from each other, we decided to circumvent this problem and designated the type of knowledge according to how it was assessed.

It is difficult to distinguish implicit from procedural knowledge, and explicit from conceptual knowledge empirically. Theoretically, it is well possible to describe and define procedural knowledge and implicit knowledge separately (the same for conceptual knowledge and explicit knowledge). It is however difficult to separately assess the knowledge in mathematics learning, a field where all the materials and concepts are consciously accessible. In other fields, such as in complex systems learning, which involves material with complex hierarchical relations that are not consciously accessible, implicit knowledge is measurable by how fast or accurate a task is solved or an aim is reached. However, in mathematics learning, material and concepts are generally accessible and therefore it is difficult to know for a solved task which knowledge was applied. Thus, this taxonomy is not uncontested, because it is difficult to determine what implicit or procedural and explicit or conceptual knowledge include. For example, in procedural tasks such as problem solving, it is difficult to distinguish whether applied knowledge is implicit or still explicit, and to which proportion procedural and conceptual knowledge are involved. For verbal concept explanations, it can be expected that conceptual knowledge has to applied; however, people may also have procedures in mind that they use as a reference for their explanations, or when solving multiple choice items (Rittle-Johnson et al., 2015).

In mathematics learning, both implicit and procedural knowledge are often treated in the same way and are typically assessed with a non-verbal measure such as problem-solving tasks (i.e., the application of knowledge). Explicit and conceptual knowledge are typically assessed with a verbal measure such as verbal explanations of principles (i.e., the explication of knowledge). This promotes using the measures synonymously and further renders it difficult to distinguish them properly. In addition, procedural and conceptual knowledge, and implicit and explicit knowledge are highly correlated and also highly intertwined within a person's total knowledge, making it difficult to distinguish between the two or measure each independently from the other (Schneider & Stern, 2010). Thus, it is also difficult to measure the types of knowledge directly and unambiguously.

In our problem-solving test, the students had to solve mixed addition and multiplication problems, a task that necessarily involves some use of conceptual differentiation. Thus, conceptual knowledge can include both implicit and explicit understandings of principles; for example, explicit knowledge can include both conceptual and procedural skills. In other words, although someone might understand and solve a problem adequately, that person might not be able to verbalize his or her knowledge (Bou-Llusar & Segarra-Ciprés, 2006; von Aufschnaiter & Rogge, 2010).

Thus, we think that the terms "procedural and conceptual" and "implicit and explicit" knowledge should be used with great caution in educational settings. Therefore, in the present study we decided to refer to the type of knowledge according to how it is assessed (i.e., problem-solving knowledge was measured via the percentage of correctly solved routine problems and verbal concept knowledge by asking for verbal explanations of the concepts).

Nevertheless, the distinction of the two kinds of knowledge remains meaningful precisely because one type of knowledge can exist and be observed to some degree without the other. Successful problem solving probably needs a larger proportion of procedural knowledge, and verbal explanation likely requires a larger proportion of conceptual understanding.

Our results confirm that the challenging distinction between two algebra concepts in problem solving cannot be acquired through pure problem-solving practice; rather, it benefits from verbal explanations that promote the understanding and sense-making process without neglecting problem-solving skills. This focus on verbal explanation and sense-making appears to be especially important for rule-based concepts in challenging fields in which the comprehension of material is crucial (Koedinger et al., 2012).

Outlook and Implications

In the present study, we examined the relative merits of explicit and implicit learning on a broad set of knowledge measures. Advantages for explicit learning were found on the verbal and non-verbal measures of the taught concepts, higher problem solving knowledge and less misconceptions. Advantages for implicit learning were found for the incidental learnable aspects. Thus, the merit of implicit learning is to compensate the disadvantage on the taught concepts with an improved learning of incidental aspects of the material.

Overall, our results suggest that the contrasted presentation of learning material clearly gains from an explicit learning approach. Verbalization helped students learn juxtaposed addition and multiplication transformations, with explicit comparison adding value to the simultaneous processing of contrasted material. We demonstrated an advantage of explicit verbal learning for similar and juxtaposed algebra principles. We assume that the explication of contrasted material is generalizable to other types of algebraic and mathematical material that comprise different principle, such as different solution methods for solving equations (e.g., Rittle-Johnson & Star, 2007) or different operations with fractions. Verbalizations extend the processing of contrasted learning material by enabling students to focus on the central features of the concepts, although this advantage might not yet be visible during the learning phase. Thus, we support verbal explanation as a teaching method that demands extra effort from learners but supports their lasting understanding of newly introduced concepts. At the same time, we showed the impact of implicit learning on incidental aspects of to-be-learned materials, for example on the alphabetical ordering convention, even though no effect was found on number-one convention. Therefore, future research has to show whether these or other incidental learning benefits can be confirmed when explicitly studying materials.

Acknowledgements We are grateful to Sara Ziegler for her assistance in the implementation of the study in the school classes and Sara Ziegler and Adrienne Suter for their assistance in the evaluation of the test materials.

Ethics Approval and Consent to Participate All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Competing Interests The authors declare that they have no competing interests.

Informed Consent Informed consent was obtained from all individual participants for whom identifying information is included in this article.

References

Aleven, V. A., & Koedinger, K. R. (2002). An effective metacognitive strategy: learning by doing and explaining with a computer-based cognitive tutor. *Cogni Sci, 26*(2), 147–179. https://doi.org/10.1207/s15516709cog2602_1.

Alfieri, L., Nokes, T. J., & Schunn, C. D. (2013). Learning through case comparisons: a metaanalytic review. *Educational Psychologist, 48*(2), 87–113. https://doi.org/10.1080/00461520.2013.775712.

Anderson, J. R. (1995). *Learning and memory: an integrated approach*. New York: Wiley.

Anderson, J. R. (1996). *The architecture of cognition*. Cambridge: Harvard University Press.

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Mahwah: Lawrence Erlbaum Associates.

Atkinson, R. K., Renkl, A., & Merrill, M. M. (2003). Transitioning from studying examples to solving problems: effects of self-explanation prompts and fading worked-out steps. *Journal of Educational Psychology, 95*(4), 774–783. https://doi.org/10.1037/0022-0663.95.4.774.

Baroody, A. J., Feil, Y., & Johnson, A. R. (2007). An alternative reconceptualization of procedural and conceptual knowledge. *Journal for Research in Mathematics Education, 38*(2), 115–131.

Birnbaum, M. S., Kornell, N., Bjork, E. L., & Bjork, R. A. (2013). Why interleaving enhances inductive learning: the roles of discrimination and retrieval. *Memory and Cognition, 41*(3), 392–402. https://doi.org/10.3758/s13421-012-0272-7.

Bou-Llusar, J. C., & Segarra-Ciprés, M. (2006). Strategic knowledge transfer and its implications for competitive advantage: an integrative conceptual framework. *Journal of Knowledge Management, 10*(4), 100–112. https://doi.org/10.1108/13673270610679390.

Carvalho, P. F., & Goldstone, R. L. (2014). Effects of interleaved and blocked study on delayed test of category learning generalization. *Frontiers in Psychology, 5*(936), 1–11. https://doi.org/10.3389/fpsyg.2014.00936.

Chi, M. T. H. (2000). Self-explaining expository texts: the dual processes of generating inferences and repairing mental models. In R. Glaser (Ed.), *Advances in instructional psychology: educational design and cognitive science* (Vol. 5, pp. 161–238). Mahwah: Erlbaum.

DeKeyser, R. (2003). Implicit and explicit learning. In C. J. Doughty & M. H. Long (Eds.), *Handbook of second language acquisition* (pp. 313–348). Oxford: Blackwell Publishing.

Eid, M., Lischetzke, T., Nussbeck, F. W., & Trierweiler, L. I. (2003). Separating trait effects from trait-specific method effects in multitrait-multimethod models: a multiple-indicator CT-C(M-1) model. *Psychology Methods, 8*(1), 38–60. https://doi.org/10.1037/1082-989X.8.1.38.

Fiedler, K. (2011). Voodoo correlations are everywhere—not only neuroscience. *Perspectives on Psychological Science, 6*(2), 163–171. https://doi.org/10.1177/1745691611400237.

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). London: Sage.

Gentner, D. (2010). Bootstrapping the mind: analogical processes and symbol systems. *Cognitive Science, 34*(5), 752–775. https://doi.org/10.1111/j.1551-6709.2010.01114.x.

Gentner, D., Loewenstein, J., & Thompson, L. (2003). Learning and transfer: a general role for analogical encoding. *Journal of Educational Psychology, 95*(2), 393–408. https://doi.org/10.1037/0022-0663.95.2.393.

Gentner, D., Loewenstein, J., Thompson, L., & Forbus, K. D. (2009). Reviving inert knowledge: analogical abstraction supports relational retrieval of past events. *Cognitive Science, 33*(8), 1343–1382. https://doi.org/10.1111/j.1551-6709.2009.01070.x.

Holland, P. W., & Wainer, H. (2012). *Differential item functioning*. New York: Routledge.

Horn, W. (1983). *Leistungsprüfsystem LPS* (2. Auflage ed.). Göttingen: Hogrefe.

Kamii, C., & Dominick, A. (1997). To teach or not to teach algorithms. *Journal of Mathematical Behavior, 16*(1), 51–61. https://doi.org/10.1016/S0732-3123(97)90007-9.

Kang, S. H. K., & Pashler, H. (2012). Learning painting styles: spacing is advantageous when it promotes discriminative contrast. *Applied Cognitive Psychology, 26*(1), 97–103. https://doi.org/10.1002/acp.1801.

Karmiloff-Smith, A. (1994). Précis of beyond modularity: a developmental perspective on cognitive science. *Behavioral and Brain Sciences, 17*, 693–745. https://doi.org/10.1017/S0140525X00036621.

Koedinger, K. R., & Aleven, V. A. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review, 19*(3), 239–264. https://doi.org/10.1007/s10648-007-9049-0.

Koedinger, K. R., Corbett, A. T., & Perfetti, C. (2012). The knowledge-learning-instruction framework: bridging the science-practice chasm to enhance robust student learning. *Cognitive Science, 36*(5), 757–798. https://doi.org/10.1111/j.1551-6709.2012.01245.x.

Körner, C. (2005). Concepts and misconceptions in comprehension of hierarchical graphs. *Learning and Instruction, 15*, 281–296. https://doi.org/10.1016/j.learninstruc.2005.07.003.

Kurtz, K. J., Miao, C. H., & Gentner, D. (2001). Learning by analogical bootstrapping. *Journal of the Learning Sciences, 10*(4), 417–446. https://doi.org/10.1207/S15327809JLS1004new_2.

Matthews, P. G., & Rittle-Johnson, B. (2009). In pursuit of knowledge: comparing self-explanations, concepts, and procedures as pedagogical tools. *Journal of Experimental Child Psychology, 104*(1), 1–21. https://doi.org/10.1016/j.jecp. 2008.08.004.

McCloskey, M., & Cohen, N. J. (1989). Catastrophic interference in connectionist networks: the sequential learning problem. *The Psychology of Learning and Motivation, 24*, 109–165. https://doi.org/10.1016/S0079-7421(08)60536-8.

Mitchell, C., Nash, S., & Hall, G. (2008). The intermixed-blocked effect in human perceptual learning is not the consequence of trial spacing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(1), 237–242. https://doi.org/10.1037/0278-7393.34.1.237.

O'Neill, K. A., McPeek, W. M., & Wild, C. L. (1993). Differential item functioning on the graduate management admission test. *ETS Research Report Series, 1993*(2), 1–46. https://doi.org/10.1002/j.2333-8504.1993.tb01546.x.

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879–903. https://doi.org/10.1037/0021-9010.88.5.879.

Rau, A. R., Aleven, V. A., & Rummel, N. (2013). Interleaved practice in multi-dimensional learning tasks: which dimension should we interleave? *Learning and Instruction, 23*(1). https://doi.org/10.1016/j.learninstruc.2012.07.003.

Renkl, A. (1997). Learning from worked-out examples: a study on individual differences. *Cognitive Science, 21*(1), 1–29. https://doi.org/10.1016/S0364-0213(99)80017-2.

Renkl, A. (2005). The worked-out example principle in multimedia learning. In R. E. Mayer (Ed.), *Cambridge handbook of multimedia learning* (pp. 229–247). Cambridge, England: Cambridge University Press.

Renkl, A. (2015). Different roads lead to Rome: the case of principle-based cognitive skills. *Learning: Research and Practice, 1*(1), 79–90. https://doi.org/10.1080/23735082.2015.994255.

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

Rittle-Johnson, B., & Star, J. R. (2007). Does comparing solution methods facilitate conceptual and procedural, knowledge? An experimental study on learning to solve equations. *Journal of Educational Psychology, 99*(3), 561–574. https://doi.org/10.1037/0022-0663.99.3.561.

Rittle-Johnson, B., & Star, J. R. (2009). Compared with what? The effects of different comparisons on conceptual knowledge and procedural flexibility for equation solving. *Journal of Educational Psychology, 101*(3), 529–544. https://doi.org/10.1037/a0014224.

Rittle-Johnson, B., & Star, J. R. (2011). The power of comparison in learning and instruction: learning outcomes supported by different types of comparisons. *Psychology of Learning and Motivation: Cognition in Education, 55*, 199–225. https://doi.org/10.1016/B978-0-12-387691-1.00007-7.

Rittle-Johnson, B., Star, J., & Durkin, K. (2012). Developing procedural flexibility: are novices prepared to learn from comparing procedures? *British Journal of Educational Psychology, 82*(3), 436–455. https://doi.org/10.1111/j.2044-8279.2011.02037.x.

Rittle-Johnson, B., Schneider, M., & Star, J. R. (2015). Not a one-way street: bidirectional relations between procedural and conceptual knowledge of mathematics. *Educational Psychology Review, 27*(4), 587–597. https://doi.org/10.1007/s10648-015-9302-x.

Rohrer, D. (2012). Interleaving helps students distinguish among similar concepts. *Educational Psychology Review, 24*(3), 355–367. https://doi.org/10.1007/s10648-012-9201-3.

Rohrer, D., & Taylor, K. (2007). The shuffling of mathematics practice problems boosts learning. *Instructional Science, 35*(6), 481–498. https://doi.org/10.1007/s11251-007-9015-8.

Schneider, M., & Stern, E. (2010). The developmental relations between conceptual and procedural knowledge: a multimethod approach. *Developmental Psychology, 46*(1), 178–192. https://doi.org/10.1037/a0016701.

Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. *Cognition and Instruction, 16*(4), 475–522. https://doi.org/10.1207/s1532690xci1604_4.

Siegler, R. S., & Stern, E. (1998). Conscious and unconscious strategy discoveries: a microgenetic analysis. *Journal of Experimental Psychology-General, 127*(4), 377–397. https://doi.org/10.1037/0096-3445.127.4.377.

Star, J. R., & Rittle-Johnson, B. (2009). It pays to compare: an experimental study on computational estimation. *Journal of Experimental Child Psychology, 102*(4), 408–426. https://doi.org/10.1016/j.jecp.2008.11.004.

Sun, R., Slusarz, P., & Terry, C. (2005). The interaction of the explicit and the implicit in skill learning: a dual-process approach. *Psychological Review, 112*(1), 159–192. https://doi.org/10.1037/0033-295X.112.1.159.

Sun, R., Mathews, R. C., & Lane, S. M. (2007). Implicit and explicit processes in the development of cognitive skills: a theoretical interpretation with some practical implications for science instruction. In E. M. Vargios (Ed.), *Educational psychology research* (pp. 1–26). New York: Nova Science Publishers.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4*(4), 295–312. https://doi.org/10.1016/0959-4752(94)90003-5.

Sweller, J. (2006). The worked example effect and human cognition. *Learning and Instruction, 16*(2), 165–169. https://doi.org/10.1016/j.learninstruc.2006.02.005.

Taylor, K., & Rohrer, D. (2010). The effect of interleaving practice. *Applied Cognitive Psychology, 24*(6), 837–848. https://doi.org/10.1002/acp.1598.

VanLehn, K., Jones, R. M., & Chi, M. T. H. (1992). A model of the self-explanation effect. *Journal of the Learning Sciences, 2*, 1–59. https://doi.org/10.1207/s15327809jls0201_1.

von Aufschnaiter, C., & Rogge, C. (2010). Misconceptions or missing conceptions. *Eurasia Journal of Mathematics, Science and Technology Education, 6*(1), 3–18.

Williams, J. J., & Lombrozo, T. (2010). The role of explanation in discovery and generalization: evidence from category learning. *Cognitive Science, 34*(5), 776–806. https://doi.org/10.1111/j.1551-6709.2010.01113.x.

Wong, R. M. F., Lawson, M. J., & Keeves, J. (2002). The effects of self-explanation training on students' problem solving in high-school mathematics. *Learning and Instruction, 12*(2), 233–262. https://doi.org/10.1016/S0959-4752(01)00027-5.

Wulf, G., & Shea, C. H. (2002). Principles derived from the study of simple skills do not generalize to complex skill learning. *Psychonomic Bulletin & Review, 9*(2), 185–211. https://doi.org/10.3758/BF03196276.

Ziegler, E., & Stern, E. (2014). Delayed benefits of learning elementary algebraic transformations through contrasted comparisons. *Learning and Instruction, 33*, 1–16. https://doi.org/10.1016/j.learninstruc.2014.04.006.

Ziegler, E., & Stern, E. (2016). Consistent advantages of contrasted comparisons: algebra learning under direct instruction. *Learning and Instruction, 41*(1), 41–51. https://doi.org/10.1016/j.learninstruc.2015.09.006.