Learning and Instruction 33 (2014) 131–146

Contents lists available at ScienceDirect

Learning and Instruction

journal homepage: www.elsevier.com/locate/learninstruc

Delayed benefits of learning elementary algebraic transformations through contrasted comparisons

Esther Ziegler*, Elsbeth Stern

Center for Research on Learning and Instruction, ETH Zurich, Switzerland

ARTICLE INFO

Article history:
Received 24 July 2013
Received in revised form
22 April 2014
Accepted 24 April 2014

Keywords:
Comparison
Contrasting
Mathematical knowledge
Concept knowledge
Algebra learning

ABSTRACT

Students studying algebra often make mistakes because of superficial similarities between addition and multiplication problems. In two experiments, we investigated whether these errors can be prevented by presenting addition and multiplication problems in such a way that students are encouraged to compare the problems at a deeper level. In Experiment 1, 72 sixth graders were assigned to two self-learning programs. In the contrast program, addition and multiplication were mixed and juxtaposed. In the sequential program, students first received only addition problems followed by multiplication problems. The results revealed that during the training, students performed worse under the contrast condition. However, in the follow-up tests (1-day, 1-week, 3-months), these findings were reversed: the contrast group clearly outperformed the sequential group. The findings were replicated under improved methodological conditions in Experiment 2 with 154 sixth graders. These experiments show that contrasted comparison of superficially similar but conceptually different material results in improved long-term learning.

1. Introduction

Research has provided strong empirical support for comparison as a fundamental learning mechanism that has a positive effect on meaningful conceptual learning in different areas (for an overview, see Gentner, 2010; Holyoak, 2005; Rittle-Johnson & Star, 2011). When two units (e.g., objects, problems, pictures) are juxtaposed, intentional comparison promotes a deep processing of the materials because their similarities and differences become highlighted. This phenomenon helps learners to abstract principles that may be used to solve novel problems (Catrambone & Holyoak, 1989; Gentner, 1983; Gick & Holyoak, 1983). In this way, comparisons were used to learn complex concepts by being presented two examples of the same concept that differed in their surfaces. Learners who studied two problems simultaneously outperformed learners who studied the two problems separately. This outperformance has for instance been demonstrated for negotiation principles using two different cover stories for the same negotiation strategy (Gentner, Loewenstein, & Thompson, 2003) or for the concept of heat flow using two different scenarios depicting heat flow (Kurtz, Miao, & Gentner, 2001). The direct comparison of superficially different but structurally equal (i.e., isomorphic) examples appears to help learners overcome contextual limitations, a crucial step in understanding complex concepts. The instruction to compare material offers learners a way to actively construct meaningful knowledge.

Positive effects of comparisons were also demonstrated for learning mathematical procedures in real-life school settings, e.g., when learning how to solve equations (Rittle-Johnson & Star, 2007) and for computational estimation (Star & Rittle-Johnson, 2009). Comparing two solution strategies led to greater learning gains compared to the sequential processing of these strategies. Chase, Shemwell, and Schwartz (2010) showed that learners who had to compare three cases of electromagnetic flux outperformed learners who studied the cases separately. Other research has shown beneficial effects of comparing examples of the same concept. According to Hattikudur and Alibali (2010), comparing equal signs with other relational symbols is more effective than simple instruction about the equal sign. Inventing the physical formula for density using contrasted cases was superior to being told the formula, followed by practicing with the cases (Schwartz, Chase, Oppezzo, & Chin, 2011). Moreover, it has been shown that comparing incorrect examples can facilitate learning (Durkin & Rittle-Johnson, 2012).

* Corresponding author. Research on Learning and Instruction, Institute for Behavioral Sciences, ETH Zurich, Universitätsstrasse 41, 8092 Zurich, Switzerland. Tel.: +41 446 325 820.

E-mail address: esther.ziegler@ifv.gess.ethz.ch (E. Ziegler).

http://dx.doi.org/10.1016/j.learninstruc.2014.04.006
0959-4752/© 2014 Elsevier Ltd. All rights reserved.

Although comparing two or three examples of the same concept has been shown to be beneficial for the meaningful learning of various school topics, this method is rarely used in classrooms, as for instance, in mathematics, where a great number of concepts and procedures must be learned and distinguished from each other. Therefore, we wish to examine whether the application of comparisons may be extended beyond learning single complex concepts to learning a broad range of concepts, using algebra as an example.

1.1. The challenge to acquire algebraic language competence

Mathematical competencies grow through acquiring knowledge of concepts and procedures which, over the years and under favorable conditions, build on each other and form a network of knowledge that is broadly applicable and transferable (Schneider & Stern, 2009; Stern, 1997). One challenge in learning mathematics is that the concepts are often very similar and highly related, e.g., in algebra, where students commonly have substantial difficulties learning the rule system (Blume & Heckman, 2000; Kamii & Dominick, 1997; Kieran, 1992).

Learning algebra requires not only making use of the formal language in rich contextual settings but also learning the language itself with its rules and conventions (Kieran, 2004; Kirshner & Awtry, 2004). In fact, a thorough knowledge of algebraic language provides a solid foundation and thus may be used flexibly in later applications, e.g., to successfully solve equations. Learning the language is at the core of elementary algebra. Algebraic language is best learned through frequent transformations of algebraic expressions, but this key practice is often neglected (Ottmar, Landy, & Goldstone, 2012). A common problem in managing algebraic expressions flexibly is confusion between the two dominant operations of addition and multiplication, e.g., incorrectly solving the problem "$a^2 + a^2$" as "a^4" instead of "$2a^2$".

Many authors have documented confusion errors occurring in basic transformations as well as in handling fractions and solving equations (Booth, 1988; Hirsch & Goodman, 2006; Kirshner & Awtry, 2004). In algebraic addition, like terms are summarized, e.g., "$x + x = 2x$". However, x and xy are different types and thus cannot be simplified by transformation; instead, they must be written "$x + xy$". In algebraic multiplication, like factors are summarized to exponents, e.g., "$x \cdot x \cdot x = x^3$", and unlike factors are joined to form a product, e.g., "$x \cdot y \cdot z = xyz$". Therefore, by transforming expressions, the continuity of terms in addition ($xy + x + xy + x = x + x + xy + xy = 2x + 2xy$) must be clearly distinguished from the splitting of factors in multiplication ($xy \cdot x \cdot xy \cdot x = x \cdot y \cdot x \cdot x \cdot y \cdot x = x^4y^2$). As core operations in algebraic transformations, addition and multiplication must be understood thoroughly. Instruction should clearly distinguish between these two structurally different and perceptually similar principles from the very beginning to prevent confusion and to offer a strong start in learning algebra.

Two reasons for frequent confusion between addition and multiplication in solving algebra problems are the tendency to focus on perceptual features of the problem and the tendency to hastily automate mathematical procedures. We expect systematic comparison to offer a means of overcoming both these misleading tracks.

1.2. Comparisons as a means of overcoming the tendency to focus on perceptual features

It is well documented that, when initially faced with new information, people tend to pay more attention to the salient perceptual features instead of noticing the more important structural characteristics (Chi, Feltovich, & Glaser, 1981; Gentner & Namy, 1999; Loewenstein & Gentner, 2001). In algebra, problems often appear very similar because they are composed of only letters, numbers, and signs. This similarity may induce misperceptions of the forms of the correct rules and mislead the learner's understanding of algebra rules (Kirshner & Awtry, 2004; Ottmar et al., 2012). Goldstone, Landy, and Son (2010) describe this confusion as a conflict between rule-based and perceptual processes. Without redirecting the learner's attention from the surface features to the crucial structural elements, confusion between addition and multiplication is preprogrammed. Hence, when teaching algebra, it is necessary to train students explicitly to focus on the syntactic structure, i.e., to make them recognize rules and procedures (Kirshner & Awtry, 2004; Ottmar et al., 2012).

Comparisons help to overcome this misleading tendency to focus on surface patterns (Chi et al., 1981; Gentner & Namy, 1999; Holyoak, 2005). Providing learners with the explicit instruction to compare objects may shift the focus from the surface to the deeper structural level, which results in a significant effort to detect and learn the underlying principles (Catrambone & Holyoak, 1989; Mason, 2004; Schwartz & Bransford, 1998); therefore, choosing the appropriate comparison material is a crucial factor. Depending on the presented material, certain features of the juxtaposed examples are accentuated. Winston (1975) introduced "near miss" for concepts that only differ in a small number of features. Such near-miss contrasts enhance the principle extracting and appear to be more resistant to interference effects, likely because critical convergence features are highlighted (Gick & Paterson, 1992).

Two examples may be more or less similar on the surface or structural level. During the comparison process, even the surface similarities may help to identify structural differences more easily because they make the differences more salient, which again promotes principle learning and transfer (Loewenstein & Gentner, 2001; Markman & Gentner, 1993). However, if surfaces are too dissimilar, children and even adults may miss the underlying concepts that the examples are supposed to demonstrate. This may be a problem if the learners do not have enough prior knowledge to align the examples (Gentner, 2010). The best alignment is enabled when the examples under comparison are similar both in their surfaces and in their relational structure (e.g., Richland, Morrison, & Holyoak, 2006). Thus, Gentner (2010, p. 769) suggests that "sequences of close, highly alignable exemplars should be the ideal learning situation". Inspired by this idea of repeatedly offering comparisons of similar materials, we wanted to examine whether the application of comparisons may be extended beyond learning single complex concepts to learning a broader range of principles. This might be especially promising in mathematics, where a great number of concepts and procedures must be learned and distinguished from each other.

1.3. Comparisons as a means to prevent learners from hastily automate mathematical procedures

Learning in many academic domains (particularly in mathematics) is often limited to purely superficial procedure memorization (Kamii & Dominick, 1997; NCTM, 2000). When students are repeatedly presented with similar types of problems, they may automate the solution procedure. When trying to solve new problems, a student may automatically retrieve an incorrect procedure because the problem resembles a different type of problem. An example of blind adherence to an automated procedure is illustrated when algebraic multiplication problems are incorrectly solved by adding the problem components instead of multiplying them. Students who repeatedly practiced addition rules such as "$xy + xy + xy = 3xy$" are prone to retrieve and apply this rule when faced with the similar multiplication problem "$xy \cdot xy \cdot xy =$" and may answer with "$3xy$" instead of "x^3y^3".

By analyzing Swiss and German mathematics textbooks, one predominantly finds a sequenced, blocked concept introduction, followed by extensive practice before proceeding to the next type of problem. Often, the introduction follows a progression of operational signs: first addition, then subtraction, multiplication, and division, usually followed by mixed sessions. People have the strong illusion that blocking or massing is more effective, and textbooks rely on this practice (Kornell & Bjork, 2008; Rohrer & Pashler, 2010), which may stem from learning experiences that memorizing one algorithm at time was easier than learning multiple algorithms simultaneously. If concepts are introduced in this sequential manner, they will eventually result in interferences (Anderson, 1983), which McCloskey and Cohen (1989) have described as "the sequential learning problem". In mathematics, concepts frequently build on precedents and thus often differ only in a few respects. Therefore, more similar concepts are more prone to interferences as well as to the false extrapolation or an overgeneralization of an automated rule (Kirshner & Awtry, 2004; Matz, 1982). An example of interferences or false extrapolation is when addition and multiplication procedures are confused in the same problem such as when one incorrectly splits the factors in addition, then correctly summarizes the letters: "$xy + xy + xy = 3 \cdot xy = 3xy$", and "$xy \cdot xy \cdot xy = x \cdot y \cdot x \cdot y \cdot x \cdot y = x^3y^3$", but the answer for "$xy + xy + xy$" is "$x + y + x + y + x + y = 3x + 3y$".

Because it is more demanding to memorize two algorithms simultaneously, comparison tends to suppress simple memorizing and instead forces students to actively engage in understanding the underlying structures of the presented examples (Kang & Pashler, 2012; Mitchell, Nash, & Hall, 2008; Sweller & Chandler, 1994). Comparison is often desirable for students to encounter difficulties during learning as a result of instructional impediments because it pushes learners to overcome difficulties they otherwise would not experience (Bjork, 1994; Kapur & Bielaczyc, 2012). Although this approach may lead to more errors at the beginning and slow the learning process, it tends to enhance long-term retention and transfer of the material (Rohrer & Pashler, 2010; Rohrer & Taylor, 2007). A desirable level of difficulty might consist of offering challenging material that makes it necessary for students to connect pieces of information to fully understand the problem. This type of connection is what Sweller and Chandler (1994) called the "intrinsic cognitive load", a term denoting the inherent complexity of materials. Because the complexity of materials may be described by the number of interacting elements, a higher cognitive load is required if several related elements must be processed simultaneously (Sweller, 1994). This requirement implies that complex material cannot be processed well in single elements. If this material was processed sequentially, the cognitive load would be lower, but the overall comprehension would be limited. Therefore, even though a comparison-based instructional design would increase the difficulty and the mental effort needed, contrasting may offer exactly this opportunity to interconnect different concepts by directing the students' attention to the relevant aspects of the material.

1.4. The current project

Based on the promising effects of comparison for learning a complex concept or procedure, our aim was to examine the effects of learning material that repeatedly fosters the comparison of principles that are perceptually similar and therefore easy to confuse. In our design, a series of contrasted examples were presented over several days in a real school setting; this material was tested against a traditional sequential presentation of the same material. We chose the two core principles of addition and multiplication in algebraic transformations, which are prone to confusion and error. We hypothesized that juxtaposing superficially similar addition and multiplication problems would increase the difficulty and the mental effort needed but would also make their differences more salient and thereby help to gain better mastery in the long term.

Prior research has shown that only a thoughtful selection of the units to be compared, along with explicit instruction telling the learners to pay attention, makes comparison beneficial (Catrambone & Holyoak, 1989; Schwartz & Bransford, 1998). Deep processing through comparison is ensured with the help of prompts that help focus the learners' attention, as in the instruction to self-explain examples (Chi, 2000; Renkl, 1997; Rittle-Johnson, 2006). Self-explaining has proven to encourage the active elaboration of learning materials and works especially well in combination with worked examples to create a basis for activities that promote cognitive activation. Worked examples provide a complete solution procedure: the problem, the solution steps, and the final solution. Because the result is already given, the goal for the students is to discover the underlying rules by themselves (e.g., Renkl, 2002). Moreover, worked examples may support an understanding of the problem and the underlying principle if learners are encouraged to deeply engage the problem. This engagement is what the instruction to self-explain may stimulate; therefore, we constructed our materials as a self-learning program with worked examples and prompts for self-explanation.

The learning material in our study comprised algebraic expressions that had to be simplified by strategies of algebraic transformations. The students were required to provide explicit explanations of the principles underlying the strategy of transformation. The learning gains were measured with two separate tests. The "algebraic transformation test" comprised algebraic expressions to be simplified, and assessed students' ability to correctly apply transformations. The second test, "algebraic transformation explanations", required students to write down the two algebra principles and demonstrate how to apply them; it assessed students' ability to explain how transformations are solved. We expected to find learning benefits for the contrasted introduction compared to the sequential introduction of addition and multiplication problems.

To explore the effect of contrasts on learning algebra, two experiments were conducted. In the first experiment, we tested the impact of a contrasted concept introduction on the acquisition of algebra knowledge compared to a sequential, consecutive introduction. In the second experiment, we replicated the findings of the first experiment with extended testing and an improved design.

2. Method

2.1. Participants

Sixth graders were chosen as participants because we wanted students with no prior algebra knowledge. In the Swiss mathematics curriculum, algebra is not introduced until secondary school, which starts at Grade 7; consequently, sixth graders are unlikely to have received any formal instruction in algebra, but, based on their arithmetic competencies they can be expected to be ready to learn elementary algebra principles. Participants were recruited from four urban and suburban public schools of the canton Zurich. 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 standard of school performance. All the students were volunteers, and the parents had to give their written consent. Every class was rewarded with 200 Swiss francs (approximately 150 euros), and each student received a small gift.

2.2. Design and procedure

In a 2 (group: contrast, sequential) $\times$ 3 (time: one day later, one week later, three months later) mixed-factorial design, we investigated effects on algebraic transformations and on explicit transformation explanations. Each student participated in four training sessions and in three follow-up sessions as listed in Table 1. Both groups participated in 90 min training sessions on four consecutive days, during which the students were asked to work through a self-study program. In the contrast group, students were presented worked examples in a contrasted mixed order; i.e., addition and multiplication problems were given simultaneously. In the sequential group, students practiced the problems in a sequential blocked order; i.e., all addition problems in a sequential manner for two days, followed by multiplication problems for two days. The two programs contained the same worked examples but differed in the order of the presentation of the examples and tasks (see Fig. 1 and Table 2).

The training occurred in groups of 10–15 students in rooms of the school. The students worked individually on their learning programs and sat at sufficient distances from their classmates so they could not look at each other's work sheets. They were instructed to work on their own and to ask the instructors directly if they had questions or problems with the material. All groups were trained by the first author, who was also educated as a primary school teacher, and who was present at all times to guide the training and the testing together with a research assistant.

2.3. Materials

2.3.1. Pretest on prior algebra knowledge: presented in training session 1

Although Swiss sixth graders have not yet received any formal algebra instruction, it is possible that some of them could spontaneously solve algebraic transformation problems by referring to their arithmetic knowledge. The test presented at the very beginning of session 1 was composed of the following 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 "$xy + xy + xy + xy$".

2.3.2. Self-study material: presented in training sessions 1–4

The instruction material was a paper–pencil version and consisted of several work sheets, each with worked algebra examples and a self-study part. In the first session, a slide presentation was given to explain how to read and write terms with letters and how to use the mathematical expression "raise to the power of". No other direct instruction was given; instead, students were asked to derive all the rules on their own from the worked examples with the help of the self-study instructions. Therefore, it was necessary to present the worked examples in increasing difficulty. The challenge of designing the learning material was to determine the appropriate order and content of the worked examples for each algebra concept that contained all the rules in a suitable sequence. The addition and multiplication problems chosen used the same numbers and letters and thus differentiated only in the use of addition and multiplication signs, which naturally impacted the solution steps and the results. This characteristic implied that the worked examples could be presented side by side in the contrast version of the teaching (see Fig. 2A). This side-by-side presentation highlighted the superficial similarity of the worked examples, making the underlying differences between addition and multiplication more salient to the students in the contrast group.

Several pilot studies were conducted, first with children on a one-on-one basis and then with small groups. The worked examples and tasks were changed several times until the definitive version of the learning steps was developed. Finally, there were two series of worked examples, which illustrated nine addition and nine multiplication learning steps (for an excerpt see Fig. 1). Every learning step exemplified one or two rules. The steps were balanced in the contrast group such that the corresponding steps in both addition and multiplication could be introduced simultaneously in the contrasted format. In the sequential group, all the addition steps were introduced first, followed by the multiplication steps. Finally, the addition and multiplication learning steps were assigned to the work sheets (see Table 2). For the contrast group, there were nine work sheets. The sequential group was assigned ten work sheets. In the sequential version, two blocks were provided on the same work sheet to keep the processing amount equal; however, the fifth and tenth work sheet contained only one learning step to keep from mixing addition with multiplication.

The work sheets were comprised of three sections: the worked examples, a self-explanation section, and trial tasks. At the top of

Table 1
Overview of the activities of each session.

	Session	Duration	Activities
Training sessions	1st day (Mon.)	2 lessons[a]	Pretest – prior algebra knowledge (5 min) Introduction – short slide presentation: (5 min) Training session 1: work sheets and learning tests 1–3
	2nd day (Tue.)	2 lessons[a]	Repetition test (5 min) Training session 2: work sheets and learning tests 4 + 5
	3rd day (Wed.)	2 lessons[a]	Repetition test (5 min) Training session 3: contrast group: work sheets and learning tests 6 + 7 sequential group: work sheets and learning tests 6–8
	4th day (Thu.)	2 lessons[a]	Repetition test (5 min) Training session 4: contrast group: work sheets and learning tests 8 + 9 sequential group: work sheets and learning tests 9 + 10
Follow-up sessions	1 day later (Fri.)	2 lessons	Follow-up test 1: "algebraic transformation test" and "algebraic transformation explanations" (45 min[b]) Survey part: personal data, logical reasoning test, and arithmetic test (45 min)
	1 week later	1 lesson	Follow-up test 2: "algebraic transformation test" and "algebraic transformation explanations" (45 min[b])
	3 months[c] later	1 lesson	Follow-up test 3: "algebraic transformation test" and "algebraic transformation explanations" (45 min[b])

Note.
[a] In total 90 min.
[b] 60 min in Experiment 2.
[c] 10 weeks later in Experiment 2.

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. An excerpt of the 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 student version; black-bold = the problems and the results.

each work sheet, students were presented with two blocks of two or three *worked examples*, either one block of addition and one block of multiplication (contrast group) or one or two consecutive blocks of the same operation (sequential group). An example for each group is presented in Fig. 2.

In the *self-explanation section*, students were instructed to write down how such problems are solved and to ascertain the underlying principle on their own. To facilitate this step, students were given questions prompting them to explain the worked examples. These questions were used to guide them in looking more carefully at the examples. In the contrast group, a block of addition and multiplication examples were always juxtaposed, and the questions prompted the students to compare the worked examples: e.g., "Compare the addition and multiplication examples", "Describe the different solutions", "Can you explain why students confuse addition with multiplication?", or "Give a tip to a classmate as to what he/she should pay attention to in order not to confuse addition with multiplication". In the sequential group, one or two consecutive blocks of the same operation were always processed. Because two consecutive steps disclose the same principle, comparison of the steps does not make sense; therefore, these questions prompted students to describe the worked examples in detail: e.g., "What have you noticed in the examples", "Describe exactly how to solve such problems", "What do you have to pay attention to?", or "Explain it to a classmate in detail". Students' self-explanations were always assessed by the research leader (first author). If the explanations were incorrect, students were required to correct them. If the explanations were too short, students were asked to write them out in greater detail before continuing to the next section. If students had problems in understanding the examples, they were helped by directing their attention to the corresponding worked examples and by instructing them to describe each step in the solution process. At the end of every work sheet were more self-explanation prompts asking students to write out a short reflection about what they had learned from filling out the work sheet. Contrast students were asked to consider possible confusing errors that could occur and how to prevent them. Sequential students were prompted to describe what they had to pay attention to. At the end of each day's training, students were told to summarize the rules they learned about the algebra concepts in a written explanation.

Table 2
Work sheets: presentation order of learning steps.

Work sheet	Group	
	Contrast	Sequential
1	A1 + M1	A1 + A2
2	A2 + M2	A3 + A4
3	A3 + M3	A5
4	A4 + M4	A6 + A7
5	A5 + M5	A8 + A9
6	A6 + M6	M1 + M2
7	A7 + M7	M3 + M4
8	A8 + M8	M5
9	A9 + M9	M6 + M7
10	–	M8 + M9

Note. A1–A9 = addition learning steps, and M1–M9 = multiplication learning steps.

A

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

B

Addition	
$\mathbf{xy + xy + xy} = 3 \cdot xy = \mathbf{3xy}$ $\mathbf{2b + 2b + 2b + 2b + 2b} = 5 \cdot 2b = \mathbf{10b}$ $\mathbf{3cx + 3cx} = 2 \cdot 3cx = \mathbf{6cx}$	A3
$\mathbf{c^2 + c^2 + c^2 + c^2} = 4 \cdot c^2 = \mathbf{4c^2}$ $\mathbf{a^4 + a^4} = 2 \cdot a^4 = \mathbf{2a^4}$ $\mathbf{x^3 + x^3 + x^3} = 3 \cdot x^3 = \mathbf{3x^3}$	A4

Fig. 2. Worked examples of the contrast group, example of work sheet 3 (A). Worked examples of the sequential group, example of work sheet 2 (B). Gray-italic = the intermediate steps to the solution that were marked red in the original student version; black-bold = the problems and the results.

There were two parts in the *trial tasks section*. In the first part, students were instructed to generate two to three of their own examples for each block of worked examples following the model of the presented examples. They were asked to write down the intermediate steps to the solution and to invent varied and interesting examples using other numbers and letters, e.g., "Invent two varied examples for each block! Write down all intermediate steps to the solution as in the worked examples!" The generated examples were assessed by the research leader and, if necessary, corrected by the students before continuing to the next part. In the second part, students were given trial problems to apply and deepen the learned algebra principles. Both groups were given 4–6 trial tasks per work sheet with instructions to solve them by writing down the intermediate steps in addition to the solution. Before the research leader corrected the trials, students were told to review these problems by themselves using the examples on the front of the page.

The sequential learners were given an additional training of 20 mixed addition and multiplication problems at the end of the fourth training day with the instruction, "You learned addition and multiplication in algebra, now you will receive mixed problems to solve." The problems were then checked and returned to the students for corrections. This extra training for the sequential group was designed such that the sequential learners would not be at a disadvantage and to mimic the teaching given in an ordinary school setting. After the sequential introduction of two operations, mathematics teachers usually give mixed tasks to make sure both concepts are understood correctly. While the same learning material was used for both groups, the presentation order of the material was unique to each group and thus presented some slight variation. However, all together, the tasks were balanced so that at the end of the training students in both groups processed exactly the same nine addition and nine multiplication worked examples and exactly the same number of problems (trial tasks, repetition tests, immediate learning tests).

When students had finished their work sheets they received some optional non-mathematical tasks to prevent the students from disturbing others, e.g., finding the differences between two pictures, or solving riddles or sudoku puzzles.

2.3.3. Immediate learning tests: presented in training sessions 1–4

To compare the immediate learning gains of both groups, two different types of algebra tests were given to all participants in sessions 1–4. At the beginning of each session, students were given a *repetition test*, on which they solved 6–8 algebra items similar to the ones they had worked on the day before to assess what they remembered. Immediately after every work sheet, students were given a *learning test*, on which they solved 3–8 algebra items per test sheet without being allowed to use the instruction material. These problems were similar to the ones presented on the work sheet the students had just finished. When students did not remember the examples, they were instructed to guess the solution and told that the next learning step would repeat the solution steps for such problems. This step allowed us to assess what they had learned during the session.

2.3.4. Control measures: presented in follow-up session 1

To control for individual characteristics that might affect algebra learning, four measures were assessed. *Logical Reasoning* was assessed with a figural and a numerical subtest of a German intelligence test (subtests 3 and 4 of the LPS by Horn (1983)). The test is based on Thurstone's primary mental abilities with a maximum score of 40 for each subtest. *Arithmetic Knowledge* was assessed with a speed test consisting of five sheets, each with 28–52 items. Students were required to solve each sheet within 90 s and were allowed to write only the results. The first three sheets contained simple arithmetic problems (additions with one single-digit addend, subtractions with a single-digit subtrahend, and single-digit multiplications). In addition, there were two sheets with more complex arithmetic problems (two-digit additions and multiplications). The arithmetic knowledge score was determined by the solution rates of correct answers. The students' mathematical school achievement was measured in the form of their *school grades on mathematics* and *school grades on German*. The grades were reported by students' teachers. In Switzerland, school grades range from 6 (best) to 1 (worst).

2.3.5. Algebraic transformation test: presented in follow-up sessions 1–3

The *algebraic transformations* test contained 58 items, which assessed the students' ability to correctly apply transformations practiced in sessions 1–4 (e.g., "$a^2 \cdot a \cdot ay \cdot 4a =$", "$5ab + b + 3b + 2ab + 2b =$", "$y \cdot y^3 \cdot y^2 \cdot y =$", "$2 + 5x + 4 + 2x + 3 =$"). Examples from every work sheet were represented, with 2–4 problems representing each learning step. The items were matched to the problems of the work sheets with other variables and numbers, i.e., every item was an algebraic term that needed to be transformed into the shortest version by applying the correct rules. The items were ordered in increasing difficulty but always overlapped by 2 or 3 learning steps, within which the items were randomly mixed. The same algebra test was applied in all three follow-up sessions, because students were not supposed to remember the items.

Scoring. Transformation knowledge was determined by the number of correct answers. In addition to the total score of the correctly solved problems, there was also a score of *careless errors*. Careless errors were such mistakes as miscounting the number of letters, e.g., "$n \cdot n \cdot n \cdot n \cdot n = n^4$", or an arithmetic error, e.g., "$b \cdot a \cdot 4 \cdot a \cdot 4 \cdot a = 62a^3b$".

In addition, all problems the students worked on, regardless of whether they were solved correctly, were analyzed according to the *alphabetical-order convention* and the *number-one convention*. Alphabetical-ordering is a convention used to sort letters alphabetically, which provides a better overview when there are many variables, e.g., "$u^2 \cdot ax \cdot u^2 \cdot u \cdot ax = a^2u^5x^2$", or "$n + b + n + x + b + n = 2b + 3n + x$". 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., "$z + n + n = 2n + z$" and not "$2n + 1z$", or "$b \cdot a \cdot 4 \cdot a \cdot 4 \cdot a = 16a^3b$", and not "$16a^3b^1$". Therefore, for convention errors, we assessed the number of answers with incorrect alphabetical-ordering as well as the number of answers with a superfluously written number 1. These conventions were not explicitly taught during the training, and mistakes involving them were not marked as errors in the algebra tests, because these conventions are not central to understanding and distinguishing algebraic additions and multiplications. Nevertheless, these conventions help to provide an overview when handling complex algebraic expressions and thus are practical to be learned. Students could learn the conventions incidentally while processing the work sheets, and the purpose of this analysis was to determine the extent to which learners directed their attention to superficial characteristics when focusing on structures. If the students did not pay attention to these two conventions, their attention was sometimes directed to these points, however not in a systematic way. This assistance was eliminated in Experiment 2 to make the convention-processing clearly incidental (see Section 4.1).

2.3.6. Algebraic transformation explanations: presented in follow-up sessions 1–3

In this test, students' ability to elucidate how to apply algebraic additions and multiplications was assessed. Students were asked to write down two separate descriptions explaining how to solve each type of problem. For each explanation, they were prompted with four hints, which were designed to help activate their knowledge: (a) "Describe in detailed steps how problems with letters are solved", (b) "Mention what one has to pay attention to", (c) "You can explain it by means of examples", and (d) "Imagine you would like to explain the rules to classmates".

Scoring. The two written transformation explanations of algebraic addition and multiplication were scored for accuracy and completeness of the answer. To do so, a coding scheme was developed, which is depicted in the Appendix. Two measures were assessed: (a) *Explicit transformation knowledge* was judged by the amount of correctly reported algebra concept features. For both addition and multiplication, the features were divided into defining features and secondary features, and the scores for each were added to the overall score of explicit transformation knowledge. (b) *Misconceptions* were judged by the number of errors in the transformation explanations including errors in the examples used in the explanations. The errors of both the addition and multiplication explanations were added for a total score of misconceptions.

Two trained raters independently coded the algebraic transformation explanations. The raters discussed and unified diverging judgments together. Inter-rater reliability for coding the answers was 93.8% for the first point of measurement, and 94.3% for the second point of measurement (exact agreement). Because of the high reliability there was only one of the two raters who coded the third point of measurement.

3. Experiment 1

The primary purpose was to compare two different methods of teaching students introductory algebra by using a contrasted self-study program for one group and a sequential self-study program for the second group. In the contrasted condition, addition and multiplication problems were mixed and juxtaposed, whereas in the sequential condition, addition problems were presented first, and multiplication problems were presented second. Students in the contrast group were expected to outperform students in the sequential group on both algebraic transformations (Hypothesis 1) and algebraic transformation explanations (Hypothesis 2). We also expected the gains in learning to persist in the long run one week and three months later (Hypothesis 3). Although we expected that the contrasted comparison program would especially help learners construct meaningful and applicable algebra knowledge, we also assumed that it would not affect their routines; therefore, we did not expect any differences between the numbers of careless errors made by each group. To find differences in learning trajectories between the two groups, we registered the solution rates achieved during the training sessions.

3.1. Participants

A total of 74 students (37 females) participated from three schools that had two parallel classes in the sixth grade so that one class of each school could be randomly assigned to the contrast and the other to the sequential group. Two students were excluded because they did not finish the training. Each group comprised 36 students (contrast group: $M = 12.9$ years, $SD = .5$; sequential group: $M = 12.6$ years, $SD = .5$). At the third point of measurement, three months later and after the summer holidays, we were only able to retest 65 students (32 in the contrast group and 33 in the sequential group) because the students had moved to secondary schooling and were dispersed in different classes and school levels.

3.2. Results

The results are presented in three sections: the students' preconditions, the effect of group on the follow-up tests, and the effect of group on immediate learning during the training. No significant gender differences were observed on any of the immediate or follow-up measures.

3.2.1. Students' preconditions: algebra pretest and control variables

Prior algebra knowledge. As expected, given the age of students chosen for the study, there was no difference between groups on prior algebra knowledge, $p = .96$. A floor effect was found, indicating that the students appeared to have almost no prior algebra knowledge before the training ($M = 1.9$ out of 8, $SD = 1.4$). The few correctly solved tasks showed that the students were solving almost only the algebraic addition problems with equal letters, similar to length measures they knew from school ("$a + a + a + a =$"; "$7b + 7b =$", and "$xy + xy + xy + xy =$"); however, they appeared to have no intuition about solving algebra problems that went beyond these simple additions.

Control variables. The control variables included logical reasoning, arithmetical knowledge, and grades in mathematics and German. A multivariate analysis of variance (MANOVA) revealed no significant difference between groups, $F(4, 67) = .44$, $p = .78$, *ns*. None of the separate univariate ANOVA tests showed a significant effect, all $p > .53$, indicating that both the contrast group and sequential group did not differ in cognitive preconditions nor school achievement.

3.2.2. Group differences on performance in the follow-up tests

Students in the contrast group performed better in most of the three main follow-up test measures: (a) transformation knowledge, (b) explicit transformation knowledge, and (c) misconceptions. Fig. 3 illustrates means and standard errors across the three points of measurement for both groups. For each of the follow-up test measures, separate mixed-factorial ANOVAs were conducted with group as a between-subject factor (contrast versus sequential) and time as a within-subject factor (T1: one day, T2: one week, T3: three months). When there was a main effect of group, separate post-hoc comparisons were made for the three points of measurements. Students in the contrast group were expected to perform better in all follow-up test measures. After a main effect of time, post-hoc tests were conducted to determine how stable the effects were over time.

Transformation knowledge. As expected, there was a main effect of group on the main score of algebraic transformations, $F(1, 63) = 3.35$, $p = .036$, $\eta^2 = .05$ in favor of the contrast group. Post-hoc tests revealed a tendency of group differences at T1, $t(70) = 1.61$, $p = .057$, $d = .33$, and a significant group difference at T2, $t(70) = 2.26$, $p = .014$, $d = .53$, whereas at T3, the difference between the groups was no longer significant at $p = .127$. These results indicate the advantage of contrast learning, although the effect was only weak-to-moderate and disappeared three months later. There was also a main effect of time, $F(2, 126) = 60.44$, $p < .001$, $\eta^2 = .49$, and no interaction. Simple contrasts revealed a significant change from T1 to T3, $F(1, 63) = 69.40$, $p < .001$, $\eta^2 = .52$, yet not from T1 to T2, $p = .053$, showing that the effects remained stable for one week and then declined. This decrease in performance after several weeks during which material is not repeated is normal and expected in learning environments. Again, as expected for careless errors, there was neither a group ($p = .66$), a time ($p = .45$), nor an interaction effect ($p = .33$) concerning the amount of careless errors, indicating that careless errors were independent of group and intervention.

Transformation knowledge: additional analysis of conventions. As previously described, there were two types of conventions that were not taught explicitly but were assessed in the learning analysis: the alphabetical-order convention and the number-one convention. We did not formulate hypotheses concerning these

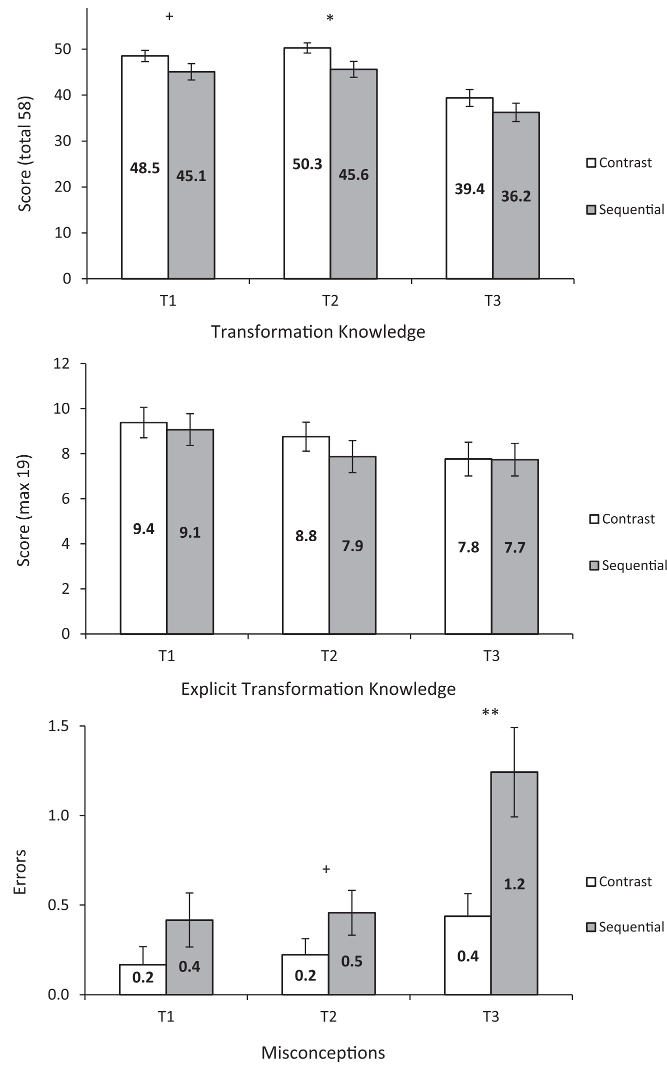

Fig. 3. Experiment 1: performance on the follow-up tests. T1 = 1 day later, T2 = 1 week later, T3 = 3 months later. $^{+}p < .10$, $^{*}p < .05$, $^{**}p < .01$. Error bars represent standard errors.

two conventions because they were not explicitly taught, and we only decided to analyze them later. For both conventions, there was no main effect of group, $p = .64$ and $p = .28$ (see Table 3 for the means and standard deviations). For the number-one convention, there was also no time effect or interaction; however, there was a main effect of time for the alphabetical-order convention, $F(1.79, 112.73) = 24.79$, $p < .001$, $\eta^2 = .28$, that was opposite of what we expected, but with no interaction. Simple contrasts revealed a significant increase at T2, $F(1, 63) = 6.96$, $p = .011$, $\eta^2 = .10$, but a reduction at T3, $F(1, 63) = 18.67$, $p < .001$, $\eta^2 = .23$, indicating an improvement for both groups at T3. These deviant time effects indicate that over the three months between the first and last follow-up test, the conventions were remembered and followed better over time.

Explicit transformation knowledge. Unexpectedly, there was no significant main effect of group in the main score of the algebraic transformation explanations, $p = .39$ (see Fig. 3); however, there was a main effect of time, $F(2, 124) = 3.47$, $p = .034$, $\eta^2 = .05$. Simple contrasts showed a decrease from T1 to T2, $F(1, 62) = 4.33$, $p = .041$, $\eta^2 = .07$, and from T1 to T3, $F(1, 62) = 5.42$, $p = .023$, $\eta^2 = .08$, indicating that the algebra concept knowledge initially acquired was not sustained. There was no interaction, showing that both groups' scores decreased equally.

Misconceptions. As expected, there was a main effect of group in the misconceptions, $F(1, 62) = 8.80$, $p = .002$, $\eta^2 = .12$ (see Fig. 3). Post-hoc tests revealed tendencies of group differences in favor of the contrast group at T1, $t(70) = 1.51$, $p = .065$, $d = .25$ and at T2, $t(70) = 3.05$, $p = .002$, $d = .50$, and a clear difference three months after the training at T3, $t(70) = 4.02$, $p < .001$, $d = .66$. There was a main effect of time, $F(1.79, 110.80) = 11.81$, $p < .001$, $\eta^2 = .16$, with the simple contrasts revealing a significant increase of errors from T1 to T3, $F(1,62) = 16.68$, $p < .001$, $\eta^2 = .21$, and a significant interaction, $F(1.79, 110.80) = 3.38$, $p = .043$, $\eta^2 = .05$, showing that the increase in the number of misconceptions during this time period was more pronounced in the sequential group.

In summary, students in the contrast group showed better results in algebraic transformations than the sequential group, with clear gains on the follow-up tests one day and one week later, but there were no longer differences in sustained knowledge three months later. The students in the contrast group also did not perform better in the section of written explanations. Conversely, the contrast group's lower score on misconceptions in the long-term indicates better knowledge about both algebraic addition and multiplication.

3.2.3. Group differences on performance during the training

The previous section revealed a clear superiority of the contrast group on most of the follow-up test measures. The question now arises: was this result predictable from the students' achievement during the training, or does it only appear delayed? By analyzing group differences in the two immediate learning tests presented during the training, we find a reversed effect with a better performance of the sequential group.

Immediate learning gains. A multivariate analysis of variance (MANOVA) revealed a significant effect of group, $F(2, 69) = 9.27$,

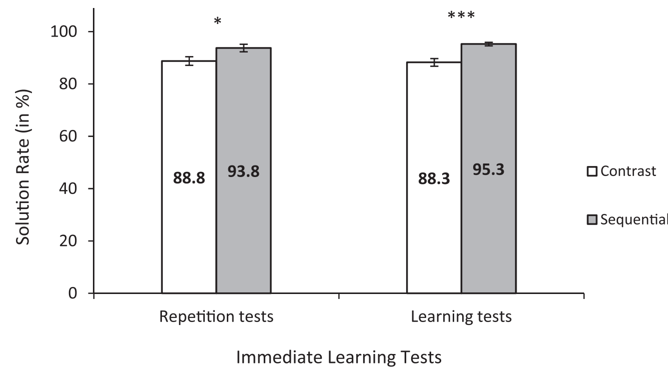

Fig. 4. Experiment 1: performance on the immediate learning testes. *$p < .05$, ***$p < .001$. Error bars represent standard errors.

$p < .001$, $\eta^2 = .21$. Surprisingly, students in the sequential group outperformed students in the contrast group on both immediate learning measures (see Fig. 4 for the means and standard errors). The separate univariate ANOVA tests on the outcome variables revealed significant treatment effects on repetition tests, $F(1, 70) = 5.15$, $p = .026$, $\eta^2 = .07$, and on learning tests, $F(1, 70) = 18.80$, $p < .001$, $\eta^2 = .21$. Therefore, with regard to short-term learning, the results show a clear advantage of the sequential group.

3.2.4. Does general reasoning ability moderate learning gains?

We wanted to test whether higher or lower-achieving students benefited more from the contrasted introduction. The results were examined for the influence of intelligence (logical reasoning ability) by contrasting the results of the upper half with the lower half of the participants. We conducted a 2 (group: contrast versus sequential) × 2 (ability: high versus low) × 3 (time: one day, one week, three months) ANOVA on the follow-up test measures.

No significant interaction was found for any of the five follow-up measures: transformation knowledge ($p = .76$), explicit transformation knowledge ($p = .41$), misconceptions ($p = .22$), the alphabetical-order convention ($p = .60$), and the number-one convention ($p = .96$). With respect to the two immediate learning measures, no interaction was found for the repetition tests was non-significant ($p = .30$). However for the learning test the interaction was significant ($p = .008$), with the graphs revealing that the low-achieving contrast learners performed especially poor. Checking the means of the learning test revealed high solution rates for the sequential learners (96.7% for the high-achievers, 93.2% for the low-achievers), implicating limited variance caused by ceiling effects. More importantly, however, the low-achieving contrast learners entirely compensated their disadvantage in the follow-up tests. Altogether, there is no evidence from our data set that the extent to which students gain from comparison instruction depends on the individual characteristics of logical reasoning.

3.3. Discussion of Experiment 1

The results from Experiment 1 showed that the contrasted training led to greater gains in algebraic transformation knowledge

Table 3
Experiment 1: means and standard deviations (in parentheses) of students' performance by group on the convention errors of the transformation tests.

	T1		T2		T3	
	Contrast	Sequential	Contrast	Sequential	Contrast	Sequential
Alphabetical-ordering	10.2 (4.0)	8.9 (4.3)	10.8 (4.6)	9.9 (4.8)	6.6 (4.1)	7.5 (4.9)
Number-one	1.3 (1.6)	1.6 (3.1)	1.2 (2.0)	2.3 (3.1)	1.9 (3.1)	1.8 (2.7)

Note. T1 = 1 day later, T2 = 1 week later, T3 = 3 months later.

compared to the sequential training, with the strongest effect found one week after the training (Hypothesis 1). Three months after training, however, this effect disappeared, although the means remained higher in the contrast group (Hypothesis 3). The test conditions at this third point of measurement (after summer holidays and dispersed in different classes) were not as controlled as they were for the first two follow-up tests. The fact that the contrast group remained slightly better at this third point of measurement was a hint that there might be a long-term advantage for the contrasted learning.

Students in the contrast group also showed fewer misconceptions in the written explanations than students in the sequential group; however, there was no group difference in the total score of explicit transformation knowledge (Hypothesis 2). During the testing, it became obvious that many students did not like writing down the written explanations and thus did not write them in very much detail. This result may show that students have more explicit knowledge than they are willing or able to express. Time restrictions and insufficient instruction might have been responsible for the fact that students' written explanations were not as informative and complete as they could have been. Moreover, writing explanations may be hard for students, as knowledge is often intuitive and only implicitly represented, which makes it difficult to formalize (von Aufschnaiter & Rogge, 2010; Bou-Llusar & Segarra-Ciprés, 2006). Therefore, it is very important to direct students to express their knowledge and to prompt them appropriately. For our next experiment, we plan to reject minimally written explanations and to request more detailed descriptions to examine this question. We also decided to give students more time during the follow-up tests so they could write their explanations more carefully and have time to review the test after finishing it.

A remarkable finding of Experiment 1 was that the contrast group, which outperformed the sequential group in transformation knowledge and misconceptions, showed clearly worse results in immediate learning during the four days of training. This pattern of delayed benefits corresponds to retention effects found for intermixed concept learning (e.g., Rohrer & Pashler, 2010). Good performance during training often leads teachers to mistakenly conclude that sequential learning is favorable; however, with regard to learning and instruction, knowledge acquisition must be considered on a long-term basis. The fewer errors made by sequential learners revealed that the training was less challenging for them and that they did not need to invest as much effort to learn the material as the contrast learners did. We also sensed through casual observation that more sequential learners finished their work sheets earlier, although we did not assess this phenomenon systematically. Therefore, the facility of learning in the training phase appears to be a poor indicator of long-term learning gains. The fact that the contrast students had higher error rates may indicate that some errors must be made for meaningful learning to occur.

In the trial tasks and the repetition tests, all the students received a "right" or "wrong" feedback on the problems they completed, and they had to correct the problems that had been marked "wrong". Corrections had to be made more often by the contrast learners because they made more errors. This feedback and the resulting corrections may have helped them deepen their processing, and allowed them to consciously distinguish between the two algebra principles. Feedback is seen as an important means of instruction (e.g., Hattie & Timperley, 2007). Although our feedback was given as a simple mark of "right" or "wrong", it entailed revising the incorrect problems and we cannot exclude that it contributed to the effect of contrasted learning.

A limitation in the sampling procedure that may have contributed to the moderate effects of Experiment 1 was that the classes were matched as a whole to the same group. Although there were no group differences in the control variables, some classes were more disciplined and motivated to achieve, whereas others needed more guidance and control. This difference may have been due to different styles of the teachers, but we did not systematically assess this phenomenon. Despite the limitations discussed above, the findings of the first experiment supported the contrasted training as a promising introductory method for teaching two similar principles in algebra with long-term benefits. Hence, we decided to replicate the first experiment under improved conditions and a controlled, within-class matching to the groups to see if the results would be the same and highlight the superiority of a contrasted and mixed training.

4. Experiment 2

The two-group experimental design, the hypotheses, the training materials, and the tests were the same as in Experiment 1 (see Sections 2.2, 2.3 and 3), whereas the way of assigning the participants to the groups and some details of the procedure were changed.

4.1. Procedure

In Experiment 2, there were slight differences in the procedure and in the instruction.

Assignment of the participants to the groups. In Experiment 1, assignment to the two conditions was performed using entire classes. To avoid selection effects, a within-classroom assignment was used to assign groups in Experiment 2. For each class, half the students were assigned to the contrast group, and the other half were assigned to the sequential group. As in Experiment 1, the students were matched according to their grades in mathematics and German.

Time allocated to the follow-up tests. Students were given an extra 15 min on the follow-up tests, i.e., 60 min instead of 45 min, so they could work through the test without time pressure, review it and make corrections.

Time between the second and third presentation of the follow-up test. The third follow-up test was conducted 10 weeks after the training instead of 3 months, so that the test would not interfere with the summer holiday.

Follow-up test instructions. To ensure the thorough processing of the materials and to improve the reliability of the tests, more detailed and precise instructions were provided. First, students were instructed to write down the intermediate steps to the solutions of the algebraic transformation problems, which only some of the students did in Experiment 1. Second, after finishing the transformation test, students were instructed to review all their answers for omitted problems and careless errors because some students skipped a problem or an entire sheet in Experiment 1. Third, if the algebraic transformation explanations did not have enough detail, students were asked to expand their answers.

Elimination of assistance. In Experiment 1, students were sometimes given help when they had difficulties. For instance, with respect to self-explanations, students were helped when the experimenter directed their attention to the corresponding worked examples and tell them to describe what to do first, second, and third. In addition, concerning the trial tasks, some students were given hints to pay attention to the alphabetical-ordering convention and the number-one convention. Again, this help was only given to some students and not provided in a controlled way. The assistance was never given during the assessment in the follow-up testing. In Experiment 2, we decided to eliminate this assistance:

students were not given any hints to notice conventions or to describe the steps needed to reach a solution.

For self-explanations, the feedback was standardized to the following: (a) “This is incorrect, look at this worked example” or (b) “This is not enough detail, can you supplement it?” For the trial tasks, the feedback was standardized to the following: (a) “This is incorrect, check it again” and (b) “This is incorrect, look at this worked example”.

4.2. Participants

A total of 157 sixth graders (84 females) from six schools participated. The participation conditions were the same as in Experiment 1. Three students did not finish the training and were excluded. For each class, students were first assigned randomly to the two groups. If there was an imbalance in the students' grades in German and mathematics, the assignment of the students was aligned. The contrast group comprised 79 students ($M = 12.4$ years, $SD = .5$), and the sequential group comprised 75 students ($M = 12.3$ years, $SD = .5$).

4.3. Results

As in Experiment 1, the results are presented in three sections. Again, no significant gender differences were observed on any of the immediate or follow-up measures.

4.3.1. Students' preconditions: algebra pretest and control variables

Prior algebra knowledge. There was no difference between groups concerning prior algebra knowledge, $p = .97$. Compared to Experiment 1, students' mean solution rates on the pretest were even lower ($M = .8$ out of 8, $SD = 1.0$). Only 23% of the students solved more than one task correctly, and none of the students solved more than three tasks. This finding strongly confirmed that the students had negligible direct prior algebra knowledge and no intuition about how to solve algebra problems.

Control variables. There was no significant difference between the two groups in their logical reasoning, arithmetical knowledge, and grades in mathematics and German, $F(4, 149) = .44$, $p = .78$, *ns*. None of the separate univariate ANOVA tests showed a significant effect, all $p > .48$.

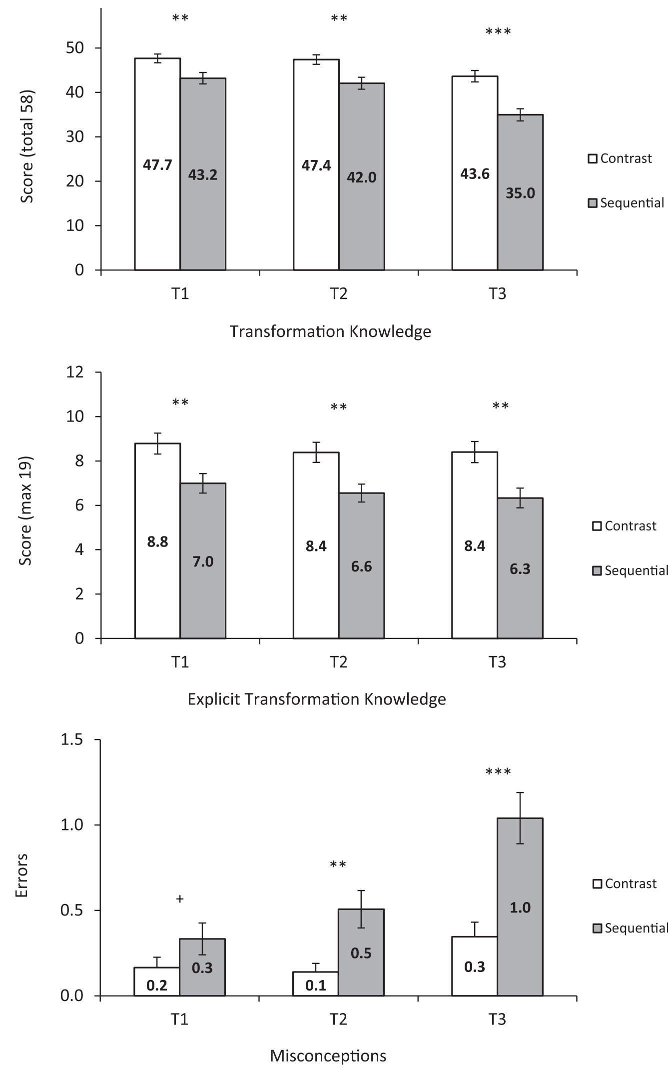

Fig. 5. Experiment 2: performance on the follow-up tests. T1 = 1 day later, T2 = 1 week later, T3 = 10 weeks later. $^{+}p < .10$, $^{**}p < .01$, $^{***}p < .001$. Error bars represent standard errors.

4.3.2. Group differences on performance in the follow-up tests

Students in the contrast group performed better in most of the three main follow-up test measures: (a) transformation knowledge, (b) explicit transformation knowledge, and (c) misconceptions. Fig. 5 illustrates means and standard errors across the three points of measurement for both groups. As expected, the results of Experiment 1 were replicated with improved significance levels. The same statistical analyses used in Experiment 1 were used in Experiment 2.

Transformation knowledge. As expected, there was a main effect group, $F(1, 151) = 15.59$, $p < .001$, $\eta^2 = .09$, in favor of the contrast group. Post-hoc tests revealed significant group differences at all points of measurements, at T1, $t(152) = 2.79$, $p = .003$, $d = .46$, at T2, $t(152) = 3.08$, $p = .002$, $d = .50$, and at T3, $t(151) = 4.65$, $p < .001$, $d = .76$, indicating an advantage of contrast learning with moderate-to-strong effects. There was also a main effect of time, $F(1.38, 207.89) = 78.48$, $p < .001$, $\eta^2 = .34$, and a significant interaction, $F(1.38, 207.89) = 7.24$, $p = .003$, $\eta^2 = .05$. Simple contrasts also showed a decline of the results over time from T1 to T2, $F(1, 151) = 4.49$, $p = .036$, $\eta^2 = .03$, and from T1 to T3, $F(1, 151) = 91.84$, $p < .001$, $\eta^2 = .38$, although the significant interaction effect shows that the decrease over time is more pronounced for the sequential group than for the contrast group. As expected, there was no difference between groups in the number of careless errors, $p = .52$.

Transformation knowledge: additional analysis of conventions. Based on the results from Experiment 1, we did not expect any differences in the convention errors. Surprisingly, however, the contrast group performed worse on these measures than the sequential group (see Table 4 for the means and standard deviations). There was a main effect of group for alphabetical ordering, $F(1, 151) = 19.56$, $p < .001$, $\eta^2 = .12$, with significant post-hoc tests at all points of measurements, at T1, $t(152) = 2.64$, $p = .009$, $d = .43$, at T2, $t(152) = 3.50$, $p = .001$, $d = .57$, and at T3, $t(151) = 5.25$, $p < .001$, $d = .86$. There was a time effect, $F(1.64, 248.29) = 33.82$, $p < .001$, $\eta^2 = .18$, with the simple contrasts showing a reduction in errors at T3 for both groups, $F(1,151) = 34.88$, $p < .001$, $\eta^2 = .19$. There was also an interaction of Group × Time, $F(1.64, 248.29) = 6.35$, $p = .004$, $\eta^2 = .04$, that indicated a less pronounced decline for the contrast group at T3. For the number-one convention, the contrast group was slightly worse on all points of measurement; however, the effect was not significant, $p = .13$. Only the time effect was found to be significant, $F(1.76, 265.77) = 12.77$, $p < .001$, $\eta^2 = .08$, with the simple contrasts showing a decline of errors at T3 for both groups, $F(1,151) = 14.58$,

Table 4
Experiment 2: means and standard deviations (in parentheses) of students' performance by group on the convention errors of the transformation tests.

	T1		T2		T3	
	Contrast	Sequential	Contrast	Sequential	Contrast	Sequential
Alphabetical-ordering	11.8 (4.2)	9.9 (4.8)	12.5 (4.2)	9.8 (5.2)	11.0 (5.1)	6.9 (4.6)
Number-one	4.7 (4.1)	3.9 (3.4)	4.7 (4.1)	4.1 (3.3)	3.9 (3.5)	2.9 (2.9)

Note. T1 = 1 day later, T2 = 1 week later, T3 = 10 weeks later.

$p < .001$, $\eta^2 = .09$. The results suggest that the sequential group performed better in maintaining the two conventions, which had not been explicitly taught.

Explicit transformation knowledge. Different from Experiment 1, there was a significant main effect of group, $F(1, 151) = 12.62$, $p = .001$, $\eta^2 = .08$ (see Fig. 5). Significant post-hoc tests at all points of measurement showed that students from the contrast group expressed more explicit knowledge in their written explanations than did students from the sequential group, at T1, $t(152) = 2.77$, $p = .003$, $d = .45$, at T2, $t(152) = 3.01$, $p = .002$, $d = .49$, and at T3, $t(151) = 3.20$, $p = .001$, $d = .52$. Unlike Experiment 1, there was no time effect and no interaction, indicating that the explicit transformation knowledge was sustained over time in both groups.

Misconceptions. Similar to Experiment 1, the contrast group had fewer misconceptions than the sequential group $F(1, 151) = 17.11$, $p < .001$, $\eta^2 = .10$ (see Fig. 5). Post-hoc tests revealed significant group differences at T2, $t(70) = 3.05$, $p = .002$, $d = .50$, and at T3, $t(70) = 4.02$, $p < .001$, $d = .66$. At T1, there was only a tendency, $t(70) = 1.51$, $p = .067$. There was a time effect, $F(1.83, 276.16) = 17.68$, $p < .001$, $\eta^2 = .11$, with the simple contrasts showing an increase from T1 to T3, $F(1,151) = 27.14$, $p < .001$, $\eta^2 = .15$, and an interaction of Group × Time, $F(1.83, 276.16) = 5.46$, $p = .006$, $\eta^2 = .04$. The graph revealed that the increase of misconceptions was more pronounced in the sequential group.

4.3.3. Group differences during the training

Experiment 1 revealed superior performance of the sequential group in the immediate tests presented during the training. We were interested in determining whether this finding would be replicated in Experiment 2.

Immediate learning gains. In accordance with Experiment 1 the MANOVA revealed a significant effect of group on immediate learning gains, $F(2, 151) = 27.75$, $p < .001$, $\eta^2 = .27$, in favor of the sequential group (see Fig. 6). Separate univariate ANOVA tests on the outcome variables revealed significant group effects on both variables: on repetition tests, $F(1, 152) = 31.84$, $p < .001$, $\eta^2 = .17$, and on learning tests, $F(1, 152) = 52.26$, $p < .001$, $\eta^2 = .26$, indicating a clear advantage for the sequential group on the short-term learning outcome.

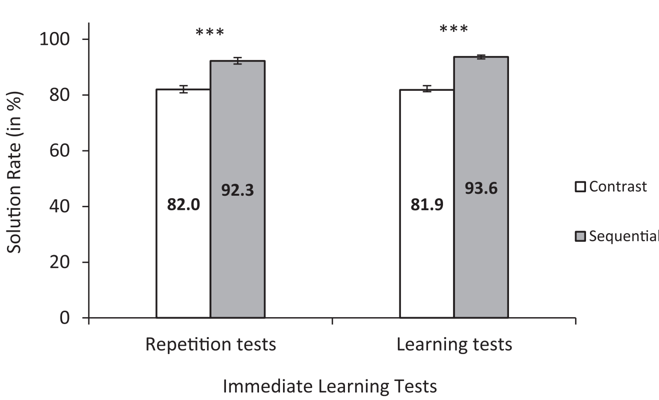

Fig. 6. Experiment 2: performance on the immediate learning tests. ***$p < .001$. Error bars represent standard errors.

4.3.4. Does general reasoning ability moderate learning gains?

Similar to Experiment 1, the examination for the influence of logical reasoning ability by median split showed no significant treatment effects for all the follow-up measures: transformation knowledge, ($p = .28$), explicit transformation knowledge ($p = .81$), misconceptions ($p = .21$), the alphabetical-order convention ($p = .35$), and the number-one convention ($p = .85$). Also the pattern of the immediate learning measures was similar to Experiment 1: no significant interaction was found for the repetition tests ($p = .83$), while the interaction for the learning test ($p = .014$) was significant. The graphs again revealed disadvantages for the low-achieving contrast learners which, however, were compensated in the follow-up. The solutions rates of the learning tests were again high for the sequential learners (94.5% for the high-achievers, 92.6% for the low-achievers) which underline potential ceiling effects. Altogether, also the results of Experiment 2 suggest that students of all abilities benefited more from the contrasted than from the sequential exposure to material.

4.4. Summary of Experiment 2

Students in the contrast group again performed better on transformation knowledge with a strong effect on the ten-week follow-up test (Hypothesis 1). After the more consequent instruction, the two not-explicitly-taught conventions were better perceived by sequential exposure to the material. Concerning explicit transformation knowledge, under the improved conditions for more detailed transformation explanations in Experiment 2, students in the contrast group clearly performed better with significant differences at all measurement points (Hypotheses 2 and 3). Most remarkably, students in the contrast group again scored clearly lower in the two immediate learning measures. Therefore, we may conclude that, as expected, the results of Experiment 2 were essentially the same as Experiment 1, but with clearer results.

5. General discussion

The aim of this study was to investigate the impact of repeated contrasted comparisons on algebra learning in a classroom setting. In two experimental training studies, we showed that juxtaposing addition and multiplication algebra problems in a self-learning program is a promising method for supporting early secondary school students in learning elementary principles of algebraic transformations. Both experiments revealed that the students who processed the contrasted, mixed program performed better in differentiating superficially similar algebra principles than the students who received the more conventional sequential teaching materials (Hypotheses 1 and 2). There was a significant decline in performance after several weeks, but this finding was expected, as learning was not continued. Nevertheless, the persistence of the group difference, despite the decline in performance with its strong effect over ten weeks, is important (Hypothesis 3). These results strengthen the robustness of previous findings regarding the comparison effect on gaining problem solving knowledge (Rittle-Johnson & Star, 2007); however, both experiments revealed that the superiority of the contrasted learning program could not be predicted from the students' performance during the training. Rather, in both experiments, the contrast group made more errors on the immediate tests compared to sequential group.

Some of the limitations of Experiment 1 were overcome in Experiment 2 by randomizing the participants on an individual level rather than on a classroom level and by optimizing the instructions and the timing of the follow-up tests. As a consequence, although the findings were similar in both experiments, they were more pronounced in Experiment 2. This finding especially applies to explicit transformation knowledge, with clear gains found for contrast learners, who concurrently perform better on all three follow-up measures. This finding may confirm that students in Experiment 1 did not write all they knew in the written explanations; therefore, in Experiment 2, when given more time and prompted to write in greater detail, students appear to express more of their knowledge – an important point to consider when designing tests.

Our findings of delayed benefits for contrasted mixed learning represent a confirmation of previous results of mixed practice, which has been demonstrated with categorization tasks or mathematical rule and formula learning (e.g., Kang & Pashler, 2012). We have contributed with an experimentally controlled intervention study, which was run under realistic classroom conditions over several days and included a ten-week follow-up test. Our results have highlighted misleading discrepancies between short-term and long-term results; i.e., the short-term benefit and the long-term disadvantage of the conventional sequential introduction of algebraic addition and multiplication. This finding emphasizes the importance of follow-up studies investigating long-term learning gains instead of relying on immediate learning gains. Moreover, it reveals the importance of organizing and assessing learning from a long-term perspective.

Another benefit of contrasted learning may be splitting the learner's attention between addition and multiplication from the very beginning. The switching between concepts and procedures helps the learner become aware of the existence of the two concepts that must be distinguished. This effect may lead to the learner's making more conscious choices when using memorized procedures and may reduce the random, overly automatic retrieval of memorized procedures when problem solving. The higher number of errors observed in the contrast group during the training period may be a concern for teachers who do not like to overburden students, especially lower-achieving students; however, our results revealed that even though contrasting is a more demanding instruction method, the added challenge also appears to provide an advantage.

In our data, there was no indication that either lower- or higher-ability students benefited more from the contrast materials. These results confirm the lack of evidence for aptitude-treatment interactions, which are scarcely found despite the increasing interest in individual differences (Scott, 2013). Studies have established the interactions of learning gains in a field with prior knowledge in this field (Day & Goldstone, 2010; Hiebert & Wearne, 1996; Rittle-Johnson, Durkin, & Star, 2009). For example, Rittle-Johnson and colleagues found that comparing is only beneficial if students have had some prior knowledge on a topic. In contrast, our participants performed very poorly on prior algebra knowledge; they appeared to have negligible prior knowledge and no intuitive algebra comprehension. The algebra material in our experiments was new for all students, and contrasting was their first contact with the new topic; therefore, we can recommend the practice of comparing and contrasting from the beginning of a student's training.

The additional analyses of students' transformation errors suggested that the two training programs shaped learners' attention in different ways. Particularly in Experiment 2, students from the sequential group adopted the not-explicitly-taught number-one and alphabetical-order conventions better. Learners of both groups could incidentally deduce these conventions from the worked examples, but the sequential learners made more use of this information. This finding suggests that sequential students had unused capacities and were better able to process these secondary superficial characteristics of the material, although these capacities may have distracted students from completely learning the more crucial core features for distinguishing the two concepts over an extended period of time.

Together with the aforementioned better performance in immediate learning, it appears that sequential learning is less demanding. This finding also corresponds with the results of reduced intrinsic cognitive load in sequential learning (Sweller & Chandler, 1994; Sweller, Van Merrienboer, & Paas, 1998) and with the demand of desirable difficulties (Bjork & Bjork, 2011; Kester, Paas, & Van Merriënboer, 2010). Desirable difficulties are those that increase the intrinsic cognitive load or complexity of materials in such a way that it helps learners to focus their attention on critical aspects of the material. The principles of algebraic addition and multiplication are perceptually very similar, comparable to near-miss contrasts (Gick & Paterson, 1992; Winston, 1975). Desirable difficulties in our algebra material are the difficulties that help students to distinguish between the two operations and the corresponding rules.

The superiority of the contrasted practice compared to the sequential practice was surprisingly clear. We ascribe the clear learning gains to the explicit comparisons the contrast learners were prompted for. The juxtaposition may have highlighted the underlying core principles (Gick & Paterson, 1992; Loewenstein & Gentner, 2001; Mitchell et al., 2008), whereas the explicit comparisons may have forced students to connect the two similar principles and to become aware of the features and rules that help to distinguish them (Sweller & Chandler, 1994). Kang and Pashler (2012) investigated the difference between the interleaved and simultaneous presentation of different concepts. Their simultaneous group showed slightly, though not significantly, better results than their interleaved group. In our instruction, not only were algebra problems processed simultaneously, but students were explicitly prompted to compare and describe addition and multiplication differences, which may have been the main reason for the strong contrast effect. Our design does not allow the full disentanglement of the impact of explicit comparison from the impact of desirable difficulties caused by mixed presentation. Therefore, we cannot exclude that the effect may be due entirely to desirable difficulty. By including a condition with mixed problem presentations, which does not explicitly request comparisons, we could have tested whether solely simultaneous processing of addition and multiplication problems would have caused learning gains. However, it was not the goal of this study to demonstrate the efficacy of comparison, as this has already been done extensively in many other studies. In our study, the prompts to compare for contrast learners were implemented to ensure the deep processing of the materials. Similarly, the sequential learners were prompted for explicit and thorough explanations of their sequential materials. The instruction "to compare" and "to explain in detail" had the same intention: to guarantee deep processing.

5.1. Limitations and future directions

Contrast learners showed clear gains in the ability to apply transformations and in the explicit availability of transformation knowledge. Because of the already loaded design, we had no additional measure to assess conceptual understanding, e.g., as questions about the principles, or ratings of definitions

(Rittle-Johnson & Star, 2007). Nevertheless, we assert that the ability to distinguish and correctly apply the two algebra principles of addition and multiplication is a clear sign of understanding the syntactic structure of algebra. This seeing of the structure is crucial for algebra learners and forms the basis for a later flexible use in applications (Kirshner & Awtry, 2004; Ottmar et al., 2012). Further research should differentiate whether and how the thorough distinction of the two principles from the beginning will influence a broader understanding of algebra.

Further controlled studies are needed to replicate the present results with other topics before deciding whether textbooks should be changed so they introduce similar concepts in a contrasted format. We intentionally restricted the investigation to algebraic additions and multiplications and did not include subtractions and divisions so as not to overload the experimental setting, but we would expect subtraction and division to be analogous to addition and multiplication, as they are each operations of the same level (Kirshner & Awtry, 2004). This phenomenon should be tested in a classroom setting. We also suggest the further examination of contrasted comparisons with many algebraic and mathematical topics, an area where we would expect broad application possibilities. It is especially important to investigate all characteristics of compared examples in greater depth.

In our experiments, contrasting was applied in a self-learning program; however, a very common instruction technique is direct instruction. Therefore, it would be interesting to investigate a learning environment in which students receive direct instruction with a simultaneous and contrasted introduction of two concepts, i.e., a contrasted direct instruction compared to a sequential direct instruction. Only further research will show whether our results are an algebra-specific or a method-specific effect and whether contrasting may be established as general method of instruction for introducing similar and frequently confused concepts.

Acknowledgments

We are grateful to Sara Ziegler for her assistance in the implementation of the Experiments in the school classes and the evaluation of the test materials. We thank Theresa Treasure for helpful comments on earlier versions of this paper.

Parts of this article report findings from the author's unpublished doctoral dissertation accepted at the ETH of Zurich, Switzerland.

Appendix

Coding scheme for the follow-up test "Algebraic transformation explanations":

For every one of the following features mentioned in the answer, one point was scored. The features were divided into defining features and secondary features.
The defining features were the principles that were considered central and essential for distinguishing algebraic addition from algebraic multiplication.
The secondary features were external features and conventions, e.g., how to report the result.

Points were equally given if: (a) a key element was explicitly mentioned, or (b) a key element became visible in the solution or the intermediate steps of the example.
If a point was described incorrectly, it was counted as a mistake.

Sub-analysis of addition:

Defining features (total 6 points) (d = defining feature, s = secondary feature):

d1: Sorting by summands (summands are the letter endings (sometimes with exponents: a, ab, a^2, …))
→ only 1/2 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
d2: Summands are not split
d3: Summands do not change in the result, exponents remain ($a^3 + a^3 = 2a^3$), no point if only with single letters
d4: Letters with different exponents are not summarized, merged, or unified ($x + x^2$)
d5: Single letters are not summarized, merged, or unified with double letters ($a + ab$)
d6: Letters and numbers are not summarized, merged, or unified ($x + 4$)

Secondary features (total 4 points):

s1: In the result, there are groups with different summands ($2bc \cdot 3b \cdot 2c$)
s2: In the results, the different summands are connected by a plus sign ($2a + 3b$)
s3: The number of letters stays directly in front of the letter ($5a$)
s4: A single letter counts as 1 ($c = 1c$)

To summarize, an exemplary instruction for the concept of algebraic addition:

1. Sorts according to identical summands. Pays attention to double letter variables and variables with exponents which are added unchanged, i.e., cx and c^2 are not separated. This underscores the identical combinations of letters.
2. Each type of unit (cx, c^2, c, 2) is added separately by adding the numbers of summands and by appending the unit unchanged. Detached numbers are added separately.
3. The result consists of the number of each different unit connected with a plus sign.

Sub-analysis of multiplication:

Defining features (total 5 points) (d = defining feature, s = secondary feature):

d1: All factors are split
→ only 1/2 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)
d2: Letters with exponents are separated into single letters ($x^2 = x \cdot x$)
d3: Double letters are separated ($ab = a \cdot b$)
d4: Numbers and letters are separated ($2z = 2 \cdot z$)
d5: Exponents are added ($c^2 \cdot c^2 \cdot c^2 = c^6$)

Secondary features:

s1: In the result, everything is merged together into one expression ($16a^2b^4$)
s2: In the result, the numbers come first
s3: The number of equal letters is noted as an exponent behind and above the letter (a^3)
s4: Numbers are multiplied

Summarized, an exemplary instruction for the concept of algebraic multiplication:

1. Splits the term into single numbers and letters. Pays attention to split units like ab, b^2, and $2a$ into their components.
2. Puts numbers together, and always puts identical letters together
3. Multiplies all the numbers and writes the amount at the first position. Then counts the number of every letter and writes their number as an exponent after the corresponding letter. Writes them one after the other without any intervening spaces.

References

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

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

Bjork, R. (1994). Institutional impediments to effective training. In D. Druckman, & R. A. Bjork (Eds.), *Learning, remembering, believing: Enhancing human performance* (pp. 295–306). Washington, DC: National Academy Press.

Bjork, E., & Bjork, R. (2011). Making things hard on yourself, but in a good way: creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56–64). New York: Worth Publishers.

Blume, G. W., & Heckman, D. S. (2000). Algebra and functions. In E. A. Silver, & P. A. Kenney (Eds.), *Results from the seventh mathematics assessment* (pp. 225–277). Reston, VA: National Council of Teachers of Mathematics.

Booth, L. R. (1988). Children's difficulties in beginning algebra. In A. F. Coxford, & A. P. Shulte (Eds.), *The ideas of algebra, K-12* (pp. 20–32). Reston, Virginia: 1988 NCTM Yearbook, National Council of Teachers of Mathematics.

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. http://dx.doi.org/10.1108/13673270610679390.

Catrambone, R., & Holyoak, K. J. (1989). Overcoming contextual limitations on problem-solving transfer. *Journal of Experimental Psychology – Learning Memory and Cognition, 15*(6), 1147–1156. http://dx.doi.org/10.1037/0278-7393.15.6.1147.

Chase, C. C., Shemwell, J. T., & Schwartz, D. L. (2010). *Explaining across contrasting cases for deep understanding in science: An example using interactive simulations*. Paper presented at the Proceedings of the 9th International Conference of the Learning Sciences.

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, NJ: Lawrence Erlbaum Associates, Inc.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*(2), 121–152.

Day, S., & Goldstone, R. L. (2010). *The effects of similarity and individual differences on comparison and transfer*. Paper presented at the Thirty-Second Annual Conference of the Cognitive Science Society. Portland, Oregon: Cognitive Science Society.

Durkin, K., & Rittle-Johnson, B. (2012). The effectiveness of using incorrect examples to support learning about decimal magnitude. *Learning and Instruction, 22*(3), 206–214. http://dx.doi.org/10.1016/j.learninstruc.2011.11.001.

Gentner, D. (1983). Structure-mapping – a theoretical framework for analogy. *Cognitive Science, 7*(2), 155–170. http://dx.doi.org/10.1016/S0364-0213(83)80009-3.

Gentner, D. (2010). Bootstrapping the mind: analogical processes and symbol systems. *Cognitive Science, 34*(5), 752–775. http://dx.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. http://dx.doi.org/10.1037/0022-0663.95.2.393.

Gentner, D., & Namy, L. L. (1999). Comparison in the development of categories. *Cognitive Development, 14*(4), 487–513. http://dx.doi.org/10.1016/S0885-2014(99)00016-7.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15*(1), 1–38. http://dx.doi.org/10.1016/0010-0285(83)90002-6.

Gick, M. L., & Paterson, K. (1992). Do contrasting examples facilitate schema acquisition and analogical transfer? *Canadian Journal of Psychology, 46*(4), 539–550. http://dx.doi.org/10.1037/h0084333.

Goldstone, R. L., Landy, D. H., & Son, J. Y. (2010). The education of perception. *Topics in Cognitive Science, 2*(2), 265–284. http://dx.doi.org/10.1111/j.1756-8765.2009.01055.x.

Hattie, J. A., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112. http://dx.doi.org/10.3102/003465430298487.

Hattikudur, S., & Alibali, M. W. (2010). Learning about the equal sign: does comparing with inequality symbols help? *Journal of Experimental Child Psychology, 107*(1), 15–30. http://dx.doi.org/10.1016/j.jecp.2010.03.004.

Hiebert, J., & Wearne, D. (1996). Instruction, understanding, and skill in multidigit addition and subtraction. *Cognition and Instruction, 14*(3), 251–283. http://dx.doi.org/10.1207/s1532690xci1403_1.

Hirsch, L. R., & Goodman, A. (2006). *Understanding elementary algebra with geometry: A course for college students* (6th ed.). Pacific Grove, CA: Brooks Cole, Thompson Learning.

Holyoak, K. J. (2005). Analogy. In K. J. Holyoak, & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 117–142). Cambridge, UK: Cambridge University Press.

Horn, W. (1983). *L-P-S Leistungsprüfsystem* (2. Auflage). Göttingen: Hogrefe.

Kamii, C., & Dominick, A. (1997). To teach or not to teach algorithms. *Journal of Mathematical Behavior, 16*(1), 51–61. http://dx.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. http://dx.doi.org/10.1002/acp.1801.

Kapur, M., & Bielaczyc, K. (2012). Designing for productive failure. *The Journal of the Learning Sciences, 21*(1), 45–83. http://dx.doi.org/10.1080/10508406.2011.591717.

Kester, L., Paas, F., & Van Merriënboer, J. J. G. (2010). Instructional control of cognitive load in the design of complex learning environments. In J. L. Plass, R. Moreno, & R. Brunken (Eds.), *Cognitive load theory* (pp. 109–130). Cambridge: Cambridge University Press.

Kieran, C. (1992). The learning and teaching of school algebra. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 390–419). New York: Macmillan Publishing Company.

Kieran, C. (2004). The core of algebra: reflections on its main activities. In K. Stacey, H. Chick, & M. Kendal (Eds.), *The future of the teaching and learning of algebra: The 12th ICMI study* (pp. 21–34). Dordrecht, The Netherlands: Kluwer.

Kirshner, D., & Awtry, T. (2004). Visual salience of algebraic transformations. *Journal for Research in Mathematics Education, 35*(4), 224–257. http://dx.doi.org/10.2307/30034809.

Kornell, N., & Bjork, R. (2008). Learning concepts and categories: is spacing the "enemy of induction"? *Psychological Science, 19*(6), 585–592. http://dx.doi.org/10.1111/j.1467-9280.2008.02127.x.

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

Loewenstein, J., & Gentner, D. (2001). Spatial mapping in preschoolers: close comparisons facilitate far mappings. *Journal of Cognition and Development, 2*(2), 189–219. http://dx.doi.org/10.1207/S15327647JCD0202_4.

Markman, A. B., & Gentner, D. (1993). Splitting the differences – a structural alignment view of similarity. *Journal of Memory and Language, 32*(4), 517–535. http://dx.doi.org/10.1006/jmla.1993.1027.

Mason, L. (2004). Fostering understanding by structural alignment as a route to analogical learning. *Instructional Science, 32*(4), 293–318. http://dx.doi.org/10.1023/B: TRUC.0000026512.88700.32.

Matz, M. (1982). Towards a process model for high school algebra errors. In D. Sleeman, & J. S. Brown (Eds.), *Intelligent tutoring systems* (pp. 25–50). London: Academic Press.

McCloskey, M., & Cohen, N. J. (1989). Catastrophic interference in connectionist networks: the sequential learning problem. *The Psychology of Learning and Motivation, 24*, 109–165.

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. http://dx.doi.org/10.1037/0278-7393.34.1.237.

NCTM. (2000). *Principles and standards for school mathematics*. Reston, VA: National Council of Teachers of Mathematics.

Ottmar, E., Landy, D., & Goldstone, R. L. (2012). *Teaching the perceptual structure of algebraic expressions: Preliminary findings from the pushing symbols intervention*. Paper presented at the Thirty-Fourth Annual Conference of the Cognitive Science Society, Sapporo, Japan.

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

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

Richland, L. E., Morrison, R. G., & Holyoak, K. J. (2006). Children's development of analogical reasoning: insights from scene analogy problems. *Journal of Experimental Child Psychology, 94*(3), 249–273. http://dx.doi.org/10.1016/j.jecp.2006.02.002.

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

Rittle-Johnson, B., Durkin, K., & Star, J. R. (2009). The importance of prior knowledge when comparing examples: influences on conceptual and procedural knowledge of equation solving. *Journal of Educational Psychology, 101*(4), 836–852. http://dx.doi.org/10.1037/a0016026.

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. http://dx.doi.org/10.1037/0022-0663.99.3.561.

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. http://dx.doi.org/10.1016/B978-0-12-387691-1.00007-7.

Rohrer, D., & Pashler, H. (2010). Recent research on human learning challenges conventional instructional strategies. *Educational Researcher, 39*(5), 406–412. http://dx.doi.org/10.3102/0013189X10374770.

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

Schneider, M., & Stern, E. (2009). The inverse relation of addition and subtraction: a knowledge integration perspective. *Mathematical Thinking and Learning, 11*(1–2), 92–101. http://dx.doi.org/10.1080/10986060802584012.

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

Schwartz, D. L., Chase, C. C., Oppezzo, M. A., & Chin, D. B. (2011). Practicing versus inventing with contrasting cases: the effects of telling first on learning and transfer. *Journal of Educational Psychology, 103*(4), 759–775. http://dx.doi.org/10.1037/a0025140.

Scott, C. (2013). The search for the key for individualised instruction. In J. A. Hattie, & E. M. Anderman (Eds.), *International guide to student achievement* (pp. 385–388). New York: Routledge.

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. http://dx.doi.org/10.1016/j.jecp.2008.11.004.

Stern, E. (1997). Early training: who, what, when, why, and how? In M. Beishuizen, K. P. Gravemeijer, & E. C. van Lieshout (Eds.), *The role of contexts and models in the development of mathematical strategies and procedures* (pp. 239–253) Utrecht: CD ß Press.

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

Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction, 12*(3), 185–233. http://dx.doi.org/10.1207/s1532690xci1203_1.

Sweller, J., Van Merrienboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251–296. http://dx.doi.org/10.1023/A:1022193728205.

Winston, P. H. (1975). Learning structural descriptions from examples. In P. H. Winston (Ed.), *The psychology of computer vision* (pp. 157–209). New York: Mc Graw-Hill.