Journal of Educational Psychology
2018, Vol. 110, No. 4, 596–610

0022-0663/18/$12.00 http://dx.doi.org/10.1037/edu0000234

When Problem-Solving Followed by Instruction Is Superior to the Traditional Tell-and-Practice Sequence

Lennart Schalk
ETH Zurich, Zurich, Switzerland, and Schwyz University of Teacher Education, Goldau, Switzerland

Ralph Schumacher, Armin Barth, and Elsbeth Stern
ETH Zurich, Zurich, Switzerland

Instruction often starts with an explanation of a concept or principle before students are presented with problems to be solved. Recent research indicates that reversing this widely used tell-and-practice sequence (T&P) so that exploratory problem-solving precedes the instructional explanation (i.e., PS-I) might be more beneficial. We aimed to replicate this advantage, but we also hypothesized based on previous research that the effectiveness of PS-I would depend on how scaffolding prompts and specific ways of representing the problems are combined. In an in vivo experimental classroom study, 213 ninth graders were randomly allocated to either a T&P or 1 of 4 PS-I conditions (in a 2 × 2 design). In all PS-I conditions, exploratory problem-solving consisted of a comparing and contrasting cases activity. However, we varied whether the students processed grounded or idealized cases (containing or stripped off contextual detail, respectively) and whether the activity was scaffolded by an invention or a self-explanation prompt. We assessed transfer performance immediately after learning and 4 weeks later. The PS-I sequences were not generally more effective than the T&P sequence, the effectiveness was influenced by an interaction of scaffolding prompts and problem representation. Immediately after learning, T&P students were only outperformed by students who learned with grounded cases and self-explanation prompts, by students who learned with grounded cases and invention prompts, and by students who learned with idealized cases and invention prompts; only the latter retained this advantage 4 weeks after learning. We discuss potential reasons and emphasize that PS-I sequences demand careful design.

Educational Impact and Implications Statement

Problem-solving followed by instruction can be a more efficient learning sequence than the conventional order in which learners are told about a concept first and solve practice problems afterward. In a classroom experiment, we investigated how two aspects of designing the problem-solving phase influenced this advantage. Students learned about the slope of a graph of a linear function by comparing and contrasting cases with different representational characteristics and different instructional prompts. Only the combination of cases that were stripped off contextual detail with the prompt to invent a general solution led to sustained advantages in transfer performance over learners who were taught following the conventional tell-and-practice sequence.

Keywords: inventing, self-explanations, compare and contrast, representational characteristics, linear functions

Presenting students with learning activities and materials that stimulate the acquisition of meaningful and transferable knowledge is among the most important professional competency requirements of teachers. However, doing so remains a challenge even for well-trained and experienced teachers. This is first and foremost the case for mathematics: Students often succeed to apply strategies and procedures to the problems they have been familiarized with during classroom practice but fail to solve structurally isomorphic but superficially different problems (e.g., Ross, 1987, 1989).

One example of a core mathematical concept of which an understanding is essential is the concept of the slope of linear functions and their graphical representation (i.e., graphs of linear functions). Linear functions and graphs are declared a key concept in mathematics curricula all over the world (e.g., Deutsche Kultusministerkonferenz, 2004; National Mathematics Advisory Panel, 2008; Schweizerische Konferenz der Erziehungsdirektoren, 2011). Understanding the slope of the graph of a linear function as rate of change facilitates the comprehension of concepts from numerous content domains; for example, for the concepts velocity, acceleration, and density in Physics, concentration in chemistry, reproduction rate in biology, and price per piece in economy. However, a deep understanding of the slope of

This article was published Online First December 21, 2017.

Lennart Schalk, ETH Zurich, Zurich, Switzerland, and Schwyz University of Teacher Education, Goldau, Switzerland; Ralph Schumacher, Armin Barth, and Elsbeth Stern, ETH Zurich, Zurich, Switzerland.

Correspondence concerning this article should be addressed to Lennart Schalk, ETH Zurich, RZ H 23, Clausiusstrasse 59, 8092 Zurich, Switzerland. E-mail: schalk@ifv.gess.ethz.ch

the graph of a linear function as the ratio of the change in the *y*-coordinates to the change in the *x*-coordinates cannot be taken for granted as a consequence of traditional classroom practice. Several studies showed that even university students were unable to interpret the slope of the graph of a linear function in unfamiliar situations (e.g., Gattis & Holyoak, 1996; Mullis, Martin, Gonzalez, & Chrostowski, 2004; Stern, Aprea, & Ebner, 2003). Whether alternatives to traditional instruction will improve an understanding of graphs of linear functions when they are first introduced is in the focus of our study. Specifically, the first goal of our study was to replicate improvements in transfer performance reported in prior research that changed the instructional sequence from a standard tell-and-practice (T&P) approach to a sequence in which exploratory problem-solving preceded an instructional explanation. The second goal of our study was to extend this research by investigating whether the beneficial effect of changing the instructional sequence depends on the design of the exploratory problem-solving phase, specifically, on how it is scaffolded and on how problems are represented.

The Potential of Preceding Explicit Instruction by Exploratory Problem-Solving

In traditional T&P instruction, an instructor would start by explaining a concept or principle (e.g., tell it to the class or provide a written instructional explanation) before presenting problems that students have to solve (i.e., practice). This sequence is typical for classrooms all over the world and across age groups (Nathan, 2012). However, several researchers have provided evidence that other sequences can produce better learning outcomes (e.g., Kapur, 2008; Schwartz & Bransford, 1998; VanLehn, 1999) because learners often lack the knowledge to properly understand a generic instructional explanation of a principle, or they may have some prior knowledge but are unable to properly activate and access it and therefore cannot relate to and elaborate the new knowledge delivered by the instructional explanation. Schwartz and Bransford (1998) reasoned that learners need a preparation before there comes the “time for telling” an instructional explanation.

In the past decade, researchers have repeatedly shown that reversing the T&P sequence by first letting students explore and try to solve problems, before they receive an explicit explanations supports learning and transfer. Schwartz, Chase, Oppezzo, and Chin (2011) created a learning condition labeled “inventing with contrasting cases,” in which students had to compare two cases exemplifying the concept of density. Learners’ comparison of the cases was scaffolded by prompts to invent a formula that would highlight the difference between both cases. Only after this exploratory problem-solving phase, learners received an instructional explanation about the concept of density and its formulaic expression. On a transfer test, the learners in the inventing with contrasting cases condition outperformed learners who were taught following the T&P sequence. Similar advantages for a reversed instructional order have been reported for learning a basic statistic concept (Kapur, 2008, 2012; Kapur & Bielaczyc, 2012). Even though students typically failed to invent a correct formula in these studies, the exploratory problem-solving activity prepared the students to benefit from an instructional explanation.

The struggle of the learners in the preparatory problem-solving activity might thus serve as a desirable difficulty (Bjork, 1994). Bjork stated that “that many of the most effective manipulations of training—in terms of post-training retention and transfer—share the property that they introduce difficulties for the learner” (p. 189). At the same time, however, Soderstrom and Bjork (2015) emphasized that “an ongoing challenge for researchers has been to identify when difficulties are desirable for learning and when they are not” (p. 193). According to the cognitive load theory (Sweller, van Merrienboer, & Paas, 1998), it is likely that difficulties are desirable when they increase the germane load in learners; that is, when they increase the mental effort devoted for processing relevant information and construction of concepts. This consideration is supported by Loibl, Roll, and Rummel (2016), who systematically reviewed studies which tested effects of learning settings in which an initial problem-solving phase was followed by an instructional phase (i.e., PS-I). Accordingly, PS-I can unfold its benefits by activating prior knowledge, making learners aware of their knowledge gaps, and highlighting deep structures of the target knowledge. Thus, PS-I sequences are challenging, but this difficulty might actually prepare learners to recognize the value of an instructional explanation and thereby benefit from them.

Design Features of PS-I Learning Environments

Scaling up scientific insights gained in lab studies to classroom practice is not always as smooth as expected by scientists (e.g., Star et al., 2015). Nevertheless, the majority of PS-I studies was run with school students in real classrooms, the contents dealt with in the studies were part of the curriculum, and the developed learning materials were ready to use. Therefore, at first sight, PS-I research seems mature for informing teacher-trainers as well as in-service-teachers when planning their classroom activities. However, we caution that there are many more decisions to make and barriers to overcome, before an idea developed by scientists will be ready for substituting traditional classroom practice. Why are we cautious?

In a solid research design, an intervention group is usually compared to a control group which has undergone a different learning activity. To trace back a potential superiority of the intervention group to the nature of its learning activity, all context and design features (e.g., time credited to the learners, design of the material, social contact with tutors) have to be kept constant in both groups. In such a design, however, potential interactions between these features and the learning activities remain uncovered. The success of the intervention group might be traced back to a positive interplay of its learning activity and context and design features. Implementing the activity under different conditions could therefore reduce or even waive its efficacy. Even worse, a negative interplay between the learning activities of the control group and context and design features might put the control group at a disadvantage; findings would then be falsely attributed to the benefits of the favored activity. Before generalizing the effects of an intervention study, one should therefore carefully consider that an instructional method can hardly be isolated from the context and design features that were held constant. Learning researchers are usually aware of the limited generalizability of findings to other content areas or age groups (e.g., Fiorella & Mayer, 2016): Learning methods that proved successful in a particular setting have to unfold their potential for various content domains and learning groups. So far, however, potential interactions between different design features of a PS-I sequence have received less attention.

Researchers have typically implemented comparing and contrasting cases activities in PS-I sequences (e.g., Kapur & Bielaczyc, 2012;

Schwartz et al., 2011). These activities can trigger structural alignment of the cases, which enables learning by schema abstraction, inference projection, difference detection, and rerepresentation (Gentner, 2010). It is, however, not sufficient to simply juxtapose cases (Alfieri, Nokes-Malach, & Schunn, 2013; Chin, Chi, & Schwartz, 2016; Renkl, 2014). In the present study, we investigated how two design features of the comparing and contrasting cases activity influence the effectiveness of a PS-I sequence. The first feature comprises the question of how to engage students in productive comparing and contrasting activities. Researchers have shown that instructional prompts are needed that trigger, scaffold, and guide the structural alignment of the cases (e.g., Roelle & Berthold, 2015; Sidney, Hattikudur, & Alibali, 2015). Two prominent prompts that we also used in the present study are to ask students (a) to write self-explanations that demand integration of the cases (Sidney et al., 2015; Williams & Lombrozo, 2013), and (b) to invent a canonical solution (e.g., an index, a formula, a principle, or a procedure) that captures commonalities and differences of the cases (e.g., Chin et al., 2016; Glogger-Frey, Fleischer, Grüny, Kappich, & Renkl, 2015; Schwartz et al., 2011). The second feature has surprisingly hitherto not received much attention from researchers who investigate comparing and contrasting cases activities. It comprises the question of how the cases themselves are represented (to foreshadow: We will distinguish between cases that are grounded in a specific context and idealized cases that strip off contextual detail). In the following, we first discuss these two design features separately, before we hypothesize a potential interaction of these features.

Two Ways to Scaffold Comparing and Contrasting Cases Activities: Self-Explanation and Invention Prompts

Asking students to generate explanations is by itself a productive learning activity that enhances learning outcomes in comparison to unguided discovery learning (Alfieri, Brooks, Aldrich, & Tenenbaum, 2011). Typically, self-explanation prompts are combined with worked examples; that is, examples or cases are presented together with their solution (Renkl, 2014). Most learners do not spontaneously produce and engage in deep self-explanations in which they monitor their learning processes, connect novel to their prior knowledge, and abstract from the examples and cases to identify an underlying principle (Atkinson, Renkl, & Merrill, 2003; Chi, Bassok, Lewis, Reimann, & Glaser, 1989). Therefore, learners should be prompted to produce principle-based self-explanations. That is, to explicitly ask learners to try to identify an underlying principle and to write down their thoughts (e.g., Rau, Aleven, & Rummel, 2015). Sidney and colleagues (2015) conducted a direct test of the necessity to scaffold comparing and contrasting activities by self-explanation prompts. In their study, undergraduates learned about fraction division. Their results indicate that merely presenting several cases simultaneously was not sufficient to enhance learning outcomes. Only if the simultaneous presentation of cases was coupled with self-explanation prompts, this promoted conceptual learning.

Another way to scaffold comparing and contrasting cases activities is to prompt learners to invent a canonical solution or a generalized description (e.g., an index, a formula, a principle, or a procedure) for the set of cases (Schwartz et al., 2011). For example, several studies (e.g., Kapur, 2012; Loibl & Rummel, 2014) used materials in which students had to compare different soccer players (i.e., cases) to determine which player was the most consistent striker. A table listed the annual amount of goals shot by three different players over several years. Students were prompted to invent an index that unequivocally indicates consistency. To solve this task, students would, in principle, have to invent the concept of variance. Students typically fail to invent it, but their exploratory problem-solving can prepare them for an explicit explanation of the concept that is provided afterward.

While self-explanation prompts are typically coupled with worked examples in which a case is presented together with its solution, it is self-evident that invention prompts can only be combined with cases that are presented without their solution. Typically, cases are developed so that only the use of the target concept or method would allow for an unequivocal description of the cases. As described above, a generalized description of consistency across all cases given (e.g., soccer players) should only be possible if students would actually invent and use the concept of variance. This was realized in the studies by Kapur (2012) and Loibl and Rummel (2014). The students knew, inter alia, how to compute the mean and the range (i.e., they had relevant prior knowledge), but the cases were developed in such a way that these two statistical values were identical for the three soccer players and therefore did not provide unequivocal evidence which player was most consistent. Only the application of the concept of variance allowed for an unequivocal description of consistency.

Prompts to invent solutions are often embedded in collaborative activities (e.g., Chin et al., 2016; Kapur, 2012), but they have also been successfully used in individual settings (e.g., Belenky & Nokes-Malach, 2012; Kapur, 2014). Few studies have directly compared the effectiveness of invention prompts in collaborative or individual learning settings; these studies do not show clear advantages for either setting (Mazziotti, Loibl, & Rummel, 2014, 2015; Sears, 2006). To foreshadow, in the present study, we used an individual learning setting.

Recently, studies revealed a more mixed picture concerning the benefits of inventing. Chin and colleagues (2016) confirmed in two studies in which students learned about projectile motion that prompting student to invent a single method led to superior learning than prompting students to list commonalities and differences of the cases. However, Glogger-Frey and colleagues (2015) found that students whose learning was scaffolded by an invention prompt performed worse in a transfer test than students who learned with worked examples. We suppose that such inconsistent findings may result from interactions with another design feature of the comparing and contrasting cases activity.

Two Ways to Represent Cases: Grounded and Idealized Problem Presentation

If an educator wants to implement a comparing and contrasting cases activity, she has to design the cases or select them from a resource. Reviewing a broad range of research, Belenky and Schalk (2014) have shown that representational characteristics of single cases, especially the amount of contextual detail, strongly influences students' performance. Belenky and Schalk distinguished between grounded and idealized cases. Grounded cases situate a principle (or a concept, a generic method, etc.) in a specific context, containing a relevant, but also sometimes irrelevant, or even seductive detail. Idealized cases, in contrast, strip out as much contextual detail as possible; that is, idealized cases represent a principle in an abstract and generic fashion. Differences between grounded and idealized cases can be subtle, but can nonetheless cause substantially different learning outcomes. Goldstone and Sakamoto (2003), for example, showed that students gained a better initial understanding if a simulation of complex adaptive systems principles was grounded (i.e., agents and objects represented in a specific context; here ants foraging for pieces of food), while students' transfer performance benefitted from an idealized simulation (i.e., agents and objects depicted as arbitrary geometrical shapes). To foreshadow, in the present study, we labeled the axes of coordinate systems with meaningful concepts (e.g., filling level in a rain barrel on the *y*-axis, and time in hours on the *x*-axis) in the grounded conditions or provided the coordinate systems without labels in the idealized conditions. That is, we rather manipulated the semantics of the cases, and not the perceptual details (note that, labeling or not labeling axes also slightly manipulates the perceptual details, but not in a strong sense as, e.g., using decorative pictures to illustrate the relation depicted in a graph). We assume that this manipulation illustrates a central decision that educators have to make when designing cases or selecting them from a resource.

Close inspection of the materials used in previous research on comparing and contrasting cases activities indeed indicates that the representational characteristics of the cases vary between studies (more about this in the next section). However, to the best of our knowledge, there are only few studies in which researchers experimentally manipulated the representational characteristics (including their similarity) in comparing and contrasting cases activities. Two studies on early language learning indicated that 18-months-olds' ability to generalize names depends on the variability of the examples encountered when learning (Perry, Samuelson, Malloy, & Schiffer, 2010), and that 3-year-olds' performance in a categorization task decreases if the exemplars contain irrelevant detail (Ankowski, Vlach, & Sandhofer, 2012). A study on algebra-like mathematical problem-solving showed that university students' learning and transfer benefitted more from juxtaposing a problem with a dissimilar than with a similar example (Lee, Betts, & Anderson, 2015). Surprisingly, even the probably most comprehensive compendium on design principles that can improve case- and example-based learning by Renkl (2014) does not encompass the feature of representational characteristics. However, from the research on the effects of representational characteristics of single problem representations or cases (as reviewed by Belenky & Schalk, 2014), one can conclude that representational characteristics do matter and influence students' learning and transfer performance. Moreover, we assume that there might be interactions between the scaffolding prompts used to prompt comparing and contrasting cases activities and the representational characteristics of cases.

Potential Interactions Between Scaffolding Prompts and Representational Characteristics of Cases

Previous research investigating self-explanation and invention prompts has used a wide variety of materials, but the representational characteristics of the cases have not been systematically varied. Given that learners have to provide a verbatim description about the cases and/or the underlying principle when writing self-explanations, this task might actually be facilitated when the prompt is coupled with grounded cases. Indeed, looking at the materials used in research on self-explanation prompts indicates that the cases are often situated in a broad context and typically provide a lot of verbatim detail (e.g., Atkinson et al., 2003; Berthold & Renkl, 2009; Rau et al., 2015). Consequently, we assume that grounding cases in a context could make it easier to write about them, to make reference to them, and to identify contextual details as superficial characteristics that vary between them and are thus irrelevant.

Prompting students to invent a canonical solution (e.g., a single formula, a generic concept) for several cases is a highly challenging task for students (Kapur & Bielaczyc, 2012). Inspecting the design of the cases in the studies in which students were prompted to invent an index for consistency reveals that this difficult task is simplified by presenting idealized cases (e.g., Kapur, 2012; Loibl & Rummel, 2014). Students' processing of the cases is aided by aligning them in a table, not much contextual information about the cases is given (only that they are soccer players). Given that the inventing task is challenging, we assume that this scaffolding prompt benefits from a coupling with cases whose design helps students to focus on the key aspects by stripping away contextual detail (e.g., without presenting distracting details about the clubs for which the players scored the goals). Contextual detail might distract learners from the key variation in the cases on which they should focus. Put differently, grounded cases might increase the extraneous load when combined with invention prompts and consequently reduce students' germane load.

The Current Study

To investigate the potential interaction between scaffolding prompts and the representational characteristics of the cases, we developed learning materials to introduce ninth graders to the concept of the slope of the graph of a linear function ($m = \frac{\Delta y}{\Delta x}$). We implemented a 2×2 factorial between-subjects design. That is, we presented either idealized or grounded cases and coupled the cases either with self-explanation or invention prompts in the learning materials. We used graphs of linear functions as cases (see Figure 1) and manipulated their representational characteristics in a simple way. In the grounded conditions, we denoted the axes of the graphs with meaningful labels; in the idealized conditions, we removed these labels. In the self-explanation conditions, we presented cases together with the respective value of the slope (i.e., a basic form of a worked example) of the linear function (exactly how it is represented in Figure 1) and prompted students to explain the slopes. In the invention conditions, we removed the values of the slopes and prompted students to invent an index that could unequivocally represent steepness.

The learning materials in these four (i.e., 2×2) conditions implemented the PS-I sequence. That is, students started with the comparing and contrasting cases activity before receiving an instructional explanation on how to compute the slope of the graph of a linear function. To evaluate whether this sequence is beneficial independent of the specific design features implemented in the different conditions, we developed an additional condition based on the T&P sequence. In the T&P sequence, students received the instructional explanation before they computed the slopes for several cases (the same number of cases as used in the experimental conditions). Overall, we thus had five experimental conditions

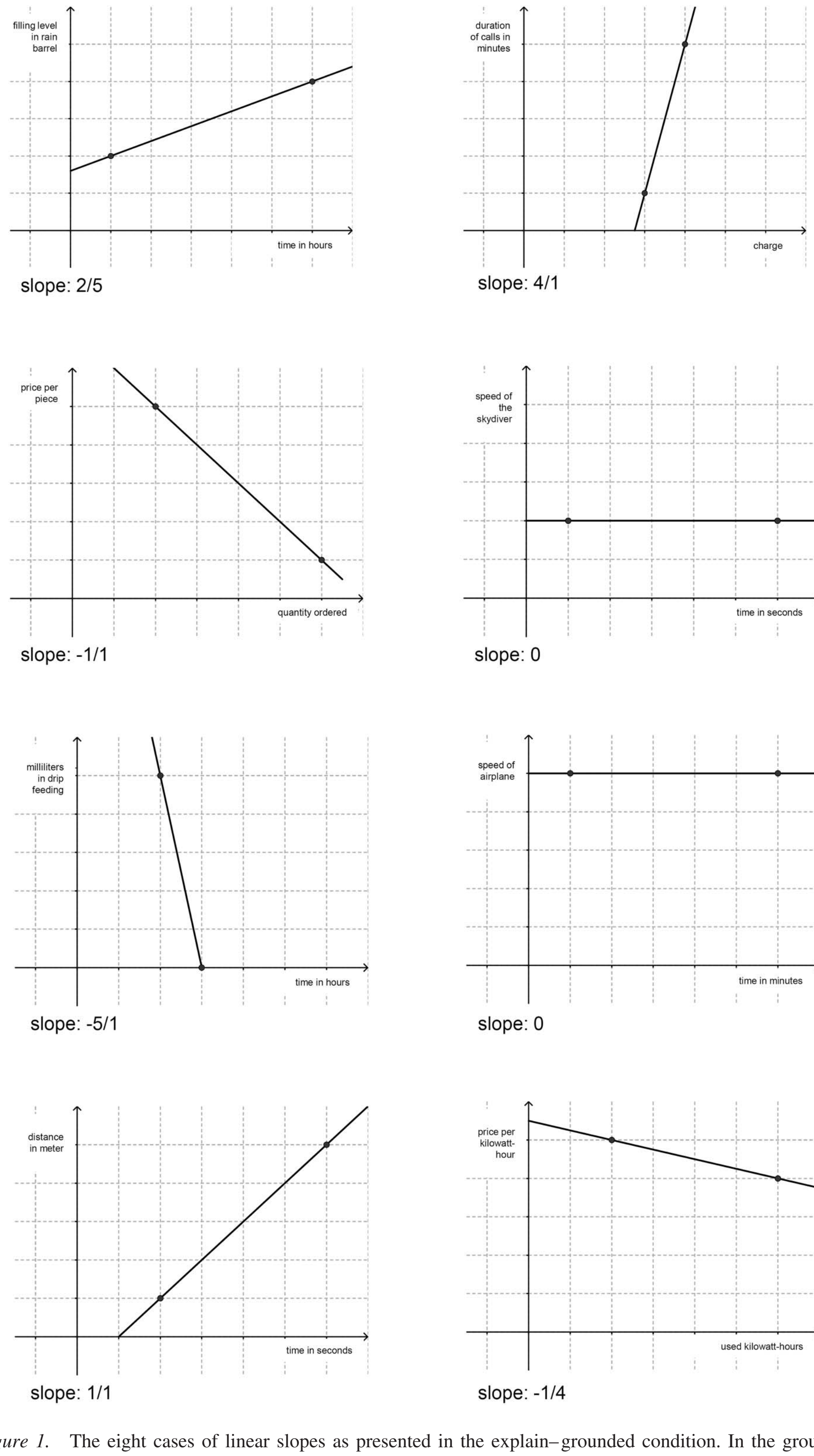


Figure 1. The eight cases of linear slopes as presented in the explain–grounded condition. In the grounded conditions, we labeled the axes of the coordinate systems to situate the slopes in a concrete context; in the idealized conditions, we removed these labels. In the explain conditions, we presented the values for the slope and learners were prompted to write explanations for how these values can be derived. In the invent conditions, we removed these values and learners were prompted to invent a method to describe the slope of the graph of a linear function.

in our study. We measured the effectiveness of the different conditions with a test on two measurement points: immediately after learning and 4 weeks later. The tasks in this test differed in context and format from the cases studied in the learning materials. Therefore, we refer to this test as a transfer test (more detail on materials and procedure is provided in the method sections).

We pursued two research questions with our study. Research Question 1 addressed whether the PS-I sequence is generally more effective than the T&P sequence, regardless of the representational characteristics and the prompts used to scaffold the comparing and contrasting cases activities. To this aim, we compared the four PS-I conditions to the T&P condition. All PS-I conditions were based on a comparing and contrasting cases activity. Furthermore, we assumed that the PS-I sequences would pose a desirable difficulty for students, as they had to try on their own (and might fail) to solve problems before receiving an instruction how to actually solve the problems. These difficulties may increase students' germane load in comparison to students who immediately receive an explanation and simply practice application of this explanation afterward (i.e., students in the T&P sequence). Given this reasoning and the previous research on the benefits of PS-I, we formulated Hypothesis 1.

> *Hypothesis 1:* At both measurement points, each of the four PS-I conditions will lead to better performance in the transfer test than the T&P condition.

Research Question 2 addressed the interaction between scaffolding prompts and representational characteristics used in the comparing and contrasting cases activity in the four PS-I conditions. On the one hand, we assumed that the invention prompt might benefit from a coupling with idealized cases rather than with grounded cases because these cases could help learners to focus on the key aspects. In the words of the cognitive load theory (Sweller et al., 1998), this kind of coupling thus increases the germane load and decreases the extraneous load (i.e., the labels of the axes are irrelevant for inventing an index for the slopes). On the other hand, we assumed that the self-explanation prompt might benefit from a coupling with grounded cases rather than with idealized cases because the concrete features of the cases (i.e., the labels) may make it easier to write about and reference the cases. Again, in the words of the cognitive load theory, this kind of coupling thus increases the germane load and decreases extraneous load (i.e., the labels help students to describe and make reference to the cases). Given these assumption, we formulated Hypothesis 2.

> *Hypothesis 2:* For both measurement points, the 2×2 analysis of variance with the factor prompts (Self-Explanation, Invention) and representational characteristics (grounded, idealized) will reveal a disordinal interaction. Specifically, the self-explanation prompt/grounded cases condition will lead to superior transfer performance than the self-explanation prompt/idealized cases condition, while the invention prompt/idealized cases condition will be superior to the invention prompt/grounded cases condition.

We tested these two hypotheses in an in vivo experimental classroom study. That is, we randomly distributed the different learning materials among the students in several ninth-grade mathematics classrooms.

Method

Participants

Overall, 213 students from a medium-sized Swiss city participated. We conducted an in vivo study and tested in 10 classrooms. Data of 24 students (11%) could not be used because they had severe language problems or missed the second test session. Thus, analyses are based on 189 students (94 girls) with a mean age of 15.6 years. The learning materials of all conditions were randomly distributed within a classroom (i.e., in all classroom, all conditions were implemented). The institutional ethics review board approved the study.

Materials

Learning materials. In all five conditions, students received the same instructional explanation. It was printed on a single page and provided a generic description of the mathematical formula for how to compute the slope of the graph of a linear function together with a graphical representation in a coordinate system (see Figure 2).

In all four PS-I (henceforth, $\text{invent}_{\text{grounded}}$, $\text{invent}_{\text{idealized}}$, $\text{explain}_{\text{grounded}}$, $\text{explain}_{\text{idealized}}$) and the T&P conditions, students learned with eight cases of linear functions. The cases comprised three graphs with a positive slope, three graphs with a negative slope, and two graphs with a slope of 0 (see Figure 1).

In the PS-I conditions, students first received a booklet that juxtaposed the eight cases on a single page. They received the instructional explanation only after they worked through this booklet. In the grounded conditions, we presented the coordinate systems used for the cases with meaningful labels. In the idealized conditions, we removed these labels.

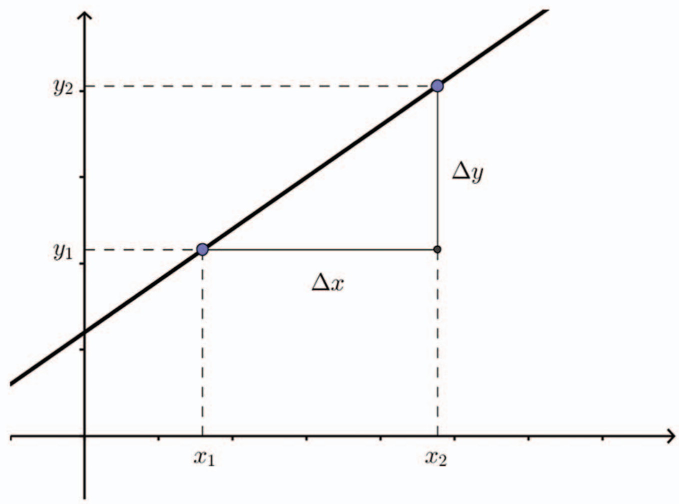
In the explain conditions, we presented the cases together with the value for the slopes. We prompted self-explanations with four questions. These questions aimed at focusing students' attention to identify how the value of the slope could be derived and computed. The first question prompted students to draw a graphical aid that illustrates how one could compute the slope and to explain their graphical aid. The second, third, and fourth questions prompted students to explain under which conditions the value of the slope becomes positive, negative, or zero, respectively.

In the invent conditions, we presented the cases without the value for the slopes. We prompted students to invent a canonical solution for how to compute the slopes that works for all cases. To this aim, we asked students to imagine that they were calling a friend who cannot see the graphs, but wants to know about them. Thus, students needed to develop a method to ensure that their friend understands exactly how the cases look like (with regard to the slope only; we did not introduce the intercept in any learning material).

In contrast to the PS-I conditions, T&P learners first read the instructional explanation. Afterward, they received a booklet presenting the eight cases one after another (four grounded and four idealized, intermixed). The cases were presented without the value of the slope (as in the invent conditions) and we asked students to compute the slopes for each case. We reminded them that they can use and apply the formula they learned about in the instructional explanation.

Graphs of linear functions are represented by straight lines in coordinate systems. A straight line can inter alia be characterized by a value that represents the "steepness" of the line.

The technical term for the steepness is **slope**. The slope of a graph of a linear function in a coordinate system is determined as follows:



The **slope** m of a straight line that connects two points (x_1, y_1) and (x_2, y_2) with $x_1 \neq x_2$ is thus defined as the following relation:

$$\boxed{m := \frac{\Delta y}{\Delta x} = \frac{y_2 - y_1}{x_2 - x_1}}$$

Straight lines that ascend from left to right have a positive slope, while straight lines that descend from left to right have a negative slope. Straight lines that are parallel to the x-axis have the slope 0 because $\Delta y = 0$ independently of which points are chosen. Straight lines that are parallel to the y-axis have no defined slope because for these lines $\Delta x = 0$ and it is not possible to divide by 0.

Figure 2. The instructional explanation (translated from the German version used in experiment) of how to compute the slope of the graph of a linear function. See the online article for the color version of this figure.

Thus, all learning materials demanded active work of the students. In the T&P condition, students had to compute solutions for the eight cases. In the explain conditions, students had to provide four explanations about the slopes of the cases. In the invent conditions, students had to develop their own solution for how to compute the slopes. In additional explorative analyses (see Results section), we evaluated how students' quality of work during learning was related to their performance on the transfer test.

Tests. We developed a test comprising 20 tasks. Twelve of the tasks required calculating solutions to given problems, the other eight tasks required to provide conceptual descriptions (see Table 1 for exemplary tasks). Tasks were situated in contexts that differed from the contexts studied in the learning materials, but we informed students that the learning materials had provided them with all information necessary to solve these tasks. Therefore, we refer to this test as a cued transfer test. Using Barnett and Ceci's (2002) taxonomy for transfer tasks, the contextual distance of the tasks to the learning materials can be described as follows. Transfer distance is near for the physical context (learning and transfer in the same room at school), functional context (learning and transfer are both clearly academic), social context (student work individually while learning and while solving the transfer test), and modality (both learning and transfer require processing written materials in similar format). Transfer distance is medium for knowledge domain (tasks of the transfer test require to solve problems situated in various context that have not been used in the learning materials) and temporal context (transfer test is conducted immediately after learning and with a delay of 4 weeks).

Table 1
Two Exemplary Tasks From the Transfer Test (Translated From the German Version Used in the Experiment)

Task	Description
A	Someone empties a bathtub; the drain of water is constant. You have a graphical representation that depicts the amount of water (in L) left in the bathtub contingent on time (in min). This representation shows a straight line with the slope −3/10, which goes through the point (2, 145); that is, after 2 min, there are still 145 L of water left in the bathtub. Please calculate and check whether it is correct that after 11 min, there is exactly 25 L of water left!
B	A straight line goes through the point (x, y). Is it right or wrong to say that the slope equals y/x? Give an explanation!

Note. Task A represents an example of the tasks that required calculating a solution; Task B represents an example of the tasks that required to provide a conceptual description.

In the test, every task was scored with one point if solved completely correct, with a half point if a partially correct answer was provided, and with 0 points if the answer was incorrect or missing (i.e., maximum score of 20). Two independent coders scored 20% of the tests (including a mathematics teacher). Interrater agreement was high (intraclass correlation coefficient = .88). Inconsistencies were resolved by discussion and one of the coders scored the remaining tests. The internal consistency across the 20 tasks was high as indicated by Cronbach's α = .79 for the immediate, and Cronbach's α = .81 for the delayed transfer test. Even though the test was conceived of as a comprehensive measure of students' ability to transfer their knowledge about the slope of the graph of a linear function, we computed additional explorative analyses to assess whether potential differences in transfer performance were more strongly influenced by performance on items that required calculating a solution or by performance on items that required conceptual descriptions (see Results section).

Demographic and motivational questionnaires. We created a demographic questionnaire to assess students' age, sex, and their math and German grades. Additionally, we assessed performance and mastery goal orientation with the standard scales (Elliot & Murayama, 2008) translated into German. The information obtained with the demographic questionnaire and the goal orientation scales served to check whether the random distribution of the materials within the classrooms resulted in experimental groups with a similar initial performance level. This check was necessary because using a pretest would have been highly questionable as it could be regarded as an invention intervention for all conditions and thus would have confounded the experimental variation in the present study. We computed additional explorative analyses to evaluate whether performance and mastery orientation influenced transfer performance of the students (see Results section).

Procedure

We tested directly after the summer break at the beginning of Grade 9. The Swiss curriculum for eighth grade requires students to learn about coordinate systems and how to depict and describe single points in these systems; linear functions are then to be introduced in ninth grade. Thus, testing after the summer break ensured that students had relevant prior knowledge that is necessary to learn about the slope of the graph of linear function, but did not yet receive instruction on this concept.

We tested students in classrooms in their standard math lessons. Testing comprised two sessions. We fully informed the teachers of the classes about the aims of the study. They agreed to not answer any questions of their students regarding the learning materials or to mention graphs of linear functions and the concept of slopes in their teaching until after the second session. This ensured that students did not receive additional instruction independent from the learning materials developed for this study. In all classrooms, we were able to spatially separate the students to prevent copying from classmates.

The first session lasted 100 min comprising two 45-min lessons (45 min is the typical duration of a lesson in Switzerland) and a 10-min break. In the first lesson, we administered the demographic and the motivational questionnaire first (5 min). Afterward, we randomly distributed the different learning materials among the students. We told the students that they would be working with different materials. These materials were color-coded to help the experimenter (all testing was conducted by the first author) communicate about the different materials. Students were told that depending on the color of their learning materials, they will get new materials after a specific time. This procedure allowed to realize all experimental conditions simultaneously in one classroom and to equate the time on task across conditions.

Specifically, students in the T&P condition started with the instructional explanation. They had 5 min to read it. Afterward, the experimenter removed the explanation and handed over the booklets with the cases on which the students worked until the end of the first lesson (i.e., 30 min). In the PS-I conditions, the experimenter gave the students the booklets with the cases first. They had 30 min to work on these materials. After 30 min, the experimenter removed the booklets and handed over the instructional explanation. Like in the T&P condition, students had 5 min to read this explanation. If one of the students in any condition finished earlier, we asked her or him to go over the materials again, imagining to prepare for an exam.

After this learning phase, a 10-min break followed. In the second lesson of the first session, we administered the transfer test. Students had 45 min to answer the questions of the transfer test.

In the second session 4 weeks later, we administered the transfer test again. Afterward, we debriefed the students. In the first session, we insured the students that data collection is completely anonymous (by using a specific coding system) and that their teachers will not get any information about their performance. We motivated students' participation by telling them that the results of the experiment will be helpful in improving the design of educational learning materials.

Results

To ensure that the random distribution of learning materials within classes resulted in comparable groups, we first compared student's math and German grades (grades range from 1 to 6 in Switzerland, with 6 representing the best grade) as well as their performance and mastery goal orientation (see Table 2 for means and standard deviations). There were no statistically significant differences regarding Math grades (six students did not provide their math grades; therefore, this analysis includes $n = 183$), $F(4, 179) = 0.308$, $p = .872$, $\eta_p^2 = .007$; German grades (seven students did not provide their German grades; therefore, this analysis includes $n = 182$), $F(4, 178) = 0.181$, $p = .948$, $\eta_p^2 = .004$; performance goal orientation, $F(4, 184) = 0.520$, $p = .721$, $\eta_p^2 = .011$; and mastery goal orientation, $F(4, 184) = 0.454$, $p = .769$, $\eta_p^2 = .010$.

Next, we investigated Hypotheses 1. We hypothesized that all PS-I conditions would outperform the T&P condition on both measurement points. We compared the immediate and delayed transfer test performance (see Table 2 and Figure 3) of the four PS-I (invent$_{\text{grounded}}$, invent$_{\text{idealized}}$, explain$_{\text{grounded}}$, and explain$_{\text{idealized}}$) and the T&P conditions in a 2 (Time: repeated assessment of the transfer test) × 5 (Condition) repeated-measures analysis of variance (ANOVA). The interaction of Time × Condition, $F(4, 184) = 2.806$, $p = .027$, $\eta_p^2 = .057$, as well as the effect of condition, $F(4, 184) = 3.225$, $p = .014$, $\eta_p^2 = .066$, were statistically significant, while the effect of time was not statistically significant, $F(4, 184) = .017$, $p = .895$, $\eta_p^2 = .000$.

The interaction indicates that the differences between conditions changed across the two measurement points. To unpack the interaction and to determine which of the four PS-I conditions outperformed T&P, we computed preplanned contrasts derived from Hypothesis 1 between conditions separately for the immediate and the delayed transfer assessment.

For the immediate transfer test, we found that not all PS-I conditions significantly outperformed T&P. Specifically, one-sided tests (justified by the directional Hypothesis 1) showed significant advantages over T&P for invent$_{\text{grounded}}$ ($p = .045$, $d = .399$), invent$_{\text{idealized}}$ ($p = .001$, $d = .768$), and explain$_{\text{grounded}}$ ($p = .023$, $d = .475$), but not for explain$_{\text{idealized}}$ ($p = .405$, $d = .119$). The only other significant difference between conditions in the immediate transfer test indicated that invent$_{\text{idealized}}$ outperformed explain$_{\text{idealized}}$ ($p = .007$, $d = .812$). For the delayed transfer test, we found significant advantages over T&P only for invent$_{\text{idealized}}$ ($p = .017$, $d = .573$), but not for invent$_{\text{grounded}}$ ($p = .373$, $d = -.073$), explain$_{\text{grounded}}$ ($p = .470$, $d = .018$), and explain$_{\text{idealized}}$ ($p = .337$, $d = 0.101$). Actually, invent$_{\text{idealized}}$ did not only outperform the T&P condition, but also the three other PS-I conditions (invent$_{\text{grounded}}$: $p = .012$, $d = .598$; explain$_{\text{grounded}}$: $p = .034$, $d = .511$, explain$_{\text{idealized}}$: $p = .010$, $d = .642$). The performance on the transfer test was stable over time within conditions (all correlations between the immediate and delayed transfer test performance ($.612 < r\text{s} < .772$ with $p\text{s} < .001$) and the stability (i.e., the correlations) did not differ between conditions (all Fisher's Z values <1.305).

To test Hypothesis 2 according to which the different representational characteristics of cases (grounded or idealized) interact with the different kind of prompts (self-explanation or invention), we computed a 2 (Time: repeated assessment of the transfer performance) × 2 (Representational Characteristics) × 2 (Kind of Prompt) repeated-measures ANOVA. This analysis showed no significant effects of the three-way Interaction Time × Representational Characteristics × Kind of Prompt, $F(1, 150) = .181$, $p = .671$, $\eta_p^2 = .001$, of the two-way Interaction Time × Kind of Prompt, $F(1, 150) = .412$, $p = .522$, $\eta_p^2 = .003$, and of the main effects time, $F(1, 150) = 1.010$, $p = .317$, $\eta_p^2 = .007$, and representational characteristics, $F(1, 150) = .219$, $p = .640$, $\eta_p^2 = .001$. The two-way interactions of Time × Representational Characteristics, $F(1, 150) = 6.821$, $p = .010$, $\eta_p^2 = .043$, and Representational Characteristics × Kind of Prompt, $F(1, 150) = 6.526$, $p = .012$, $\eta_p^2 = .042$, as well as the main effect of kind of prompt, $F(1, 150) = 4.028$, $p = .047$, $\eta_p^2 = .026$, were statistically significant.

The two-way interaction of time and representational characteristics emerged because the performance of learners who studied grounded cases dropped slightly over time (from $M_{\text{Immediate}} = 12.090$ to $M_{\text{Delayed}} = 11.231$) while the performance of learners who studied idealized cases slightly increased over time (from $M_{\text{Immediate}} = 11.733$ to $M_{\text{Delayed}} = 12.114$). The two-way inter-

Table 2
Number of Participants and Means (Standard Deviations) per Condition for All Questionnaires and Tests

Number of participants, questionnaires, and tests	T&P	Invent: Grounded	Invent: Idealized	Explain: Grounded	Explain: Idealized
n	35	39	39	39	37
Math grade	4.84 (.56)	4.98 (.63)	4.94 (.64)	4.85 (.80)	4.91 (.54)
German grade	4.77 (.41)	4.78 (.41)	4.85 (.42)	4.83 (.45)	4.81 (.49)
Performance orientation (max. 21)	15.57 (3.21)	14.94 (3.80)	15.69 (3.43)	15.77 (3.43)	16.00 (3.24)
Mastery orientation (max. 21)	17.94 (2.46)	18.00 (2.13)	18.33 (1.40)	17.72 (2.45)	17.81 (2.31)
Immediate transfer					
Total (max. 20)	10.53 (3.29)	11.96 (3.86)	13.14 (3.50)	12.22 (3.81)	10.32 (3.44)
Conceptual description items (max. 8)	3.54 (1.42)	4.24 (1.78)	4.31 (1.81)	4.19 (1.59)	3.57 (1.56)
Calculation items (max. 12)	6.99 (2.46)	7.72 (2.74)	8.83 (2.21)	8.03 (2.71)	6.76 (2.70)
Delayed transfer					
Total (max. 20)	11.34 (3.63)	11.05 (4.25)	13.27 (3.08)	11.41 (4.12)	10.96 (3.91)
Conceptual description items (max. 8)	3.67 (1.51)	3.76 (1.64)	4.51 (1.71)	4.01 (1.62)	4.05 (1.62)
Calculation items (max. 12)	6.84 (2.49)	6.79 (3.07)	8.01 (2.26)	6.69 (3.15)	6.28 (2.87)

Note. max. = maximum; T&P = tell-and-practice condition; invent = invention conditions; explain = self-explanation conditions; grounded = conditions in which graphs were situated in a specific context by labeling the axes; idealized = conditions in which graphs were presented without labeled axes.

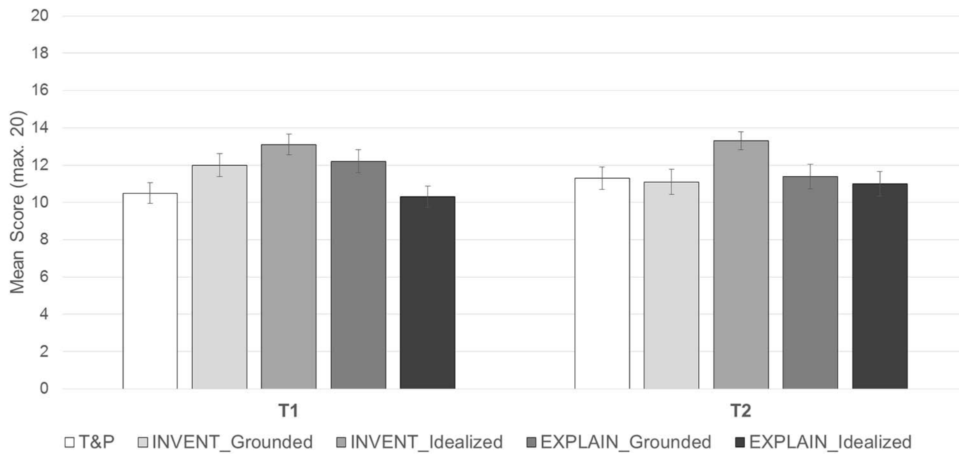


Figure 3. Results of the transfer test. Transfer was assessed immediately after learning (T1) and 4 weeks later (T2). T&P = tell-and-practice condition; INVENT = invention conditions; EXPLAIN = self-explanation conditions; grounded = conditions in which graphs were situated in a specific context by labeling the axes; idealized = conditions in which graphs were presented without labeled axes. Error bars represent standard errors of the means.

action of Representational Characteristics × Kind of Prompt qualifies the main effect of kind of prompt. Namely, learners in the two invent conditions showed a better transfer performance than learners in the two explain conditions, but the interaction additionally indicates that learners in the invent conditions showed a better performance when they learned with idealized ($M_{\text{Idealized}}$ = 13.205) than with grounded cases (M_{Grounded} = 11.506) while learners in the explain conditions showed an inferior performance when they learned with idealized ($M_{\text{Idealized}}$ = 10.642) than with grounded cases (M_{Grounded} = 11.814). This pattern is exactly the disordinal interaction, we formulated in Hypothesis 2.

While the significant representational characteristics × kind of prompt interaction and the visual pattern in Figure 3 match our second hypothesis, it is also obvious from inspecting Figure 3 that the pattern is not particularly pronounced, especially in the delayed transfer performance. Preplanned comparisons between conditions for both transfer assessments support this impression. We predicted that the invent prompt would lead to better performance if coupled with idealized compared to when coupled with grounded cases; and that the explain prompt would lead to better performance if coupled with grounded compared to when coupled with idealized cases. For the invent prompt, the data supported our prediction only at the delayed, $t(76) = 2.640$, $p = .010$, $d = .598$, but not at the immediate transfer assessment, $t(76) = 1.412$, $p = .162$, $d = .320$. For the explain prompt, the data supported our prediction only at the immediate, $t(74) = 2.266$, $p = .026$, $d = .523$, but not at the delayed transfer assessment, $t(74) = .481$, $p = .632$, $d = .110$. Furthermore, it did not make a statistically detectable difference whether grounded cases were coupled with invent or explain prompts, immediate transfer: $t(76) = .295$, $p = .769$, $d = .068$; delayed transfer: $t(76) = .379$, $p = .706$, $d = .086$; but learners showed a better transfer performance when idealized cases were coupled with invent than with explain prompts, immediate transfer: $t(74) = 3.534$, $p = .001$, $d = .813$; delayed transfer: $t(74) = 2.809$, $p = .006$, $d = .642$.

Taken together, the data provides partial support for our hypotheses: problem-solving before instruction can lead to significant and stable transfer performance advantages over T&P, but this advantage depends on the design of the problem-solving activity. Before discussing our results in more detail, we report additional explorative analyses that might illuminate potential mechanisms underlying our findings.

Additional Explorative Analyses

We explored with additional analyses (a) whether the two motivational aspects that we measured, that is, performance and mastery orientation, did have a differential effect varying across conditions for the two transfer assessments; (b) whether students' activity during learning correlated with their transfer performance; and (c) whether differences in transfer performance were mainly driven by the tasks that required calculating solutions to given problems or by the tasks that required to provide conceptual descriptions.

Motivational aspects. To explore how the motivational aspects influenced transfer performance, we recomputed the 2 (Time: repeated assessment of the transfer test) × 5 (Condition) repeated-measures ANOVAs reported above, but now additionally included the scores for performance and mastery orientation as covariates. This analysis did not show significant main effects of either motivational aspect nor interactions of these aspects with the other factors (all $ps > .05$) while the effect of condition remained for the marginal estimated means (i.e., the transfer performance controlled for motivation), $F(4, 174) = 3.615$, $p = .007$, $\eta_p^2 = .077$. Thus, performance and mastery orientation did not play an important role in determining the outcome of our study.

Students' activity during learning. In the introduction, we speculated that PS-I sequences might serve as a desirable difficulty and thus support learning in comparison to T&P which might be less challenging. To shed some light on this speculation, we analyzed how learners performed during working with the learning materials. That is, learners had to either solve practice problems (T&P), invent a solution (invent$_{\text{grounded}}$; invent$_{\text{idealized}}$), or write explanations (explain$_{\text{grounded}}$; explain$_{\text{idealized}}$). For T&P, we coded whether students solved the practice problems correctly or not. For the invent conditions, we used a coding scheme with five categories to capture the quality of the invented solutions: (a) missing (i.e., student did not invent any solution), (b) wrong–incoherent (i.e., student invented several solutions for the different cases and all of these solutions are wrong), (c) wrong–coherent (i.e., student invented a general solution for all cases, but it does not work), (d) correct–incoherent (i.e., student invented several solutions for subgroups of cases that do only work for these subgroups), and (e) correct–coherent (i.e., student invented a general solution for all cases that works). For the explain conditions, we used a coding scheme with four categories: (a) missing (i.e., student did not produce any explanation), (b) incorrect (i.e., student provided an explanation that does not work at all), (c) partially wrong (i.e., student provided an explanation that works for some cases, but not as a general solution), and (d) correct (i.e., student provided an explanation that works generally). Two independent raters (student coworkers who were blind to hypotheses) coded the answers of 10 randomly selected participants per condition. Interrater agreement was 100% for T&P, 95% for the invent conditions, and 86.25% for the explain conditions. Disagreements were solved by discussion and one of the raters coded all remaining answers of participants.

The solution rate in the T&P condition for the eight practice problems was high (M = 6.914; SD = 1.269). The solution rate was not statistically related to the performance on the immediate, Spearman's $\rho(35) = -.013$, $p = .942$, and delayed transfer assessments, Spearman's $\rho(35) = .010$, $p = .955$.

Also, students were quite successful in inventing a solution; most students came up with a solution that was scored as either (4) correct—incoherent or (5) correct coherent (invent$_{\text{grounded}}$: M = 4.462, SD = 1.047; invent$_{\text{idealized}}$: M = 4.436, SD = 0.912). The quality of the invented solutions did not differ between both conditions, $t(76) = .115$, $p = .909$, $d = .026$. In the invent$_{\text{grounded}}$ condition, the quality of the invented solution showed a positive statistically significant correlation with performance on the immediate, Spearman's $\rho(39) = .421$, $p = .008$, and the delayed transfer assessment, Spearman's $\rho(39) = .485$, $p = .002$. In contrast, in the invent$_{\text{idealized}}$ condition, the quality of the invented solution did not show a statistically significant correlation with performance on the immediate, Spearman's $\rho(39) = .294$, $p = .070$, and the delayed transfer assessment, Spearman's $\rho(39) = .197$, $p = .228$.

Students had to write four explanation in the explain conditions; thus, they could obtain a maximum score of 16. The explanation of students were of rather high quality (explain$_{\text{grounded}}$: M = 13.846, SD = 1.615; explain$_{\text{idealized}}$: M = 13.351, SD = 1.975). The quality of explanations did not differ between both conditions, $t(74) = 1.198$, $p = .235$, $d = .274$. In the explain$_{\text{grounded}}$ condition, the quality of explanations was not correlated with performance on the immediate, Spearman's $\rho(39) = .187$, $p = .255$, but with performance on the delayed transfer assessment, Spearman's $\rho(39) = .442$, $p = .005$. In the explain$_{\text{idealized}}$ condition, the quality of the explanations was also not correlated with performance on the immediate, Spearman's $\rho(37) = .073$, $p = .666$, but with performance on the delayed transfer assessment, Spearman's $\rho(37) = .327$, $p = .048$. We provide a tentative interpretation of these explorative analyses for students' activity during learning in the discussion.

Kinds of transfer items. We explored whether the performance differences between conditions were more strongly affected by the two kind of items of the transfer test. That is, some items required to write conceptual descriptions, others required to calculate solutions. To this aim, we recomputed the 2 (Time: repeated assessment of the transfer test) × 5 (Condition) repeated-measures ANOVA for the transfer test twice: once for the conceptual items and once for the calculation items (see Table 2 for means). Note that, splitting up the test, reduced the internal consistency of the scales (for calculation items in the immediate transfer test, Cronbach's $\alpha = .627$; for the delayed transfer test, $\alpha = .627$; for conceptual items in the immediate transfer test, $\alpha = .747$; for the delayed transfer test, $\alpha = .788$).

The ANOVA for conceptual items showed no effects of time, $F(1, 184) = .098$, $p = .755$, $\eta_p^2 = .001$, and condition, $F(4, 184) = 1.576$, $p = .183$, $\eta_p^2 = .033$, but a significant Time × Condition interaction, $F(4, 184) = 2.947$, $p = .022$, $\eta_p^2 = .060$. Therefore, we computed two ANOVAs for each measurement point. For the immediate transfer test, the ANOVA showed a significant effect of condition, $F(4, 184) = 3.970$, $p = .004$, $\eta_p^2 = .079$. For the delayed transfer test, the ANOVA showed no significant effect of condition, $F(4, 184) = 2.061$, $p = .088$, $\eta_p^2 = .043$. The ANOVA for procedural items showed significant effects of time, $F(1, 184) = 20.750$, $p < .001$, $\eta_p^2 = .101$, and condition, $F(4, 184) = 3.229$, $p = .014$, $\eta_p^2 = .066$, but the Time × Condition interaction was not significant, $F(4, 184) = 1.531$, $p = .195$, $\eta_p^2 = .032$. For the immediate transfer test, the ANOVA showed a significant effect of condition, $F(4, 184) = 3.970$, $p = .004$, $\eta_p^2 = .079$. For the delayed transfer test, the ANOVA showed no significant effect of condition, $F(4, 184) = 2.061$, $p = .088$, $\eta_p^2 = .043$. The pattern of means both for conceptual and procedural items matches the pattern for the total transfer score (see Table 2). Taken together, these explorative analyses for the transfer item kinds converge with the analyses for the overall transfer performance. That is, the difference became smaller over time while the differences seem not to be specifically driven by one of the two kinds of transfer items.

Discussion

For an in vivo experimental study run in ninth-grade mathematics classrooms, we developed learning materials that introduced a central concept in linear algebra (namely, the concept of the slope of the graph of linear functions) to investigate specific implementations of the PS-I sequence. The goal of the study was twofold: First, we wanted to conceptually replicate the superiority of PS-I learning to the traditional T&P approach that has been demonstrated in various previous studies (for an overview see Loibl et al., 2016). Second, we hypothesized that the success of the PS-I sequence might depend on the interaction of the two design features: (a) "scaffolding prompts" that are used to engage learners in productive comparing and contrasting cases activities and (b) the "representational characteristics" of the cases. To this aim, we scaffolded processing of the cases with either self-explanation or invention prompts and varied the representational characteristics by using either grounded or idealized cases in the PS-I learning materials. With this 2 × 2 design, we tested whether the effectiveness of PS-I learning materials depended on the combination of these two design features.

How Design Features of PS-I Learning Environments Took Effect

Our first hypothesis, according to which the PS-I sequence would generally benefit transfer performance in comparison to the T&P sequence, did not receive full empirical support. Not all four PS-I conditions performed significantly better than the T&P condition, in which students received the instructional explanation before they solved practice problems. Immediately after learning, only participants who studied grounded cases coupled with self-explanation prompts (explain$_{\text{grounded}}$), participants who studied grounded cases coupled with invention prompts (invent$_{\text{grounded}}$), and participants who studied idealized cases coupled with invention prompts (invent$_{\text{idealized}}$) outperformed T&P learners, but not participants who studied idealized cases coupled with explanation prompts (explain$_{\text{idealized}}$). In the repeated assessment of the transfer performance 4 weeks after learning, even only the learners who learned with the PS-I materials in which idealized cases were coupled with invention prompts (invent$_{\text{idealized}}$) showed a statistically significant advantage over T&P learners. This result suggests that there is no general superiority of PS-I compared to traditional T&P instruction. All of our PS-I conditions required learners to compare and contrast cases. If students successfully engage in structural alignment of the cases (Gentner, 2010), this typically results in superior learning outcomes in comparison to other forms of case study (Alfieri et al., 2013). However, integrating these comparing and contrasting cases activities in real classrooms has been shown be challenging for mathematics teachers (Star et al., 2015). Complementing these findings, our results indicate that it matters *how* a comparing and contrasting cases activity is designed when it is implemented in real classroom instruction.

Accordingly, we found the interaction pattern formulated in Hypothesis 2 in the 2×2 design of the PS-I learning materials for both times of transfer assessment. However, this interaction was not particularly pronounced especially in the delayed transfer performance. We predicted that grounded cases would lead to a better transfer performance if coupled with self-explanation prompts while idealized cases would lead to a better transfer performance if coupled with invention prompts. When learning was scaffolded with self-explanation prompts, we found the predicted difference only in the immediate transfer performance, but not in the delayed transfer performance. When learning was scaffolded with invention prompts, we found the predicted difference only in the delayed but not in the immediate transfer performance. Furthermore, whether grounded cases were coupled with invention or self-explanation prompts did not statistically matter. However, it mattered whether idealized cases were coupled with invention or self-explanation prompts: The combination of idealized cases and the invention prompt led to a superior transfer performance. Additionally, students who learned with idealized cases showed a slight increase in transfer performance over time while students who learned with grounded cases showed a slight performance decrease over time. Hence, the interaction was more complicated than we expected. Nevertheless, the results clearly highlight that it is important to carefully design comparing and contrasting cases activities in PS-I sequences.

The present results thus complement and extend recent research on the benefits of comparing and contrasting cases. Sidney and colleagues (2015) showed that juxtaposing cases is not sufficient to trigger structural alignment in which learners identify similarities and differences and abstract from the cases to construct generalizable knowledge representation. Self-explanation prompts were necessary to scaffold learners' processing of simultaneously presented cases. Our findings indicate that the representational characteristics of the cases also matter. Only if grounded cases were coupled with self-explanation prompts, a significant advantage performance in comparison to the control condition was detected and this advantage was only present directly after learning. Four weeks later, this advantage disappeared. We assume that the self-explanation prompts initially helped learners to abstract from the context provided in the grounded cases, but that the contextual details were encoded anyway. Over time, these details seem to have become an integral part of the knowledge representation encoded from the learning materials, and thus hamper transfer performance in the long run. This assumption is backed up by findings that learners encode knowledge representations that preserve the format and the details of presentation (e.g., De Bock, Deprez, Van Dooren, Roelens, & Verschaffel, 2011; Schalk, Saalbach, & Stern, 2016).

Coupling a prompt to invent a canonical solution with idealized cases was the only PS-I sequence that led to stable advantages over the T&P sequence. The prompt to invent a canonical solution that demands to abstractly and coherently describe a set of cases is a highly challenging task. Students often fail to find the right solution, but this failure can prepare them for an instructional explanation (e.g., Kapur, 2008). At first glance, our results might seem to be in conflict with findings by Glogger-Frey and colleagues (2015). They showed that learners benefitted more from learning with worked examples than from inventing. However, they also showed that learners in their inventing condition recalled more surface features (Experiment 2). Their cases actually contained various surface features: colored and detailed pictures of clowns organized in compartments of busses represented density. Thus, according to our distinction between grounded and idealized cases, the cases used by Glogger-Frey and colleagues would be considered grounded cases. Thus, our results rather refine the findings of Glogger-Frey and colleagues by indicating that the efficiency of prompts depends on the representational characteristics of the cases.

In the introduction and when deriving our hypotheses, we speculated that PS-I sequences might pose a desirable difficulty (Bjork, 1994) for learners by increasing the germane load (Sweller et al., 1998) in comparison to the T&P sequence and consequently enhance transfer performance. Despite its plausibility, this generic claim is not entirely supported by our data. Explorative analyses in which we correlated students' activity during learning with their transfer performance allow speculating about mechanisms underlying our findings but do not reveal a fully consistent picture.

In general, students' activity was of good quality across conditions. In the T&P condition, the mean performance in solving the practice problems was high. But, the solution rate of practice problems was not correlated with transfer performance at both measurement points. Given that T&P did result in inferior transfer performance in comparison to some of the PS-I conditions, one could argue and reinstate a widely held practical and theoretical assumption (for an overview see Nokes-Malach & Mestre, 2013) that simply computing the solutions to eight similar problems is not the best instructional technique to support students' transfer performance. In both explain conditions, the quality of self-explanation did not correlate with immediate, but with the delayed transfer performance. Thus, their constructive activity when formulating the self-explanations did not matter much initially (here grounded cases helped students in their performance on the transfer test). But, 4 weeks later when the delayed transfer test was administered, this activity mattered. Students might have forgotten the direct instruction (i.e., simply reading about how to compute the slope of linear functions) over the 4 weeks, but seemed to still remember their self-explanation activity—the ones who provided higher quality explanations now also performed better (while the representational characteristics of the cases did not substantially influence performance any more). In the invent conditions, the quality of invented solutions correlated with transfer performance on both measurement points when learners worked with grounded cases, but not when they worked with idealized cases. Remember that the invent$_{\text{grounded}}$ learners were inferior in transfer performance compared to the invent$_{\text{idealized}}$ learners, and invent$_{\text{idealized}}$ outperformed all other conditions on the delayed transfer test. Thus, one may speculate that the invent$_{\text{idealized}}$ condition was a constructive learning opportunity for all students. A possible theoretical explanation is that this PS-I design posed a desirable difficulty for all learners which unequivocally increased their germane load. In contrast, in the invent$_{\text{grounded}}$ condition, transfer performance depended more on the quality of the invented solutions. As said, the invention prompt is challenging and grounded cases might provide unnecessary details. If some learners are not able to ignore the details, these details may increase their extrinsic load and by this decrease their germane load. Thus, their quality of inventions suffers and with this their transfer performance. Importantly, these speculative explanations of how students' activity during learning is related to their transfer performance are based on explorative analyses and thus have to be taken with a grain of salt.

Limitations and Directions for Future Research

With the present study, we opted to achieve high ecological validity. Therefore, we conducted the study in vivo; that is, learners were tested in their mathematics classrooms. However, this benefit comes at a cost.

First, we had less control over how exactly learners processed the learning materials. A precise and tightly coupled measurement of cognitive load would be helpful to better track students learning processes when working on comparing and contrasting cases activities and when processing the instructional explanation. It has, for example, been shown that taking up student generated solutions and comparing them to a canonical solution in a direct instruction phase (i.e., when the teacher provides an instructional explanation) increases learning outcomes (Loibl & Rummel, 2014). We could not adapt the instructional explanation to individual student's solutions because we implemented the experimental variation within classrooms (to realize a real experiment with randomized distribution of learning materials). Furthermore, while we held the time of the different sequences constant across conditions, we cannot be sure that all students actually productively used the time. The explorative correlational analyses of students' activity during learning and their learning outcomes seem to indicate that the invent$_{\text{idealized}}$ condition was beneficial for all learners. However, the comparison of the correlations across the explain and invent conditions was inevitably based on different coding schemes for categorizing self-explanations and invented solutions. It is possible that these differences in the coding schemes influenced the correlational pattern. Therefore, further research is needed that directly tests whether specific PS-I sequences are beneficial for all learners while others may benefit only specific groups of students. This would require assessing student's domain-general reasoning abilities and their domain-specific knowledge and regressing their learning gains in PS-I sequences on these facets.

Second, it is possible that the superior performance of students who studied idealized cases coupled with the invention prompt might result from motivational effects. In explorative analyses, we showed that mastery and performance orientation (measured before instruction) did not substantially influence transfer performance, but we did not assess any kind of motivation after students had processed the learning materials. Glogger-Frey and colleagues (2015) reported that the invention prompt can increase curiosity and interest (Experiment 1). Thus, learners from the invent$_{\text{idealized}}$ condition in our study might have looked for more information about slopes on their own in the time between the immediate and delayed transfer assessment. We however assume that it is unlikely given that learners in the invent$_{\text{grounded}}$ condition did not increase their performance from the immediate to the delayed transfer assessment. Nevertheless, the motivational aspects of PS-I sequences are underexplored. If specific PS-I sequences benefit, for example, curiosity and interest, this might positively influence students' subsequent learning behavior. Here, studies are needed in which students' knowledge and motivational development is monitored over more extended instructional units (potentially comprised of repeated PS-I sequences). This would allow investigation of whether and to what extent subsequent learning behavior is indeed positively nudged by inventions prompts.

Third, our transfer tasks were situated in novel contexts that had no superficial similarity to any learning material, but we reminded them that the contents of the learning materials might help them to solve the tasks (i.e., our test should be considered a cued transfer test). This makes it difficult to clearly identify how students transferred their knowledge. Our explorative analyses for the two kinds of transfer tasks (i.e., items either requiring calculating solutions or conceptual descriptions), showed that the pattern for these subgroups of task matched the overall transfer performance pattern. A wider range of tasks could help to better describe the knowledge structure that students acquire during processing of different instructional sequences. For example, Schalk and colleagues (2016) varied the transfer distance and could thus more precisely describe transfer performance (also see De Bock et al., 2011; Nokes, 2009; for precise assessments of different kinds of transfer). Thus, future studies that aim to more deeply explore how PS-I sequences benefit knowledge construction and abstraction, should include a range of transfer tasks with varying distance to the learning materials.

Fourth, replications of the present results are needed. Directly applying our material to samples from other countries would deliver information about the robustness of the impact of the prompts and design features we studied on learning about graphs of linear functions. Beyond that, the broader generalizability of our findings needs to be investigated. We focused on two design aspects for which there are profound theoretical reasons and empirical support that they influence learners' processing and encoding of the cases. We believe that our description of the design factors allows researcher to develop conceptually similar learning materials and test the robustness of our findings across different contents and age groups. While not all results of the present study are particularly strong, they nevertheless clearly indicate that design aspects of PS-I activities cause differences in performance and thus provide a good starting point for future research.

Another aspect that has not yet received any attention in the PS-I literature concerns the relational complexity of the concept to be learned (for an extended discussion of relational complexity see Goldwater & Schalk, 2016). For example, the variance concept used in several previous PS-I studies (e.g., Kapur, 2012; Loibl & Rummel, 2014) is probably more challenging to learn than the concept of the slope of the graph of a linear function used in the present study which in turn might be more challenging than the density concept used in several studies (Glogger-Frey et al., 2015; Schwartz et al., 2011). Comparing the effect sizes across these studies seems to indicate that a less structured invention prompt could be more effective, the more relationally complex the concept is. However, this is at odds with research on the relational complexity (or element interactivity) of single worked examples which indicates that learning about more complex concepts benefits from more structure (Chen, Kalyuga, & Sweller, 2015). Thus, we suggest that future empirical research which systematically varies the complexity of concepts, the kind of prompts, and the representational characteristics of the cases in PS-I settings will be highly valuable both from a theoretical and practical perspective.

Conclusion

PS-I sequences have repeatedly been shown to benefit learning and transfer performance in comparison to a T&P sequence. In the present study, we implemented differently designed comparing and contrasting cases activities which are generally regarded to be an effective exploratory problem-solving activity in PS-I sequences to prepare students for a subsequent explanation. But, recent research together with the present study indicates that this beneficial effect is easily achieved. The effectiveness depends on design aspects such as the representational characteristics and the prompts used to scaffold the structural alignment of the cases. We showed that across two points of transfer assessment (immediately after learning and 4 weeks later), only learners who studied idealized cases scaffolded by invention prompts outperformed learners who first received an instructional explanation before solving practice tasks. Thus, making a PS-I sequence more efficient than a standard T&P sequence requires careful design of the learning materials.

References

Alfieri, L., Brooks, P. J., Aldrich, N. J., & Tenenbaum, H. R. (2011). Does discovery-based instruction enhance learning? *Journal of Educational Psychology, 103,* 1–18. http://dx.doi.org/10.1037/a0021017

Alfieri, L., Nokes-Malach, T. J., & Schunn, C. D. (2013). Learning through case comparisons: A meta-analytic review. *Educational Psychologist, 48,* 87–113. http://dx.doi.org/10.1080/00461520.2013.775712

Ankowski, A. A., Vlach, H. A., & Sandhofer, C. M. (2012). Comparison versus contrast: Task specifics affect category acquisition. *Infant and Child Development, 22,* 1–23. http://dx.doi.org/10.1002/icd.1764

Atkinson, R. K., Renkl, A., & Merrill, M. M. (2003). Transitioning from studying examples to solve problems: Combining fading with prompting fosters learning. *Journal of Educational Psychology, 95,* 774–783. http://dx.doi.org/10.1037/0022-0663.95.4.774

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin, 128,* 612–637. http://dx.doi.org/10.1037/0033-2909.128.4.612

Belenky, D. M., & Nokes-Malach, T. J. (2012). Motivation and transfer: The role of mastery-approach goals in preparation for future learning. *Journal of the Learning Sciences, 21,* 399–432. http://dx.doi.org/10.1080/10508406.2011.651232

Belenky, D. M., & Schalk, L. (2014). The effects of idealized and grounded materials on learning, transfer, and interest: An organizing framework for categorizing external knowledge representations. *Educational Psychology Review, 26,* 27–50. http://dx.doi.org/10.1007/s10648-014-9251-9

Berthold, K., & Renkl, A. (2009). Instructional aids to support a conceptual understanding of multiple representations. *Journal of Educational Psychology, 101,* 70–87. http://dx.doi.org/10.1037/a0013247

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). Cambridge, MA: The MIT Press.

Chen, U., Kalyuga, S., & Sweller, J. (2015). The worked examples effect, the generation effect, and element interactivity. *Journal of Educational Psychology, 107,* 689–704. http://dx.doi.org/10.1037/edu0000018

Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 13,* 145–182. http://dx.doi.org/10.1207/s15516709cog1302_1

Chin, D. B., Chi, M., & Schwartz, D. L. (2016). A comparison of two methods of active learning in physics: Inventing a general solution versus compare and contrast. *Instructional Science, 44,* 177–195. http://dx.doi.org/10.1007/s11251-016-9374-0

De Bock, D., Deprez, J., Van Dooren, W., Roelens, M., & Verschaffel, L. (2011). Abstract or concrete examples in learning mathematics? A replication and elaboration of Kaminski, Sloutsky, and Heckler's study. *Journal for Research in Mathematics Education, 42,* 109–126.

Deutsche Kultusministerkonferenz. (2004). *Bildungsstandards im Fach Mathematik fr̈ den Mittleren Schulabschluss*. M̌nchen, Germany: Wolters Kluwer.

Elliot, A. J., & Murayama, K. (2008). On the measurement of achievement goals: Critique, illustration, and application. *Journal of Educational Psychology, 100,* 613–628. http://dx.doi.org/10.1037/0022-0663.100.3.613

Fiorella, L., & Mayer, R. E. (2016). Eight ways to promote generative learning. *Educational Psychology Review, 28,* 717–741. http://dx.doi.org/10.1007/s10648-015-9348-9

Gattis, M., & Holyoak, K. J. (1996). Mapping conceptual to spatial relations in visual reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 231–239. http://dx.doi.org/10.1037/0278-7393.22.1.231

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

Glogger-Frey, I., Fleischer, C., Gr̈ny, L., Kappich, J., & Renkl, A. (2015). Inventing a solution and studying a worked solution prepare differently for learning from direct instruction. *Learning and Instruction, 39,* 72–87. http://dx.doi.org/10.1016/j.learninstruc.2015.05.001

Goldstone, R. L., & Sakamoto, Y. (2003). The transfer of abstract principles governing complex adaptive systems. *Cognitive Psychology, 46,* 414–466. http://dx.doi.org/10.1016/S0010-0285(02)00519-4

Goldwater, M. B., & Schalk, L. (2016). Relational categories as a bridge between cognitive and educational research. *Psychological Bulletin, 142,* 729–757. http://dx.doi.org/10.1037/bul0000043

Kapur, M. (2008). Productive failure. *Cognition and Instruction, 26,* 379–424. http://dx.doi.org/10.1080/07370000802212669

Kapur, M. (2012). Productive failure in learning the concept of variance. *Instructional Science, 40,* 651–672. http://dx.doi.org/10.1007/s11251-012-9209-6

Kapur, M. (2014). Productive failure in learning math. *Cognitive Science, 38,* 1008–1022. http://dx.doi.org/10.1111/cogs.12107

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

Lee, H. S., Betts, S., & Anderson, J. R. (2015). Not taking the easy road: When similarity hurts learning. *Memory & Cognition, 43,* 939–952. http://dx.doi.org/10.3758/s13421-015-0509-3

Loibl, K., Roll, I., & Rummel, N. (2016). Towards a theory of when and how problem-solving followed by instruction supports learning. *Educational Psychology Review*. Advance online publication. http://dx.doi.org/10.1007/s10648-016-9379-x

Loibl, K., & Rummel, N. (2014). Knowing what you don't know makes failure productive. *Learning and Instruction, 34,* 74–85. http://dx.doi.org/10.1016/j.learninstruc.2014.08.004

Mazziotti, C., Loibl, K., & Rummel, N. (2014). Does collaboration affect learning in a productive failure setting? In J. L. Polman, E. A. Kyza, D. K. O'Neill, I. Tabak, W. R. Penuel, A. S. Ju-row, . . . L. D'Amico (Eds.), In *Proceedings of the 11th international conference of the learning sciences* (Vol. 3, pp. 1184–1185). Boulder, CO: International Society of the Learning Sciences. Retrieved from https://www.isls.org/icls/2014/downloads/ICLS 2014 Volume 3 (PDF)-wCover.pdf

Mazziotti, C., Loibl, K., & Rummel, N. (2015). Collaborative or individual learning within productive failure: Does the social form of learning make a difference? In O. Lindwall, P. Häkkinen, T. Koschman, P. Tchounikine, & S. Ludvigsen (Eds.), *Exploring the material conditions of learning: The Computer Supported Collaborative Learning (CSCL) Conference* (Vol. 2, pp. 570–575). Gothenburg, Sweden: International Society of the Learning Sciences. Retrieved from https://www.isls.org/cscl2015/papers/CSCL2015ProceedingsVolume2.pdf

Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., & Chrostowski, S. J. (2004). *TIMSS 2003 international mathematics report: Findings from IEA's Trends in International Mathematics and Science Study at the fourth and eighth grades*. Retrieved from https://timss.bc.edu/PDF/t03_download/T03INTLMATRPT.pdf

Nathan, M. J. (2012). Rethinking formalisms in formal education. *Educational Psychologist, 47,* 125–148. http://dx.doi.org/10.1080/00461520.2012.667063

National Mathematics Advisory Panel. (2008). *Foundations for success: Final report of the National Mathematics Advisory Panel*. Retrieved from http://www.ed.gov/about/bdscomm/list/mathpanel/report/final-report.pdf

Nokes, T. J. (2009). Mechanisms of knowledge transfer. *Thinking & Reasoning, 15,* 1–36. http://dx.doi.org/10.1080/13546780802490186

Nokes-Malach, T. J., & Mestre, J. P. (2013). Toward a model of transfer as sense-making. *Educational Psychologist, 48,* 184–207. http://dx.doi.org/10.1080/00461520.2013.807556

Perry, L. K., Samuelson, L. K., Malloy, L. M., & Schiffer, R. N. (2010). Learn locally, think globally. Exemplar variability supports higher-order generalization and word learning. *Psychological Science, 21,* 1894–1902. http://dx.doi.org/10.1177/0956797610389189

Rau, M., Aleven, V., & Rummel, N. (2015). Successful learning with multiple graphical representations and self-explanation prompts. *Journal of Educational Psychology, 107,* 30–46. http://dx.doi.org/10.1037/a0037211

Renkl, A. (2014). Toward an instructionally oriented theory of example-based learning. *Cognitive Science, 38,* 1–37. http://dx.doi.org/10.1111/cogs.12086

Roelle, J., & Berthold, K. (2015). Effects of comparing contrasting cases on learning from subsequent explanations. *Cognition and Instruction, 33,* 199–225. http://dx.doi.org/10.1080/07370008.2015.1063636

Ross, B. H. (1987). This is like that: The use of earlier problems and the separation of similarity effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 629–639. http://dx.doi.org/10.1037/0278-7393.13.4.629

Ross, B. H. (1989). Distinguishing types of superficial similarities: Different effects on the access and use of earlier problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 456–468. http://dx.doi.org/10.1037/0278-7393.15.3.456

Schalk, L., Saalbach, H., & Stern, E. (2016). Approaches to foster transfer of formal principles: Which route to take? *PLoS ONE, 11,* e0148787. http://dx.doi.org/10.1371/journal.pone.0148787

Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. *Cognition and Instruction, 16,* 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,* 759–775. http://dx.doi.org/10.1037/a0025140

Schweizerische Konferenz der Erziehungsdirektoren. (2011). *Grundkompetenzen f̈r die Mathematik*. Retrieved from http://edudoc.ch/record/96784/files/grundkomp_math_d.pdf

Sears, D. (2006). *Effects of innovation versus efficiency tasks on collaboration and learning* (Doctoral dissertation). Stanford University, Stanford, CA. Retrieved from https://www.stat.auckland.ac.nz/~iase/publications/dissertations/06.Sears.Dissertation.pdf

Sidney, P. G., Hattikudur, S., & Alibali, M. W. (2015). How do contrasting cases and self-explanation promote learning? Evidence from fraction division. *Learning and Instruction, 40,* 29–38. http://dx.doi.org/10.1016/j.learninstruc.2015.07.006

Soderstrom, N. C., & Bjork, R. A. (2015). Learning versus performance: An integrative review. *Perspectives on Psychological Science, 10,* 176–199. http://dx.doi.org/10.1177/1745691615569000

Star, J. R., Pollack, C., Durkin, K., Rittle-Johnson, B., Lynch, K., Newton, K., & Gogolen, C. (2015). Learning from comparison in algebra. *Contemporary Educational Psychology, 40,* 41–54. http://dx.doi.org/10.1016/j.cedpsych.2014.05.005

Stern, E., Aprea, C., & Ebner, H. G. (2003). Improving cross-content transfer in text processing by means of active graphical representation. *Learning and Instruction, 13,* 191–203. http://dx.doi.org/10.1016/S0959-4752(02)00020-8

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

VanLehn, K. (1999). Rule learning events in the acquisition of a complex skill: An evaluation of cascade. *Journal of the Learning Sciences, 8,* 71–125. http://dx.doi.org/10.1207/s15327809jls0801_3

Williams, J. J., & Lombrozo, T. (2013). Explanation and prior knowledge interact to guide learning. *Cognitive Psychology, 66,* 55–84. http://dx.doi.org/10.1016/j.cogpsych.2012.09.002

Received January 31, 2017
Revision received August 10, 2017
Accepted August 10, 2017 ■