Micro productive failure and the acquisition of algebraic procedural knowledge

Esther Ziegler¹ · Dragan Trninic¹ · Manu Kapur¹

Received: 9 August 2019 / Accepted: 15 May 2021 / Published online: 29 May 2021
© The Author(s) 2021

Abstract

Productive failure has shown positive effects on conceptual and transfer measures, but no clear effects on procedural measures. It is therefore an open question whether, and to what extent, productive failure methods may be used to enhance the learning of procedural skills. A typical productive failure study focuses on a single, complex concept; in contrast, procedural knowledge generally consists of a series of less-complex procedural steps. In this study, failure occasions were adapted to specifically fit procedural knowledge by introducing procedural problems prior to the formal instruction of relevant principles. These procedural problems offered brief but multiple occasions for failure, which we call *micro productive failure*. A total of 85 sixth-graders were introduced to algebraic expression simplification by providing problem-solving prior to instruction (PS-I condition), compared to providing problem-solving after instruction (I-PS condition). Findings reveal a stable effect of offering micro productive failure occasions for procedural learning; however, as anticipated, there were no effects on conceptual or transfer measures.

Keywords Productive failure · Micro productive failure · Procedural skills · Procedural learning · Problem solving · Algebra learning

Introduction

The ordering of instruction prior to problem solving—that is, first instruction, then problem solving (I-PS)—has been long practiced without question. The pedagogical focus consisted in facilitating students' acquisition process by various means before giving them problems to solve; for example, by implementing some sort of guidance (Alfieri et al., 2011; Mayer, 2004), by adapting cognitive load (Mayer and Moreno 2003; Sweller, 1994), or by manipulating the presentation of materials (Alfieri et al., 2013; Ziegler et al., 2021; Rohrer and Pashler 2010). More recently, researchers have begun to methodically investigate the effects of changing the ordering of the instructional and the problem-solving phases (Darabi et al., 2018; Kapur, 2016; Loibl et al., 2017; Schwartz and Bransford 1998).

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

¹ Institute for Learning Sciences and Higher Education, ETH Zurich, Clausiusstrasse 59, 8092 Zurich, Switzerland

Instead of changing the instructional or problem-solving phases, these studies changed their *order*, asking students to engage in problem solving prior to formal instruction (PS-I). When given the opportunity to generate their own solutions prior to instruction, as one might expect, students typically fail to produce the correct solution; however, they benefit more from the subsequent instruction (Kapur, 2016; Schwartz and Bransford 1998). This hidden benefit of initial failure gave the instructional method its characteristic name of productive failure (Kapur, 2008; Kapur and Bielaczyc 2012).

Across a number of studies, productive failure is shown to primarily impact conceptual understanding and transfer, with no clear effects on procedural knowledge (Loibl et al., 2017). See Table 1 for an overview of relevant studies. Conceptual understanding refers to knowledge explicitly available, that is, the comprehension of principles and of interrelations between knowledge elements; in contrast, procedural knowledge refers to more implicit knowledge, that is, the ability to successfully perform a procedure (Rittle-Johnson et al., 2015). Transfer refers to the degree to which learned principles can be applied to different types of material and problems (Salomon and Perkins 1989). Conceptual and procedural knowledge are highly intertwined, rendering it difficult to distinguish and assess them independently (Ziegler et al. 2018; Schneider and Stern 2010). Nonetheless, different instructional approaches appear to foster one type of knowledge more than the other. In productive failure studies, gains on conceptual knowledge and transfer have primarily been shown in studies that included contrasting cases, and in studies where including procedural and conceptual instructional elements were incorporated through discussion; other studies showed more mixed results (Loibl et al., 2017).

A closer look at procedural knowledge outcomes reveals that—while a few studies indicate an advantage of PS-I (Kapur, 2010, 2011; Kapur and Bielaczyc 2012; Loehr et al., 2014), and a few studies indicate an advantage of I-PS (Fyfe et al., 2014; Loibl and Rummel 2014a, 2014b)—the majority of studies did not find a significant differences between the two approaches (DeCaro and Rittle-Johnson 2012; Gloger-Frey et al., 2015; Kapur, 2012, 2014; Kapur and Bielaczyc 2011, 2012; Loehr et al., 2014; Loibl and Rummel 2014a, 2014b; Roll et al., 2009; Schwartz et al., 2011). We offer that the interventions applied in those studies are not well-suited for procedural learning, and are instead designed to foster conceptual knowledge and transfer. Analysis of materials used in productive failure studies reveals that participants encountered problem-solving materials with a focus on conceptual understanding—for example, understanding the concepts of average speed, variance, equivalence, density, motion, collision, statistics process control, control of variable strategy, or evaluation of learning strategies (Loibl et al., 2017; Sinha and Kapur 2019). Across all studies, students were exposed to a single, intense problem-solving exploration phase, which lasted 15–60 min, before a formal instruction phase (or vice versa). We offer that this intense exploration phase may be better suited for developing conceptual understanding than training procedural skills.

To the best of our knowledge, no one has attempted to investigate the implementation of productive failure to (multi-stage) procedural problem-solving material, which is recurrent in mathematical algorithms. Failure with procedural material is sometimes described in terms of “impasses” students experience when they do not know how to proceed (Siegler and Jenkins 1989; VanLehn, 1999). The impasses drive students to seek a solution and repair missing or erroneous knowledge. In such studies, if students do not detect a solution, they may be provided information in the form of instructional feedback. In contrast, in productive failure studies, students are prompted to generate a solution even if their solution is incorrect, which may in turn encourage students to be creative and to explore more broadly. Because the effects of productive failure with complex conceptual problem-solving

Table 1 Results of PS-I vs. I-PS comparison, separately for procedural, conceptual and transfer measures

Result of the comparison	For procedural measures	For conceptual measures	For transfer measures
PS-I > I-PS	Kapur (2010, 2011), Kapur and Bielaczyc (2012); Loehr et al. (2014)	DeCaro and Rittle-Johnson (2012), Kapur (2010, 2011, 2012, 2014), Kapur and Bielaczyc (2011, 2012), Loibl and Rummel (2014a, 2014b), Roll et al. (2011)	Belenky and Nokes-Malach (2012), Kapur (2010, 2012, 2014), Kapur and Bielaczyc (2011); Roll et al. (2009); Schwartz and Bransford (1998), Schwartz et al. (2011), Schwartz and Martin (2004)
PS-I = I-PS	DeCaro and Rittle-Johnson (2012), Glogger-Frey et al. (2015), Kapur (2012, 2014), Kapur and Bielaczyc (2011, 2012), Loehr et al. (2014), Loibl and Rummel (2014a, 2014b), Roll et al. (2009), Schwartz et al. (2011)	Fyfe et al. (2014), Loehr et al. (2014), Loibl and Rummel (2014b)	Belenky and Nokes-Malach (2012), Kapur and Bielaczyc (2011), Matlen and Klahr (2013), Roll et al. (2009), Schwartz and Martin (2004)
I-PS > PS-I	Fyfe et al. (2014), Loibl and Rummel (2014a, 2014b)	Hsu et al. (2015), Loehr et al. (2014)	Glogger-Frey et al. (2015)

The references are studies that show the respective results. References in italics indicate that sub-studies show mixed results. The table is adapted from Loibl et al. (2017, see Table 3)

material are strong, and because conceptual understanding works hand-in-hand with procedural problem-solving (Ziegler et al. 2018; Rittle-Johnson et al., 2015), we conjectured that failure occasions may be similarly beneficial to the learning of procedural problem-solving routines.

Therefore, we offer that adapting productive failure for procedural learning may be accomplished by inserting failure occasions specifically adapted to the characteristics of procedural materials. Mathematical procedures generally consist of several principles or algorithms. Therefore, several productive failure occasions spread over the course of instruction (i.e., multiple cycles) might be better suited for procedural material than a single, intense productive failure problem-solving session as typically done in productive failure studies. Algebraic expression simplification is a suitable content for this task, being a procedure consisting of a series of principles, each of which is traditionally introduced then followed by an application to various problems.

We chose to focus on basic algebraic manipulation procedures, because they form a core part of algebra. For example, to solve algebraic equations, students need basic knowledge of simplifying algebraic expressions (Ziegler et al. 2019, 2021; Ottmar et al., 2012; Star and Rittle-Johnson 2008). We acknowledge the importance of conceptual understanding and transfer, yet argue that developing strong procedural knowledge in algebra is an important aim of instruction. Indeed, procedural knowledge lays a strong foundation for conceptual understanding and complex problem-solving activities (Fuson et al., 2005; Kershner and Awtry 2004). Furthermore, we note that algebraic procedural knowledge need not be blindly automated. The difficulty in learning algebraic procedures lies not in the complexity of rules to be learned and automated, but in the discrimination between different types of procedures. A flexible application of procedures therefore requires a degree of conceptual understanding (Ziegler et al. 2018; Sweller et al., 1998), and flexible procedural knowledge may be understood as an integral component of conceptual understanding and transfer.

As algebraic manipulation procedures consist of a series of principles, our aim is to exploit opportunities for learning from failure just before a new principle is introduced. This is accomplished by asking students to solve problems that require principles they have not yet learnt. In contrast to former productive failure designs on conceptual problem-solving material, where the problem solving prior to instruction unfolds over a more intensive and longer block of time, our problems are brief (<3 min) and do not require the generation of multiple solutions, but a single solution attempt. We refer to our design as learning from micro productive failure, because students are given micro opportunities to learn from failure prior to learning a specific algebraic manipulation principle. For an overview of the differences between productive failure and micro productive failure designs, see Table 2.

Initially withholding instruction means that—in most cases—student-generated solutions will be incorrect. Productive failure research has demonstrated that some failures can be productive during initial learning (Kapur, 2008; Schwartz and Bransford 1998). Broader research on desirable difficulties supports that conclusion as well (Bjork, 1994; Schmidt and Bjork 1992). Crucial for failure to be productive is the extent to which failure triggers students' active engagement with to-be-learned material, which may increase compensatory processes to tackle the increased difficulty (Kapur, 2014; Schmidt and Bjork 1992). In this way, the mechanisms initiated by difficulties prepare students to benefit more from the subsequent instruction. These mechanisms comprise searching for relevant prior knowledge (e.g., Schwartz et al., 2011; Sidney and Alibali 2015), comparing and contrasting to notice and encode critical features (Durkin and Rittle-Johnson 2012), noticing limits and inconsistencies to increase the likelihood of students selecting correct over incorrect

Table 2 Differences between typical productive failure and micro productive failure designs

	Productive failure occasions	Micro productive failure occasions
Duration	Long, 15–60 min	Brief, 2–3 min
Exploration phases	Single phase, in-depth	Multiple cycles, iterative
Material/concepts	Conceptual material one(two) complex challenging concept(s)	Procedural material a (multi-stage) procedure with a series of steps to be learned
Task	Generating multiple solutions to a problem	Generating a single solution to a problem, multiple times
Novelty	Numerous studies (see e.g., Kapur, 2016; Loibl et al. 2017)	Present study

knowledge and procedures (DeCaro and Rittle-Johnson 2012; Loibl et al., 2017; Siegler, 2002), and expectancy-violations or surprise effects, which come into effect after failed predictions and increase attention (Brod et al., 2018). In the present study, students were required to generate a solution even if they claimed not to know how to solve the problem, which may have led to an activation of prior knowledge and served to bring awareness to knowledge gaps. As the presented material involves contrasted algebraic addition and multiplication, the comparison and encoding of features and the noticing of limits and inconsistencies may have been facilitated even through single—rather than multiple—solutions to problems.

In summary, the aim of this study is to examine whether multiple cycles of micro productive failures are helpful when learning procedural problems. Algebraic procedural skills are not difficult per se, but require a careful distinction between structurally different yet superficially similar procedures—in the present study, this distinction refers to the learning of algebraic addition and algebraic multiplication. As we noted, most previous research suggests that procedural knowledge does not always benefit from productive failure (Kapur, 2016; Loibl et al., 2017). Our study material is designed to specifically fit the characteristic of procedural problem solving that consists of a series of addition and multiplication principles. Therefore, we expect micro productive failure to improve the discrimination of algebraic principles and their constituent procedural features. Due to the present focus on procedural problem solving, and the brief duration of failure occasions, we expect advantages on the procedural measures, but not on the conceptual measure. Thus, our expectations are distinct from previous studies on productive failure that focused on more complex, conceptual problem-solving material. In this study, providing micro problem-solving opportunities prior to instruction (PS-I condition) is compared to providing the same problem-solving occasions after instruction (I-PS condition, typical of more traditional instruction). To examine whether possible effects are stable over time, in addition to the posttest one day after the intervention, a one-week delayed posttest was conducted.

Method

Participants

Participants were 85 sixth-graders of classes from urban and suburban public schools in Switzerland. The students were volunteers, with their parents giving consent. The teachers of the participating classes received 150 CHF (approx. 150 USD) for their cooperation and time investment to organize the rooms and timing for the study, and each student was rewarded a small gift.

Even though the two conditions and all the research questions in the present study are new, the hypotheses and predictions were informed by similar theory, design, and materials as an existing study (Ziegler et al. 2019). We, therefore, expected similar, moderate effect sizes with a lower bound guess of about partial eta squared = 0.06 on the posttest outcome measures. A power analysis for a repeated measures ANOVA showed that an effect of this size, with a probability above 0.95, required a sample size of at least 54 students (Faul et al., 2007).

Teachers provided students' grades. After accounting for their mathematics and German grades, students (50.6% female) were randomly matched to one of two conditions within their classes. German grades were included to take into account that students had to provide

verbal explanations during the intervention; that is, the grades were included to control equal preconditions of groups. Four students missed intervention days and were therefore excluded from the analyses. In the PS-I condition, there were 41 students ($M = 12.42$ years, $SD = 0.49$), and 40 students in the I-PS condition ($M = 12.58$ years, $SD = 0.50$).

Design and procedure

The study used a 2×2 factorial design in which a between-participants instructional method factor (PS-I vs. I-PS) was crossed with a time factor (one-day later vs. one-week later).

- a. Students participated in three intervention sessions, and in two posttest sessions (for an overview, see Table 3). All sessions were 90-min in duration. The intervention was provided by the first author and a research assistant in groups of 10–14 students in school classrooms. The students worked on their own and petitioned the instructors if they had questions concerning the algebra material. In the initial session, a five-minute slide presentation showed how to use the expression “raise to the power of.” In Switzerland, algebra is not introduced by sixth grade; that is, the students in our study had not previously received formal algebra instruction.

Intervention materials: process measures

The intervention materials consisted of a paper–pencil self-study packet on simplifying algebraic expressions that included six units, each with a problem-solving sheet, an instructional worksheet, and an immediate learning test. Our instructional material was based on worked examples, a powerful form of direct instruction (e.g., Renkl, 1997; Van Merriënboer, 1997). In Appendix 2, a complete set of materials of unit 3 is given for both conditions.

During the intervention units, four process measures were assessed, which were indicators of performance during the different phases of the intervention: (1) a problem-solving (PS) measure (Sect. 2.3.1.); two measures during instruction: (2) example generation and (3) practice problems (Sect. 2.3.2.); and (4) an immediate learning test measure (Sect. 2.3.3.). The intervention materials used the same contrasted material validated in a former study, that is, the same instructional worksheets and immediate learning tests (Ziegler et al., 2019). For the current study, the material was extended with problem-solving sheets either before instruction and practice sections in the PS-I condition, or after instruction and practice sections in the I-PS condition.

Problem-solving sheets (process measure 1)

The six problem-solving sheets consist each of four algebraic procedural problems. These sheets correspond to the experimental manipulation. In both I-PS and PS-I conditions, identical sheets were processed but in different sequencing. In the PS-I condition, students processed the problems before receiving formal instruction in the subsequent worksheet. When students claimed not to know how to solve a problem, they were instructed

Table 3 Order of the activities in each session

Intervention sessions	1st day	Introduction	PS-I condition	I-PS condition
1st day		Unit 1 ^a	Algebra pretest	Instruction ^b I 1 (20–25 min) <i>a. worked examples^c</i> <i>b. example generation^c</i> <i>c. practice^d</i> Problem solving PS 1 (2–3 min)
			Slide presentation	
			Problem solving PS 1 (2–3 min)	
			Problem solving PS 1 (2–3 min)	
2nd day		Unit 2	Instruction ^b I 1 (20–25 min) <i>a. worked examples^c</i> <i>b. example generation^c</i> <i>c. practice^d</i>	Instruction ^b I 1 (20–25 min) <i>a. worked examples^c</i> <i>b. example generation^c</i> <i>c. practice^d</i> Problem solving PS 1 (2–3 min)
			Immediate learning test 1 (4–5 min)	
			Problem solving PS 2 (2–3 min)	
			Problem solving PS 2 (2–3 min)	
2nd day		Unit 3	Instruction ^b I 2 (20–25 min) <i>a. worked examples^c</i> <i>b. example generation^c</i> <i>c. practice^d</i>	Instruction ^b I 2 (20–25 min) <i>a. worked examples^c</i> <i>b. example generation^c</i> <i>c. practice^d</i> Problem solving PS 2 (2–3 min)
			Immediate learning test 2 (4–5 min)	
			Problem solving PS 3	
			Problem solving PS 3	
2nd day		Unit 3	Instruction ^b I 3 <i>a. Worked examples^c</i> <i>b. Example generation^c</i> <i>c. Practiced</i>	Instruction ^b I 3 <i>a. Worked examples^c</i> <i>b. Example generation^c</i> <i>c. Practiced</i> Problem solving PS 3
			Problem solving PS 3	
			Problem solving PS 3	
			Problem solving PS 3	

Table 3 (continued)

	PS-I condition	I-PS condition
3rd day	Unit 4	Instruction ^b I 4 a. <i>worked examples</i> ^c b. <i>example generation</i> ^c c. <i>practice</i> ^d Problem solving PS 4
	Unit 5	Instruction ^b I 5 a. <i>worked examples</i> ^c b. <i>example generation</i> ^c c. <i>practice</i> ^d Problem solving PS 5
	Unit 6	Instruction ^b I 6 a. <i>Worked examples</i> ^c b. <i>Example generation</i> ^c c. <i>Practice</i> ^d Problem solving PS 6

Table 3 (continued)

Posttest sessions		PS-I condition		I-PS condition
		1 day later	Posttest ^e	
	1 week later		Posttest ^e	Isomorphic problem-solving test
				Transfer test
				Conceptual explanation test
				Isomorphic problem-solving test
			Survey part	Transfer test
				Conceptual explanation test
				Personal data and reasoning ability test

^aA unit corresponded to one lesson. ^bThe instruction consists of a worksheet with a worked example section, an example generation section, and a practice section. ^cOn the frontside of the worksheets. ^dOn the backside of the worksheets. ^eThe posttest lasted two lessons

to generate their best solution. Thus, these problems served as an opportunity to attempt a solution with a high likelihood of failure. In the I-PS condition, students processed the identical problems *after* receiving instruction on the principles in the preceding instructional worksheet (see Sect. 2.3.2, below). Therefore, for I-PS students, these problems served as additional practice of the new learned problems. The total solution rate of correctly solved problems served as the first process measure to compare the implementation of the two conditions as either micro productive failure or additional practice; that is, it served as a manipulation check of the intervention.

Instructional worksheets (process measures 2 and 3)

Each instructional worksheet provided a comprehensive instruction suitable for the classroom, consisting of three sections: worked examples, example generation, and practice.

Worked example section At the top of each worksheet, two blocks of 2–3 *worked examples* were presented, one block of addition examples and one block of the same kind of multiplication examples in a contrasted way.

The worked examples consisted of two series of solved algebra problems, illustrating six addition and six multiplication learning steps in the order of increasing difficulty (Fig. 1). The six learning units consisted of the following principles: (1) dealing with like terms with single variables, e.g., “ $a + a =$ ”, or “ $a \cdot a =$ ”; (2) dealing with like terms with multiple variables, e.g., “ $ab + ab =$ ”, or “ $ab \cdot ab =$ ”; (3) dealing with like terms with exponents, e.g., “ $a^2 + a^2 =$ ”, or “ $a^2 \cdot a^2 =$ ”; (4) dealing with like terms with coefficients, e.g., “ $2a + 3a =$ ”, or “ $2a \cdot 3a =$ ”, and “ $ab + 2ab =$ ”, or “ $ab \cdot 2ab =$ ”; (5) dealing with unlike terms with single variables and constants, e.g., “ $a + b + a =$ ”, or “ $a \cdot b \cdot a =$ ” and “ $1 + a + 1 + a =$ ”, or “ $1 \cdot a \cdot 1 \cdot a =$ ”; and (6) dealing with unlike terms with multiple variables or variables with exponents, e.g., “ $ab + a + ab + a =$ ”, or “ $ab \cdot a \cdot ab \cdot a =$ ”, and “ $a^2 + a + a^2 + a =$ ”, or “ $a^2 \cdot a \cdot a^2 \cdot a =$ ”.

<p>A3</p> $xy + xy + xy = 3 \cdot xy$ $= 3xy$ $2b + 2b + 2b + 2b + 2b = 5 \cdot 2b$ $= 10b$ $3cx + 3cx = 2 \cdot 3cx$ $= 6cx$	<p>M3</p> $xy \cdot xy \cdot xy = x \cdot y \cdot x \cdot y \cdot x \cdot y$ $= x \cdot x \cdot x \cdot y \cdot y \cdot y$ $= x^3 \cdot y^3 = x^3y^3$ $2b \cdot 2b \cdot 2b \cdot 2b \cdot 2b = 2 \cdot b \cdot 2 \cdot b \cdot 2 \cdot b \cdot 2 \cdot b \cdot 2 \cdot b$ $= 2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 \cdot b \cdot b \cdot b \cdot b \cdot b$ $= 32 \cdot b^5 = 32b^5$ $3cx \cdot 3cx = 3 \cdot c \cdot x \cdot 3 \cdot c \cdot x$ $= 3 \cdot 3 \cdot c \cdot c \cdot x \cdot x$ $= 9 \cdot c^2 \cdot x^2 = 9c^2x^2$
<p>A4</p> $c^2 + c^2 + c^2 + c^2 = 4 \cdot c^2 = 4c^2$ $a^4 + a^4 = 2 \cdot a^4 = 2a^4$ $x^3 + x^3 + x^3 = 3 \cdot x^3 = 3x^3$	<p>M4</p> $c^2 \cdot c^2 \cdot c^2 \cdot c^2 = c \cdot c \cdot c \cdot c \cdot c \cdot c \cdot c \cdot c = c^8$ $a^4 \cdot a^4 = a \cdot a \cdot a \cdot a \cdot a \cdot a \cdot a \cdot a = a^8$ $x^3 \cdot x^3 \cdot x^3 = x \cdot x \cdot x \cdot x \cdot x \cdot x \cdot x \cdot x \cdot x \cdot x \cdot x \cdot x = x^9$
<p>A5</p> $2x + 5x + 2x = 9x$ $3bc + bc + 6bc = 3bc + 1bc + 6bc$ $= 10bc$ $y^3 + 4y^3 = 1y^3 + 4y^3$ $= 5y^3$	<p>M5</p> $2x \cdot 5x \cdot 2x = 2 \cdot x \cdot 5 \cdot x \cdot 2 \cdot x$ $= 2 \cdot 2 \cdot 5 \cdot x \cdot x \cdot x$ $= 20 \cdot x^3 = 20x^3$ $3bc \cdot bc \cdot 6bc = 3 \cdot b \cdot c \cdot b \cdot c \cdot 6 \cdot b \cdot c$ $= 3 \cdot 6 \cdot b \cdot b \cdot b \cdot c \cdot c \cdot c$ $= 18 \cdot b^3 \cdot c^3 = 18b^3c^3$ $y^3 \cdot 4y^3 = y \cdot y \cdot y \cdot 4 \cdot y \cdot y \cdot y$ $= 4 \cdot y \cdot y \cdot y \cdot y \cdot y \cdot y$ $= 4 \cdot y^6 = 4y^6$

Fig. 1 Excerpt of worked examples: addition learning steps (A3–A5), and multiplication learning steps (M3–M5). Grey-cursive = the intermediate steps to the solutions that were marked red in the original version; black-bold = the problems and the results

For multiplicative principles 5 and 6, the commutativity principle was introduced, e.g., $ab = a \cdot b = b \cdot a$.

Contrasting similar examples has shown positive effects in a series of former studies (Ziegler and Stern 2014; Ziegler et al. 2019; Rittle-Johnson and Star 2007; Star and Rittle-Johnson 2009). Students were instructed to self-explain the worked examples by detecting the underlying principles on their own and writing down how the worked algebra problems were solved (Renkl, 1997; Siegler, 2002). We used contrasted worked examples to highlight crucial aspects of learning material and thus help students learn the underlying principles. Students were prompted to look carefully at the worked examples and to conceptualize the algebra principles. The prompts required the students to verbally explain the worked examples and to explain the differences (e.g., Ziegler et al. 2019; Ziegler and Stern 2014). Students' conceptual explanations were assessed by the two assistants. If the explanations were incorrect, students were asked to correct them by attending at the worked examples again. If the explanations were too short, students were required to describe them in additional detail before moving on to the next section (a minimum of 4 lines of text, approx. 40 words, was required).

Example generation section In this section, students were required to generate two examples for the addition block and two examples for the multiplication block of worked examples. They were required “to invent varied and interesting examples using other numbers and letters”, and “to write down the intermediate solution steps.” Example generation is a recognized instructional method for students to process material in a deep way by applying principles to their own problems (Dahlberg and Housman 1997). The solution rate of correctly generated examples served as the second process measure to compare the effects of the two conditions on learning the material.

Practice section On the flipside of the worksheet, students were given six to eight problems per worksheet to practice the introduced principles. The practice problems were assessed by the two assistants. The ratio of correctly solved practice problems served as the third process measure to compare the effects of the two conditions on learning the material. Before moving on to the immediate learning tests, students received corrective feedback on practice problems and had to correct problems with the help of the worked examples.

Immediate learning tests (process measure 4)

After *each* of the six worksheets, students were asked to solve an immediate learning test, which served as the fourth process measure to assess their learning during the intervention sessions. Each learning test consisted of 6–8 algebra problems similar to those presented on the respective worksheet. Students' immediate learning was determined by the ratio of correctly solved problems averaged over the six immediate learning tests. Results were checked separately for the six single immediate learning tests. As the statistical results between the two conditions were highly similar, we averaged the results of all six tests to a total percentage of correctly solved problems.

Posttest assessments

Students participated in two posttest assessments of algebraic procedural knowledge. The same posttest was used one day and one week after the last day of the intervention.

The posttest assessments consisted of three outcome measures, which we describe below.

Isomorphic problem solving

The isomorphic problem-solving test contained 24 mixed addition and multiplication problems assessing students' ability to simplify algebraic expressions practiced in the intervention sessions: dealing with like terms with exponents, for example, " $c^2 \cdot c^2 \cdot c^2 =$ ", " $b^3 + b^3 =$ "; dealing with like terms with multiple variables, for example, " $ex + ex + x + x + ex =$ " and " $ax \cdot ax \cdot x =$ "; dealing with like terms with coefficients, for example, " $6\ mn + 6\ mn =$ ", " $3x^3 \cdot x^3 \cdot 2x^3 =$ "; and dealing with unlike terms, for example, " $z + n + n =$ ", " $b \cdot a \cdot 4 \cdot a \cdot 4 \cdot a =$ ", " $ax \cdot ax \cdot a =$ ", or " $c^2 + c + c^2 =$ ". Students' isomorphic problem solving was determined by the percentage of correct solutions to the 24 problems. The internal reliability of the isomorphic test was high, Cronbach's $\alpha = 0.89$.

Conceptual explanation

The conceptual explanation test assessed students' ability to explain how to apply algebraic addition and multiplication. Students were asked to write down two separate explanations about how each type of problem is solved. This was prompted with activating hints to describe, in detail, how algebraic expressions are simplified. For example, students were told, "imagine you would like to explain the rules to classmates", or "mention what one has to pay attention to." Students' explanations were rated and scored by two independent raters based on a coding scheme (Appendix 1). The coding scheme comprised a broad range of principles and features that could be mentioned in the descriptions. Two scores were assessed, representing the number of correctly reported concept features for addition and for multiplication. The two scores were added to the overall score of conceptual explanation. Inter-rater reliability was Cohen's $\kappa = 88.2\%$ for the first posttest assessment, and Cohen's $\kappa = 87.0\%$ for the second posttest assessment.

Transfer

The transfer test contained 15 mixed addition and multiplication problems assessing students' ability to transfer the acquired knowledge to new types of problems with principles not practiced in the intervention sessions. Note that this transfer test measured the transfer of knowledge of expression simplification, which is primarily procedural, and is not necessarily comparable to transfer tests that measure the transfer of deeper conceptual knowledge. The conceptual transfer in this study consisted of applying the learned principles to new types of problems, for example, dealing with like terms with three variables, " $3bcz \cdot 3bcz =$ ", and " $cdx + cdx =$ ", dealing with unlike terms with different exponents, " $a^3 \cdot a^2 \cdot a =$ ", and " $m^2 + m^4 + m^2 + m^4 =$ ", dealing with mixed unlike terms, " $a^2 \cdot a \cdot ay \cdot 4a =$ ", and " $u^2 \cdot ax \cdot u^2 \cdot u \cdot ax =$ ", or dealing with unlike terms with multiple variables with mixed exponents, for example, " $2bz^2 + bz + 4bz^2 + 3bz =$ ", and " $2c^2y \cdot c^2y^3 \cdot 3c^2y =$ ". Students' *transfer score* was determined by the percentage of correct solutions to the 15 problems. In our sample, the internal reliability of the transfer test was high, Cronbach's $\alpha = 0.88$.

Control measures

To control for students' pre-existing differences, four measures were used. *Prior algebra knowledge* was assessed with a pretest containing eight rudimentary algebra problems at the beginning of the intervention sessions, e.g., " $a+a+a+a=$ ", " $c \cdot c \cdot c=$ ", " $xy+xy+xy+xy=$ ", " $ab \cdot ab=$ ". The items were similar to the problems introduced in the first two intervention worksheets, but did not cover the more difficult problems of the subsequent worksheets, which students are unlikely to know before formal algebra instruction. *Reasoning ability* was assessed with figural and numerical subtests from the German intelligence test LPS (Horn, 1983) at the end of the last posttest session. Students' *grades in mathematics and German* were reported by the teachers.

Results

Control measures

A multivariate analysis of variance (MANOVA) on the control measures revealed no significant effect of condition (Table 4), $F(4,76)=0.16$, $p=0.958$, $\eta^2=0.01$, indicating a fair assignment of the students to the two conditions. A floor effect on the algebra pretest indicated that the students had almost no prior algebra knowledge. Out of eight algebra problems, on average the students solved less than one problem correctly ($M=0.82$, $SD=1.22$).

Effect of condition on the process measures

A multivariate analysis of covariance (MANCOVA) was calculated to examine the effect of condition on the four process measures *problem solving* (PS), *example generation* (I), *practice problems* (I), and *immediate learning tests* by including mathematics grade and reasoning ability as the two covariates (see Fig. 2). There was an effect of mathematics grade on the process measures, $F(4,74)=6.36$, $p<0.001$, $\eta^2_p=0.26$, whereas there was no effect of reasoning ability, $F(4,74)=0.32$, $p=0.862$, $\eta^2_p=0.02$. There was a main effect of condition on the process measures after controlling for mathematics grade and reasoning ability, $F(4,74)=81.12$, $p<0.001$, $\eta^2_p=0.81$.

Separate univariate ANOVAs on the process measures revealed an expected significant effect in favor of the I-PS condition on *problem solving*, $F(1,77)=243.42$, $p<0.001$, $\eta^2_p=0.76$, indicating that the implementation of the two conditions worked as anticipated, that is, expected success for the I-PS condition and expected failure for the PS-I condition. There was no significant effect on *example generation*, $F(1,77)=0.77$, $p=0.383$,

Table 4 Students' Performance by condition on the control variables

	PS-I <i>M</i> (<i>SD</i>)	I-PS <i>M</i> (<i>SD</i>)
Prior algebra knowledge (max. 8 points)	0.88 (1.12)	0.75 (1.33)
Reasoning ability (max. 40 points)	27.82 (4.26)	27.41 (3.83)
Mathematics grade ^a	4.88 (0.57)	4.91 (0.60)
German grade ^a	4.88 (0.55)	4.89 (0.59)

^aRange from 6 (best) to 1 (worst)

Fig. 2 Performance on the process measures. We use (PS) for assessment during the problem-solving phase, and (I) for assessment during the instruction phase. The error bars represent standard errors. Note that ordinate commences at 20 percent

$\eta^2_p = 0.01$, indicating that the initial failure did not hurt the instruction to generate examples. On subsequent *practice problems*, the effect reversed to a significant effect in favor of the PS-I condition, $F(1,77) = 8.22$, $p = 0.005$, $\eta^2_p = 0.10$, but no significant effect on *immediate learning tests*, albeit the means were, on average, still higher for the PS-I condition, $F(1,77) = 0.95$, $p = 0.333$, $\eta^2_p = 0.01$.

The algebra pretest, comprising problems of the first two worksheets, might have worked as a productive failure occasion for all the students. Therefore, we ran sub-analyses on the four process measures, comparing the results of the first two worksheets to the results of the subsequent four worksheets. The results of the sub-analyses were similar to the results across the six worksheets for three of the four measures. For practice problems, we found a significant difference between the PS-I and the I-PS conditions on the first two worksheets, $t(79) = 2.35$, $p = 0.021$, $d = 0.52$, distinct from later worksheets and the above reported practice problem outcomes. Thus, there might have been an influence of the pretest on the initial practice problems, but no influence on the subsequent practice problems.

Effect of condition on the posttest outcome measures

Separate mixed-factorial ANOVAs were conducted for each of the algebra knowledge measures, with the factors of condition (PS-I versus I-PS) and time (one-day versus one-week), and by including mathematics grade and reasoning ability as the two covariates. Overall, students in the PS-I condition performed better than students in the I-PS condition on isomorphic problem solving, but not on transfer and conceptual explanation (Fig. 2).

Isomorphic problem solving

The ANCOVA revealed that the covariate mathematics grade was significantly related to isomorphic problem solving, $F(1,72) = 9.11$, $p = 0.004$, $\eta^2_p = 0.11$, but reasoning ability was not, $F(1,72) = 0.84$, $p = 0.363$, $\eta^2_p = 0.01$. There was a main effect of condition on students' isomorphic problem solving in favor of the PS-I condition after controlling for mathematics grade and reasoning ability, $F(1,72) = 4.35$, $p = 0.041$, $\eta^2_p = 0.06$ (Fig. 3A). But there was no main effect of time, $F(1,72) = 0.25$, $p = 0.619$, $\eta^2_p = 0.00$, and also no interaction of condition $\times$ time, $F(1,72) = 0.13$, $p = 0.715$, $\eta^2_p = 0.00$, indicating a consistent advantage of the PS-I condition in the one-day and one-week delay.

Fig. 3 Performance on the posttest outcome measures. The error bars represent standard errors. ^aNote that ordinate commences at 60% (a), at 3 points (b), and at 40 percentage (c)

Conceptual explanation

The ANCOVA revealed that the covariate mathematics grade was significantly related to isomorphic problem solving, $F(1,72)=20.45$, $p<0.001$, $\eta^2_p=0.22$, but reasoning ability was not, $F(1,72)=3.40$, $p=0.069$, $\eta^2_p=0.05$. There was no main effect of condition on *conceptual explanation* after controlling for mathematics grade and reasoning ability, $F(1,72)=0.35$, $p=0.558$, $\eta^2_p=0.01$, and no main effect of time, $F(1,72)=0.14$, $p=0.713$, $\eta^2_p=0.00$ (Fig. 3C). However, there was an interaction of condition $\times$ time, $F(1,72)=4.46$, $p=0.038$, $\eta^2_p=0.06$. Post hoc analyses showed a higher but nonsignificant mean score of the I-PS condition compared to the PS-I condition at T1, $t(79)=-0.91$, $p=0.365$, $d=0.22$. The higher mean score of the I-PS condition turned to a lower but nonsignificant mean score in the one week delay at T2, $t(74)=0.98$, $p=0.329$, $d=0.23$.

Transfer

The ANCOVA revealed that the covariate mathematics grade was significantly related to isomorphic problem solving, $F(1,72)=18.47$, $p<0.001$, $\eta^2_p=0.20$, but reasoning ability was not, $F(1,72)=0.05$, $p=0.830$, $\eta^2_p=0.00$. There was no main effect of condition on *transfer* after controlling for mathematics grade and reasoning ability, $F(1,72)=0.01$, $p=0.936$, $\eta^2_p=0.00$ (Fig. 3b). There was also no main effect of time, $F(1,72)=0.79$, $p=0.378$, $\eta^2_p=0.01$, and no interaction of condition $\times$ time, $F(1,72)=1.53$, $p=0.220$, $\eta^2_p=0.02$.

Discussion

Our study examined the effect of micro productive failure on the learning of a procedural skill, simplifying algebraic expressions. Our results showed that a minimal intervention impacted the development of algebraic procedural knowledge. On isomorphic procedural problems, there was a stable positive effect of opportunities of micro productive failure. However, as expected due to the focus on procedural problem-solving material and the brief duration of failure occasions, the intervention did not show effects on the conceptual measure or on conceptual transfer.

This finding of micro productive failure on procedural problem-solving material is contrary to most earlier findings of productive failure on conceptual problem-solving material (DeCaro and Rittle-Johnson 2012; Kapur, 2014; Loibl et al., 2017). While those findings revealed clear effects on conceptual and transfer but not on procedural measures, our findings showed effects on procedural but not on the verbal conceptual measures or on conceptual transfer. The higher, albeit non-significant, mean of the I-PS condition on the verbal conceptual measure in the one-day delay turned to a lower mean in the one-week delay, an interaction that was significant. Because this result did not remain stable over time, this does not change our recommendation of starting with micro productive failure occasions. Indeed, our findings suggest that any effect might look markedly different even after one week, and that assessing learning in retention serves as an important boundary condition.

Mechanisms were not measured directly; however, our process measures indicated greater difficulty in the PS-I condition, which then outperformed the I-PS condition in the delay. This supports our expectation that productive failure mechanisms were involved. Having students generate a single solution, even if incorrect, and even when students explicitly claimed not to know how to solve the problem, may have been sufficient to activate prior knowledge and support the awareness of knowledge gaps (Schwartz et al., 2011; Sidney and Alibali 2015), at least when repetition was involved. This afforded students the opportunity to become aware of the limits of their own knowledge, and compare and contrast their solutions with the correct ones presented immediately after (DeCaro and Rittle-Johnson 2012; Durkin and Rittle-Johnson 2012). In this sense, micro failure occasions may have served as an opportunity to compare and contrast with the correct solution and notice the critical features (Roediger et al., 2011), leading to better learning of the correct solution (Roediger and Karpicke 2006). Better performance on the isomorphic problem solving also supports the argument that students learnt to select and deploy the correct solutions over the incorrect ones they had generated (Siegler, 1994, 2002).

The predicted absence of effect on conceptual and transfer measures can be explained by comparing our design with the productive failure design principles articulated by Kapur and Bielaczyc (2012). As argued in the Introduction, the productive failure design principles in the problem-solving tasks used in this study were different those used by Kapur and colleagues. In their studies, the problems were conceptual problem-solving tasks intended to provoke multiple and elaborated solution methods. In this study, students in the PS-I condition were prompted to generate a single solution of a problem before being presented the new principle, an approach we consider particularly appropriate for learning the principles of algebraic manipulation procedures. The continuous and repeated generation of even a single solution positively influenced learning, albeit the effect was on isomorphic problems only. Arguably, every request to generate a solution—that is, each micro failure occasions—served as an impasse that had to be overcome and thereby may have improved attention during subsequent instruction (Roediger and Karpicke 2006; VanLehn, 1999). Furthermore, for the students in the PS-I condition, their many repetitions might have led to increased expectation during instruction—in the sense of paying more attention to the crucial principles they initially failed on (Roediger et al., 2011). Some authors have described this mechanism as a surprise effect where, upon realizing that their solution was incorrect, students' subsequent enhanced attention contributes to deeper learning (Brod et al., 2018).

In terms of differences between our study and previous work, another point of consideration is that our material was not typical routine procedural material, but required a careful distinction between addition and multiplication procedures. Thus, our material is arguably more demanding than automated implicit procedural skills, requiring at least some conceptual or structural understanding of the material (Kieran, 1992). This might be an additional reason for the advantages of productive failure on procedural problem solving in our study. Finally, our conceptual explanation measure was different from typical measures assessing conceptual understanding in productive failure. The conceptual knowledge test in this study asked students to explain how each type of problem was solved—an approach that is, arguably, not equivalent to conceptual assessments used in previous studies. This difference in conceptual measures could have led to the divergent results from previous studies.

To summarize, micro productive failure is an instructional design introducing brief interventions that, in the present study, enhanced the learning of procedural skills. However, because they were brief and did not include reasoning about multiple and diverse solutions, these interventions had no impact on conceptual or transfer measures. Previous productive failure studies showed that the greater the diversity of student production, the more they are likely to learn from instruction (Kapur, 2014; Kapur and Bielaczyc 2012). Our failure occasions were brief interventions and consisted of generating a single solution, but repeatedly for each new principle (multiple cycles), which might be a better fit for learning a procedural skill, as seen in this study.

Further studies should examine an extension of micro failure intervention, for example by prompting students to generate more than one solution (Kapur and Bielaczyc 2012), or to verbally explain solution attempts (Ziegler et al., 2018), thus focusing on deeper learning. Whether presenting micro procedural units rather than larger conceptual problem chunks makes PS-I effective for procedural problem solving remains an open question, since this factor was not manipulated in our study. Studies varying the factor of problem-solving material (procedural versus conceptual) in addition to the productive failure factor (micro versus traditional) would be a similarly challenging but important next step. Finally, we note that a minimal intervention was enough to increase the acquisition of algebraic procedural knowledge. Minimal interventions that involve generating to-be-introduced principles or rules are potentially applicable in many instructional settings. Future research

could apply micro failure interventions to a broad range of mathematical topics that require procedural knowledge and skills, contributing to the development of more complex knowledge.

Nevertheless, some caution is warranted before attempting to apply these results to the classroom. Based on research that revealed correlations between procedural learning and poor mathematics achievement, one might conclude that procedural learning is generally favored for acquiring procedural knowledge. Yet, the pitfalls on overreliance on procedural competence have been known for some time (e.g., Erlwanger, 1973), indicating that procedural fluency without conceptual understanding is of limited value. Complicating the matter further is a general lack of agreement in the field on how to define and operationalize conceptual understanding (see Crooks and Alibali 2014, for a literature review). That is to say, the interactions between procedural and conceptual knowledge are nuanced and multi-layered, and need to be carefully studied if precise implications for the classroom are to be made. As noted by an anonymous reviewer, further research in this area would do well to explore whether and to what extent is conceptual understanding needed for procedural fluency—what is the nature of this conceptual understanding and what does it afford if it is present?

Like all interventions, the present study is limited: it examined a short, four-day instruction of algebra compared to the intense algebra instruction students receive at school over the years. A complete algebra instruction will always touch all facets of algebra: from understanding of variables and manipulating terms, to applying and understanding algebra.

Appendix 1

Coding scheme for the conceptual explanation test.

For every one of the following features mentioned in the answer, one point was scored. The features were the principles that were considered central and essential for distinguishing algebraic addition from algebraic multiplication. Points were equally given if: (a) a key element was explicitly mentioned, or (b) a key element became visible in the solution or the intermediate steps of the example. If a point was described incorrectly, it was counted as a mistake

Sub-analysis of Addition (total 6 points):

1: Sorting by summands (summands are the letter endings (sometimes with exponents: a, ab, a^2 , ...))

→ only 1/2 point: if only single letters were mentioned or used (a, b, c, ...)

→ 2 points^a: if sufficiently detailed in a way that examples became redundant or dispensable

2: Summands are not split

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

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

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

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

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

1. Sorts according to identical summands. Pays attention to double letter variables and variables with exponents, which are added unchanged, i.e. cx and c^2 are not separated. This underscores the identical combinations of letters

2. Each type of unit (cx , c^2 , c , 2) is added separately by adding the numbers of summands and by appending the unit unchanged. Detached numbers are added separately

3. The result consists of the number of each different unit connected with a plus sign

Sub-analysis of Multiplication (total 5 points):

1: All factors are split

→ only ½ point: if only single letters are mentioned or used ($2x \cdot 3y \cdot 3x$)

→ 2 points^a: if sufficiently detailed in a way that examples became redundant or dispensable

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

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

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

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

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

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

^aEven if 2 points are given, the total score cannot exceed 6 points in addition and 5 points in multiplication because the same information was only counted once either in a convincing comprehensive argument or in mentioning sub-points

Appendix 2

PS-I condition: Complete set of materials of Unit 3.

Problem solving sheet *Problem solving PS*
(2–3 min)

$$\begin{aligned} z^2 + z^2 + z^2 &= \\ y^3 + y^3 &= \\ x^2 \cdot x^2 \cdot x^2 \cdot x^2 &= \\ e^3 \cdot e^3 &= \end{aligned}$$

Instructional worksheet *Instruction I*
(20–25 min)

a. worked examples

$$\begin{aligned} c^2 + c^2 + c^2 + c^2 &= 4 \cdot c^2 = 4c^2 \\ a^4 + a^4 &= 2 \cdot a^4 = 2a^4 \\ x^3 + x^3 + x^3 &= 3 \cdot x^3 = 3x^3 \end{aligned}$$

$$\begin{aligned} c^2 \cdot c^2 \cdot c^2 \cdot c^2 &= \\ c^2 &= c \cdot c \cdot c \cdot c \\ c &= c^8 \\ a^4 \cdot a^4 &= a \cdot a \cdot a \cdot a \cdot a \cdot a \cdot a \cdot a = a^8 \\ x^3 \cdot x^3 \cdot x^3 &= x \cdot x \cdot x \cdot x \cdot x \cdot x \cdot x \cdot x \cdot x = x^9 \end{aligned}$$

With the following prompts:

Compare the examples in the left and the right column together. What do you notice? Describe what is different!
Compare the addition and multiplication problems and describe the different steps needed to solve them

b. example generation

Invent 2 varied examples, along with the solution steps, for each box. Write down all the intermediate steps to the solution like those given in the examples

c. practice

$$\begin{aligned} a^2 + a^2 + a^2 + a^2 + a^2 &= \\ e^4 + e^4 &= \\ z^2 \cdot z^2 \cdot z^2 \cdot z^2 &= \\ s^3 \cdot s^3 &= \end{aligned}$$

I-PS condition: Complete set of materials of Unit 3.

Acknowledgements We are grateful to Sara Ziegler for her assistance with the school classes, and Sara Ziegler and Nadine Schmid for their assistance with the evaluation of the test materials. We thank the students, teachers, and principals of the participating schools for their support of this project. Tanmay Sinha and Nadja Beeler (ETH Zurich) provided valuable feedback on the manuscript. Finally, we are grateful to the reviewers for their insightful comments and suggestions.

Funding Open Access funding provided by ETH Zurich.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

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), 1–18. <https://doi.org/10.1037/A0021017>
- Alfieri, L., Nokes, T. J., & Schunn, C. D. (2013). Learning through case comparisons: A metaanalytic review. *Educational Psychologist*, 48(2), 87–113. <https://doi.org/10.1080/00461520.2013.775712>
- Belenky, D. M., & Nokes-Malach, T. J. (2012). Motivation and transfer: The role of mastery-approach goals in preparation for future learning. *The Journal of the Learning Sciences*, 21(3), 399–432. <https://doi.org/10.1080/10508406.2011.651232>
- Bjork, R. (1994). Institutional impediments to effective training. In D. Druckman & R. A. Bjork (Eds.), *Learning, remembering, believing: Enhancing human performance* (pp. 295–306). National Academy Press.
- Brod, G., Hasselhorn, M., & Bunge, S. A. (2018). When generating a prediction boosts learning: The element of surprise. *Learning and Instruction*, 55(1), 22–31. <https://doi.org/10.1016/j.learninstruc.2018.01.013>
- Crooks, N. M., & Alibali, M. W. (2014). Defining and measuring conceptual knowledge in mathematics. *Developmental Review*, 34(4), 344–377. <https://doi.org/10.1016/j.dr.2014.10.001>
- Dahlberg, R. P., & Housman, D. L. (1997). Facilitating learning events through example generation. *Educational Studies in Mathematics*, 33(3), 293–299. <https://doi.org/10.1023/A:1002999415887>
- Darabi, A., Arrington, T. L., & Sayilir, E. (2018). Learning from failure: A meta-analysis of the empirical studies. *Educational Technology Research and Development*, 66(5), 1101–1118. <https://doi.org/10.1007/s11423-018-9579-9>
- DeCaro, M. S., & Rittle-Johnson, B. (2012). Solving math problems prepares students to learn from instruction. *Journal of Experimental Child Psychology*, 113(4), 552–568. <https://doi.org/10.1016/j.jecp.2012.06.009>
- Durkin, K., & Rittle-Johnson, B. (2012). The effectiveness of using incorrect examples to support learning about decimal magnitude. *Learning and Instruction*, 22(3), 206–214. <https://doi.org/10.1016/j.learninstruc.2011.11.001>
- Erlwanger, S. H. (1973). Benny's conception of rules and answers in IPI mathematics. *Journal of Children's Mathematical Behavior*, 1(2), 7–26.
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/BF03193146>
- Fuson, K. C., Kalchman, M., & Bransford, J. D. (2005). *Mathematical understanding: An introduction*. National Academies Press.
- Fyfe, E. R., DeCaro, M. S., & Rittle-Johnson, B. (2014). An alternative time for telling. *British Journal of Educational Psychology*, 84(3), 502–519. <https://doi.org/10.1111/bjep.12035>
- Glogger-Frey, I., Fleischer, C., Grüny, 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(1), 72–87. <https://doi.org/10.1016/j.learninstruc.2015.05.001>
- Horn, W. (1983). *Leistungsprüfsystem LPS* (2nd ed.). Hogrefe.
- Hsu, C.-Y., Kalyuga, S., & Sweller, J. (2015). When should guidance be presented in physics instruction? *Archives of Scientific Psychology*. <https://doi.org/10.1037/arc0000012>
- Kapur, M. (2008). Productive failure. *Cognition and Instruction*, 26(3), 379–424. <https://doi.org/10.1080/07370000802212669>

- Kapur, M. (2010). Productive failure in mathematical problem solving. *Instructional Science*, 38(6), 523–550. <https://doi.org/10.1007/s11251-009-9093-x>
- Kapur, M. (2011). A further study of productive failure in mathematical problem solving: Unpacking the design components. *Instructional Science*, 39(4), 561–579. <https://doi.org/10.1007/s11251-010-9144-3>
- Kapur, M. (2012). Productive failure in learning the concept of variance. *Instructional Science*, 40(4), 651–672. <https://doi.org/10.1007/s11251-012-9209-6>
- Kapur, M. (2014). Productive failure in learning math. *Cognitive Science*, 38(5), 1008–1022. <https://doi.org/10.1111/cogs.12107>
- Kapur, M. (2016). Examining productive failure, productive success, unproductive failure, and unproductive success in learning. *Educational Psychologist*, 51(2), 289–299. <https://doi.org/10.1080/00461520.2016.1155457>
- Kapur, M., & Bielaczyc, K. (2011). Classroom-based experiments in productive failure. In L. Carlson, C. Hoelscher, & T. F. Shipley (Eds.), *Proceedings of the 33rd annual conference of the Cognitive Science Society* (pp. 2812–2817). Austin, TX: Cognitive Science Society.
- Kapur, M., & Bielaczyc, K. (2012). Designing for productive failure. *The Journal of the Learning Sciences*, 21(1), 45–83. <https://doi.org/10.1080/10508406.2011.591717>
- Kieran, C. (1992). The learning and teaching of school algebra. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 390–419). Macmillan Publishing Company.
- Kirshner, D., & Awtry, T. (2004). Visual salience of algebraic transformations. *Journal for Research in Mathematics Education*, 35(4), 224–257. <https://doi.org/10.2307/30034809>
- Loehr, A. M., Fyfe, E. R., & Rittle-Johnson, B. (2014). Wait for it... delaying instruction improves mathematics problem solving: A classroom study. *Journal of Problem Solving*, 7(1), 36–49. <https://doi.org/10.7771/1932-6246.1166>
- Loibl, K., Roll, I., & Rummel, N. (2017). Towards a theory of when and how problem solving followed by instruction supports learning. *Educational Psychology Review*, 29(4), 693–715. <https://doi.org/10.1007/s10648-016-9379-x>
- Loibl, K., & Rummel, N. (2014a). The impact of guidance during problem-solving prior to instruction on students' inventions and learning outcomes. *Instructional Science*, 42(3), 305–326. <https://doi.org/10.1007/s11251-013-9282-5>
- Loibl, K., & Rummel, N. (2014b). Knowing what you don't know makes failure productive. *Learning and Instruction*, 34(1), 74–85. <https://doi.org/10.1016/j.learninstruc.2014.08.004>
- Matlen, B. J., & Klahr, D. (2013). Sequential effects of high and low instructional guidance on children's acquisition of experimentation skills: Is it all in the timing? *Instructional Science*, 41(3), 621–634. <https://doi.org/10.1007/s11251-012-9248-z>
- Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning. *American Psychologist*, 59, 14–19. <https://doi.org/10.1037/0003-066X.59.1.14>
- Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, 38(1), 43–52. https://doi.org/10.1207/S15326985EP3801_6
- Ottmar, E., Landy, D., & Goldstone, R. L. (2012). *Teaching the perceptual structure of algebraic expressions: Preliminary findings from the pushing symbols intervention*. Paper presented at the Thirty-Fourth Annual Conference of the Cognitive Science Society, Sapporo, Japan.
- Renkl, A. (1997). Learning from worked-out examples: A study on individual differences. *Cognitive Science*, 21(1), 1–29. [https://doi.org/10.1016/S0364-0213\(99\)80017-2](https://doi.org/10.1016/S0364-0213(99)80017-2)
- Rittle-Johnson, B., Schneider, M., & Star, J. R. (2015). Not a one-way street: Bidirectional relations between procedural and conceptual knowledge of mathematics. *Educational Psychology Review*, 27(4), 587–597. <https://doi.org/10.1007/s10648-015-9302-x>
- Rittle-Johnson, B., & Star, J. R. (2007). Does comparing solution methods facilitate conceptual and procedural knowledge? An experimental study on learning to solve equations. *Journal of Educational Psychology*, 99(3), 561–574. <https://doi.org/10.1037/0022-0663.99.3.561>
- Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, 17(3), 249–255.
- Roediger, H. L., Putnam, A. L., & Smith, M. A. (2011). Ten benefits of testing and their applications to educational practice. *Psychology of Learning and Motivation*, 55(1), 1–36. <https://doi.org/10.1016/B978-0-12-387691-1.00001-6>
- Rohrer, D., & Pashler, H. (2010). Recent research on human learning challenges conventional instructional strategies. *Educational Researcher*, 39(5), 406–412. <https://doi.org/10.3102/0013189X10374770>
- Roll, I., Alevén, V., & Koedinger, K. R. (2009). Helping students know 'further'—increasing the flexibility of students' knowledge using symbolic invention tasks. In N. A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st annual conference of the Cognitive Science Society* (pp. 1169–1174). Austin, TX: Cognitive Science Society.

- Roll, I., Alevén, V., & Koedinger, K. R. (2011). Outcomes and mechanisms of transfer. In L. Carlson, C. Hoelscher, & T. F. Shipley (Eds.), *Proceedings of the 33rd annual conference of the Cognitive Science Society* (pp. 2824–2829). Austin, TX: Cognitive Science Society.
- Salomon, G., & Perkins, D. N. (1989). Rocky roads to transfer: Rethinking mechanism of a neglected phenomenon. *Educational Psychologist*, 24(2), 113–142. https://doi.org/10.1207/s15326985ep2402_1
- Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science*, 3(4), 207–217. <https://doi.org/10.1111/j.1467-9280.1992.tb00029.x>
- Schneider, M., & Stern, E. (2010). The developmental relations between conceptual and procedural knowledge: A multimethod approach. *Developmental Psychology*, 46(1), 178–192. <https://doi.org/10.1037/a0016701>
- Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. *Cognition and Instruction*, 16(4), 475–522. https://doi.org/10.1207/s1532690xci1604_4
- Schwartz, D. L., Chase, C. C., Oppezzo, M. A., & Chin, D. B. (2011). Practicing versus inventing with contrasting cases: The effects of telling first on learning and transfer. *Journal of Educational Psychology*, 103(4), 759–775. <https://doi.org/10.1037/a0025140>
- Schwartz, D. L., & Martin, T. (2004). Inventing to prepare for future learning. *Cognition and Instruction*, 22(2), 129–184. https://doi.org/10.1207/s1532690xci2202_1
- Sidney, P. G., & Alibali, M. W. (2015). Making connections in math: Activating a prior knowledge analogue matters for learning. *Journal of Cognition and Development*, 16(1), 160–185. <https://doi.org/10.1080/15248372.2013.792091>
- Siegler, R. S. (1994). Cognitive variability: A key to understanding cognitive development. *Current Directions in Psychological Science*, 3(1), 1–5. <https://doi.org/10.1111/1467-8721.ep10769817>
- Siegler, R. S. (2002). Microgenetic studies of self-explanation. In N. Granott & J. Parziale (Eds.), *Microdevelopment: Transition processes in development and learning* (pp. 31–58). Cambridge University Press.
- Siegler, R. S., & Jenkins, E. (1989). *How children discover new strategies*. Lawrence Erlbaum Associates.
- Sinha, T., & Kapur, M. (2019). When productive failure fails. In A. K. Goel, C. M. Seifert, & C. Freska (Eds.), *Proceedings of the 41th annual conference of the Cognitive Science Society* (pp. 2811–2817). Montreal, Canada.
- Star, J. R., & Rittle-Johnson, B. (2008). Flexibility in problem-solving: The case of equation solving. *Learning and Instruction*, 18(6), 565–579. <https://doi.org/10.1016/j.learninstruc.2007.09.018>
- Star, J. R., & Rittle-Johnson, B. (2009). It pays to compare: An experimental study on computational estimation. *Journal of Experimental Child Psychology*, 102(4), 408–426. <https://doi.org/10.1016/j.jecp.2008.11.004>
- Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, 4(4), 295–312. [https://doi.org/10.1016/0959-4752\(94\)90003-5](https://doi.org/10.1016/0959-4752(94)90003-5)
- Sweller, J., Van Merriënboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10(3), 251–296. <https://doi.org/10.1023/A:1022193728205>
- Van Merriënboer, J. (1997). *Training complex cognitive skills: A four-component instructional design model for technical training*. Educational Technology Publications.
- VanLehn, K. (1999). Rule-learning events in the acquisition of a complex skill: An evaluation of cascade. *The Journal of the Learning Sciences*, 8(1), 71–125. https://doi.org/10.1207/s15327809jls0801_3
- Ziegler, E., & Stern, E. (2014). Delayed benefits of learning elementary algebraic transformations through contrasted comparisons. *Learning and Instruction*, 33(1), 131–146. <https://doi.org/10.1016/j.learninstruc.2014.04.006>
- Ziegler, E., Edelsbrunner, P. A., & Stern, E. (2018). The relative merits of explicit and implicit learning of contrasted algebra principles. *Educational Psychology Review*, 30(1), 531–558. <https://doi.org/10.1007/s10648-017-9424-4>
- Ziegler, E., Edelsbrunner, P. A., & Star, J. R. (2019). Preventing Interference: Reordering Complexity in the Learning of New Concepts. *Journal of Educational Psychology*, 111(7), 1202–1219. <https://doi.org/10.1037/edu0000347>
- Ziegler, E., Edelsbrunner, P. A., & Stern, E. (2021). The benefits of combining teacher-direction with contrasted presentation of algebra principles. *European Journal of Psychology of Education*, 36(1), 187–218. <https://doi.org/10.1007/s10212-020-00468-3>