Educ Psychol Rev
DOI 10.1007/s10648-016-9379-x

REVIEW ARTICLE

Towards a Theory of When and How Problem Solving Followed by Instruction Supports Learning

Katharina Loibl[1] **· Ido Roll**[2] **· Nikol Rummel**[3]

Abstract Recently, there has been a growing interest in learning approaches that combine two phases: an initial problem-solving phase followed by an instruction phase (PS-I). Two often cited examples of instructional approaches following the PS-I scheme include Productive Failure and Invention. Despite the growing interest in PS-I approaches, to the best of our knowledge, there has not yet been a comprehensive attempt to summarize the features that define PS-I and to explain the patterns of results. Therefore, the first goal of this paper is to map the landscape of different PS-I implementations, to identify commonalities and differences in designs, and to associate the identified design features with patterns in the learning outcomes. The review shows that PS-I fosters learning only if specific design features (namely contrasting cases or building instruction on student solutions) are implemented. The second goal is to identify a set of interconnected cognitive mechanisms that may account for these outcomes. Empirical evidence from PS-I literature is associated with these mechanisms and supports an initial theory of PS-I. Finally, positive and negative effects of PS-I are explained using the suggested mechanisms.

Keywords Contrasting cases · Invention · Learning mechanisms · Problem solving · Productive Failure · Student solutions · Compare and contrast

Introduction

Recently, there has been a growing interest in learning approaches that include two phases: an initial problem-solving phase followed by an instruction phase (PS-I). Two commonly cited examples of instructional approaches that apply the PS-I structure include Productive Failure

✉ Katharina Loibl
katharina.loibl@ph-freiburg.de

[1] University of Education Freiburg, Kunzenweg 21, 79117 Freiburg, Germany

[2] University of British Columbia, Vancouver, Canada

[3] Ruhr-Universität Bochum, Bochum, Germany

Published online: 18 July 2016

(e.g., Kapur 2016; Kapur and Bielaczyc 2012) and Invention (e.g., Schwartz and Martin 2004). The growing interest in these approaches is reflected in a striking number of recent research publications (e.g., PsychINFO lists 11 journal papers that include "Productive Failure" in the title published in 2010–2014) as well as articles aiming at a broader audience (e.g., article in the *Time Magazine*: Paul 2012). Despite this strong interest, to the best of our knowledge, there has not yet been a comprehensive attempt to summarize the features that define PS-I, study their impact on learning, and explain the patterns of results. In this paper, we map the landscape of PS-I approaches, relate characteristic design features to learning, and discuss potential cognitive mechanisms triggered by PS-I. Notably, by choosing this focus, we do not aim to compare PS-I to alternative approaches in general but to take a closer look at the mechanisms at play in PS-I.

PS-I approaches include a problem-solving phase and an instruction phase. During the problem-solving phase, students attempt to solve a problem requiring the application of a yet to-be-learned concept and usually fail to solve the problem successfully (Kapur 2010, 2012). For example, in a study by Kapur (2012) on the topic of variability, students were given data regarding different athletes and were asked to identify the most consistent athlete during the problem-solving phase. Subsequently, students were taught the canonical solution explicitly. We therefore refer to this second phase as the explicit instruction phase. In the example given above, students were instructed about standard deviation and then applied it to the problem at hand. In another example (Glogger-Frey et al. 2015), student-teachers were first given samples of learning diaries and were asked to invent criteria for evaluating the application of learning strategies in these samples. Subsequent explicit instruction introduced the desired criteria for evaluations of learning diaries.

By including an explicit instruction phase, PS-I differs from other inductive methods such as inquiry or (guided) discovery learning (cf. Loibl and Rummel 2014b). In inquiry or guided discovery learning, the ultimate goal is that students discover the underlying model or concept on their own, with various forms of support. In contrast, the problem-solving phase in PS-I is not designed to facilitate the acquisition of the target concept (cf. Kalyuga and Singh 2015) as the concept is taught during the subsequent explicit instruction phase. By asking students to engage in problem solving prior to being taught the target knowledge, PS-I differs from other instructional methods with upfront instruction. In summary, the uniqueness of PS-I lies not in its components themselves (i.e., inductive problem solving and explicit instruction); it is their *combination* and their *order* that define an instructional approach as PS-I. While the combination and the order of the two phases are common to all PS-I approaches, the specific implementation of the phases may differ across manipulations.

The goals of the current paper are twofold. First, PS-I implementations are mapped and commonalities and differences in their *instructional design* are identified. The review focuses explicitly on studies that investigate effects of the two-phase PS-I approach and does not consider studies that evaluate only one of the two phases (see Section 2 for details on the selection process). In order to identify patterns in the effectiveness of different PS-I implementations, we clustered the reviewed studies according to design features and analyzed the *learning outcomes* within each cluster. Second, we identified a set of interconnected *cognitive mechanisms* that may account for these outcomes (Section 3). We grounded these proposed mechanisms in broader literature. While some of the mechanisms can directly be associated with PS-I, others are more speculative and are so far supported only indirectly by referring to related research. The paper puts forward a starting point for working towards a theory of PS-I; we discuss where more research is needed to evaluate certain conjectures.

Commonalities and Difference in PS-I Approaches and Their Impact on Learning

While the overall structure (problem solving followed by explicit instruction) is common to all PS-I approaches, differences exist in the way PS-I approaches implement the two phases. Therefore, a clear definition of the differences across PS-I is needed. First, we identify design features that span the space of PS-I implementations. Second, we relate these features to patterns in learning outcomes.

We searched the databases PsycINFO and ERIC for published papers from the last 10 years that use one of the following terms: “productive failure,” “invention activities,” “initial problem solving,” or a combination of “problem solving” and either “direct instruction” or “explicit instruction.” Last, we analyzed influential studies cited by most of the reviewed papers that preceded this period (Schwartz and Bransford 1998; Schwartz and Martin 2004). We then narrowed the list to include only studies that compared the two-phase PS-I approach to an I-PS approach (i.e., an approach with the reverse order of the two phases; often termed Direct Instruction condition). The I-PS control condition is needed as a baseline to identify differences in the outcomes of different PS-I implementations. More precisely, positive or negative effects of PS-I in comparison to the control condition offer indirect insight into the impact of certain design elements that are not directly compared with one another. Note that the main aim of this review was not to determine whether PS-I is superior or inferior than other instructional designs in general but to take a closer look at patterns related to different PS-I implementations. A broader comparison of PS-I to other instructional designs would require a more detailed analysis of the learning goals related to all design features of the different designs (cf. Kalyuga and Singh 2015). As we are interested in the impact of differences in the PS-I implementations on learning outcomes, we further narrowed the list to include only studies that measured learning. Notably, many of the selected studies were conducted by Kapur et al.. However, as evident in Table 1, PS-I has also been investigated by a larger number and variety of research groups.

Table 1 presents the selected papers. The first column assigns a *number* to each paper; the second column includes the *reference*. Columns 3 and 4 provide short descriptions of the implemented conditions (*PS-I group* and *control group*), and column 5 provides a summary of the *learning outcomes and effect sizes* (if reported) regarding procedural knowledge, conceptual knowledge, and transfer. Procedural knowledge refers to the ability to correctly apply a learned procedure (Rittle-Johnson and Schneider 2014). For example, students are asked to apply the learned formula of standard deviation to problems isomorphic to the ones discussed during explicit instruction. Conceptual knowledge refers to a deep understanding of the taught concept and its components (Rittle-Johnson and Schneider 2014). Conceptual knowledge can be reflected in principle-based reasoning or in the ability to connect different representations. For example, students who learned standard deviation may be asked to predict and explain what will happen to the standard deviation if all numbers in the sample were increased by a constant. Transfer refers to the ability to adapt the learned concept to a new situation or a different type of problem. For example, students who learned standard deviation may be asked to answer a problem that requires a more advanced topic such as normalization (for example, in order to be able to compare apples and oranges; Barnett and Ceci 2002). In column 6, we describe the *population* and in column 7 the learning *topic*.

The two rightmost columns refer to two specific variations in the implementation of the problem-solving phase and the explicit instruction phase: whether the problem-solving

Table 1 Overview of recent findings on PS-I approaches concerning learning outcomes

No.	Reference	PS-I group	Control group (CG)	Learning outcomes and effect sizes	Population	Topic	Problem with contrasting cases	Building instruction on student solutions
1	Belenky and Nokes-Malach (2012)	Problem solving followed By instruction (+ short practice)	Instruction followed by problem solving	Transfer: PS-I = CG for high mastery orientation, PS-I > CG for low mastery orientation	104 undergraduates, USA	Variability	Yes	No
2	DeCaro and Rittle-Johnson (2012)	Problem solving with accuracy feedback followed by instruction (+ short practice)	Instruction followed by problem solving with accuracy feedback	Procedural skills: PS-I = CG; Conceptual knowledge: PS-I > CG, $\eta_p^2 = .03$	159 2nd–4th graders, USA	Equivalence	No (but no rich problems)	No
3	Fyfe, DeCaro, and Rittle-Johnson (2014)	Problem solving with accuracy feedback followed by instruction	Instruction followed by problem solving with accuracy feedback	Procedural skills: PS-I < CG, $\eta_p^2 = .04$; conceptual knowledge: PS-I = CG (overall; for one subscale PS-I < CG, $\eta_p^2 = .04$)	122 2nd–3rd graders, USA	Equivalence	No (but no rich problems)	No
4	Glogger-Frey et al. (2015)	Inventing evaluation criteria (study 1) or problem solving (study 2) followed by instruction	Studying worked examples followed by instruction	Study 1: transfer: PS-I < CG, $d = -0.72$ study 2: procedural skills and near transfer: PS-I = CG; far transfer: PS-I < CG, $d = -0.71$	2 studies: 42 pre-service teacher and 40 8th graders, Germany	Study 1: evaluation of learning strategies Study 2: density	Yes	No
5	Hsu et al. (2015)	Problem solving followed by worked example	Worked example followed by problem solving	Study 1: conceptual knowledge: PS-I < CG, $\eta_p^2 = .03–.04$ Study 2: conceptual knowledge: PS-I (<) CG, $\eta_p^2 = .02–.03$	2 studies: 260 10th–12th graders and 129 10th graders, Taiwan	Physics (study 1: motion; study 2: collision)	No	No

Table 1 (continued)

No.	Reference	PS-I group	Control group (CG)	Learning outcomes and effect sizes	Population	Topic	Problem with contrasting cases	Building instruction on student solutions
6	Kapur (2010)	Problem solving followed by instruction (+ short practice)	Instruction followed by problem solving	Procedural skills: PS-I > CG, ES = .42; conceptual knowledge: PS-I > CG, ES = .98; transfer: PS-I > CG	75 7th graders, Singapore	Average speed	No	Yes
7	Kapur (2011)	Problem solving followed by instruction (+ short practice)	(a) Instruction followed by problem solving (b) Scaffolded problem solving followed by instruction	Procedural skills: PS-I > CG ($a < b$), $\eta_p^2 = .10$; conceptual knowledge (incl. representational flexibility): PS-I > CG ($a < b$), $\eta_p^2 = .02$–$.10$	109 7th graders, Singapore	Average speed	No	Yes
8	Kapur (2012)	Problem solving followed by instruction (+ short practice)	Instruction followed by problem solving	Procedural skills: PS-I = CG; conceptual knowledge: PS-I > CG, $\eta_p^2 = .44$; transfer: PS-I > CG, $\eta_p^2 = .23$	133 9th graders, Singapore	Variability	No	Yes
9	Kapur (2014b)	Problem solving followed by instruction (+ short practice)	Instruction followed by problem solving (additional condition in study 2: evaluation of erroneous solutions followed by instruction)	Procedural skills: PS-I = CG; conceptual knowledge: PS-I > CG, $d = 2.00/2.25$; transfer: PS-I > CG, $d = 1.52/1.29$	2 studies: 75 and 111 9th graders, India	Variability	No (only in instruction phase)	No
10	Kapur and Bielaczyc (2011)	Problem solving followed by	3 studies: (1) Instruction followed by problem solving	Procedural skills: PS-I = CG; conceptual	3 studies: 74, 54, and 57 9th graders, Singapore	Variability	No	Yes

Table 1 (continued)

No.	Reference	PS-I group	Control group (CG)	Learning outcomes and effect sizes	Population	Topic	Problem with contrasting cases	Building instruction on student solutions
		instruction (+ short practice	(2) Evaluation of predesigned student solutions followed by instruction (3) Strong instruction (explanation on solution components) followed by problem solving	knowledge: PS-I > CG, $\eta_p^2 = .31/.19/.13$; near transfer: study 2: PS-I > CG $\eta_p^2 = .08$, study 3: PS-I = CG				
11	Kapur and Bielaczyc (2012)	Problem solving followed by instruction (+ short practice)	Instruction followed by problem solving	Procedural skills: study 1: PS-I > CG, study 2 and 3: PS-I = CG Conceptual knowledge (incl. representational flexibility): PS-I > CG	3 studies: 75, 114, and 113 7th graders, Singapore	Average speed	No	Yes
12	Loehr et al. (2014)	Problem solving followed by instruction (in study 2 additionally followed by self-checking initial solutions)	Instruction followed by problem solving (in study 2 additionally followed by self-checking solutions)	Study 1: procedural skills: PS-I = CG; conceptual knowledge: PS-I < CG, $\eta_p^2 = .02$ Study 2: procedural skills: PS-I > CG, conceptual knowledge: PS-I = CG	2 studies: 41 and 47 2nd graders, USA	Equivalence	No (but no rich problems)	No
13	Loibl and Rummel (2014a)	(a) Problem solving followed by instruction	(a) Instruction followed by problem solving	Procedural skills: study 1: PS-I_b < CG_a = CG_b,	2 studies: 98 and 240 10th graders, Germany	Variability	No	(a) No (b) Yes

Table 1 (continued)

No.	Reference	PS-I group	Control group (CG)	Learning outcomes and effect sizes	Population	Topic	Problem with contrasting cases	Building instruction on student solutions
		without student solutions (only implemented in study 2) (b) Problem solving followed by instruction building on student solutions	(b) Instruction building on student solutions followed by problem solving	$\eta_p{}^2 = .09$; study 2: PS-I_a = PS − I_b = -CGa = CG_b Conceptual knowledge: study 1: PS-I_b = CG_b > CG_a, $\eta_p{}^2 = .26$; study 2: PS-I_b > CG_b > PS--Ia = CG_a, $\eta_p{}^2 = .08$				
14	Loibl and Rummel (2014b)	(a) Problem solving followed by instruction (b) Problem solving with contrasting cases followed by instruction	Instruction followed by problem solving	Procedural skills: study 1: PS-I_a = PS − I_b < CG, $\eta_p{}^2 = .03$, study 2: PS-I_a = PS-I_b = CG; conceptual knowledge: PS-I_a = PS-I_b > CG, $\eta_p{}^2 = .30$	2 studies: 104 and 175 10th graders, Germany	Variability	(a) No (b) Yes	Yes
15	Matlen and Klahr (2013)	Designing experiments followed by explanation	Predesigned experiments and explanation followed by designing experiments	Procedural skills and transfer: PS-I = CG	57 3rd graders, USA	Control of variable strategy	No	No
16	Roll et al. (2009)	Problem solving followed by instruction and practice	Ranking contrasting cases followed by instruction and practice	Procedural skills and transfer with learning resource: PS-I = CG Transfer without learning resource: PS-I > CG	105 7th graders, USA	Central		
	tendency/graphing and variability	Yes	No					
17	Roll et al. (2011)	Problem solving followed by	Evaluation of predesigned student	Conceptual knowledge and debugging items:	92 7th graders, USA	Variability	Yes	Yes

Table 1 (continued)

No.	Reference	PS-I group	Control group (CG)	Learning outcomes and effect sizes	Population	Topic	Problem with contrasting cases	Building instruction on student solutions
		instruction and practice	solutions followed by instruction and practice	PS-I > CG; transfer: PS-I = CG				
18	Schwartz and Bransford (1998)	Analyzing contrasting cases followed by instruction	3 studies: (1) Reading text followed by lecture (2) Summarizing text (3a) Summarizing text followed by lecture (b) Analyzing contrasting cases twice (without lecture)	Transfer: PS-I > CG	3 studies: 24, 18, and 36 undergraduates, USA	schema concepts and encoding concepts	Yes	No
19	Schwartz et al. (2011)	Problem solving followed by instruction (+ short practice)	Instruction followed by problem solving	Procedural skills: PS-I = CG; transfer: PS-I > CG, d = .53–.66	2 studies: 128 and 120 8th graders, USA	density	Yes	No
20	Schwartz and Martin (2004)	Problem solving followed by instruction (+ short practice)	Instruction followed by problem solving	Transfer with learning resource: PS-I > CG; transfer without learning resource: PS-I = CGs	2 studies: 95 and 102 9th graders, USA	Variability	Yes	No

[a] The very same data reported in this paper is also reported in DeCaro, DeCaro, and Rittle-Johnson (2015). Therefore, we only included one of these papers

Table 2 Examples illustrating variants of PS-I

Phase	Variant 1	Variant 2
Problem solving	*Rich problem* Which soccer player scores more consistently? • Player A: 14, 9, 14, 10, 15, 11, 15, 11, 16, 12, 16, 12, 17, 13, 17, 13, 18, 14, 19, 14 • Player B: 13, 9, 16, 14, 10, 11, 13, 14, 15, 19, 14, 12, 15, 14, 17, 13, 14, 18, 14, 15 • Player C: 13, 18, 15, 10, 16, 10, 17, 10, 12, 14, 19, 14, 18, 9, 10, 18, 11, 10, 18, 18	*Contrasting cases* Which soccer player scores more consistently? Contrasting cases #1: • Player A: 9, 10, 10, 11 • Player B: 5, 10, 10, 15 Contrasting cases #2: • Player B: 5, 10, 10, 15 • Player C: 5, 5, 15, 15
Explicit instruction	*Building on student solutions* Many of you calculated the sum of the deviation from one year to the next. Some used the deviations as calculated and summed them up, others took absolute values. This led to different results. What is the benefit of one solution method or the other? … One solution that experts often use is standard deviation: $SD = \sqrt{\frac{(x_i - mean)^2}{N}}$	*Canonical solution only* One solution that experts often use is standard deviation: $SD = \sqrt{\frac{(x_i - mean)^2}{N}}$

phase includes contrasting cases and whether the instructional phase builds on student solutions.

Table 2 offers greater detail and examples for the variations that are introduced in Table 1. The problem-solving phase has two main variants, which we introduce in column 8 of Table 1 (*problem with contrasting cases*). In some PS-I designs (often termed Productive Failure), data is presented as part of a rich cover story that does not highlight the deep features of the topic and for which the solution cannot be intuitively guessed (e.g., Kapur 2010, 2011, 2012; Loibl and Rummel 2014a).

Table 3 Classification of studies and summary of learning outcomes

Contrasting cases	*Building on student solutions*	*Procedural assessment*	*Conceptual assessment*	*Transfer assessment*
no	yes	+: 6, 7, (11) =: 8, 10, (11), (13b), (14a) -: (13b), (14a)	+: 6, 7, 8, 10, 11, 13b, 14a =: -:	+: 6, 8, (10) =: (10) -:
yes	no	+: =: 4, 16, 19 -:	+: =: -:	+: (1), (16), 18, 19, (20) =: (1), (16), (20) -: 4
yes	yes	+: =: 14b -:	+: 14b, 17 =: -:	+: =: 17 -:
no	no	+: (12) =: 2, 9, (12), 13a, 15 -: 3,	+: 2, 9 =: 3, (12), 13a -: 5, (12)	+: 9 =: 15 -:

Note. +, =, - indicate positive, null, or negative effect respectively for PS-I as compared to the control condition. Numbers refer to the row numbers in Table 1. Numbers in parentheses refer to papers with multiple studies or assessments that found mixed results.

By deep structure, we refer to the underlying principles of the domain (Chi et al. 1981). In the rich story example displayed in the left column of Table 2, each dataset includes 20 values, and datasets differ from each other across a variety of features. Thus, students' attention is not explicitly directed to the deep features. In contrast, in other PS-I designs (often termed Invention), the relevant information is given to students in the form of contrasting cases (e.g., Belensky and Nokes-Malach 2012; Glogger-Frey et al. 2015; Roll et al. 2009). Contrasting cases consist of small sets of data, examples, or strategies presented side-by-side (e.g., Schwartz and Martin 2004; Schwartz and Bransford 1998). These minimal pairs differ in one deep feature at a time ceteris paribus, thereby highlighting the target features. In the example provided in the right column of Table 2, the datasets of player A and player B differ with regard to the range, while other features (e.g., mean, number of data points) are held constant. The next pair of datasets addresses another feature: player B and C have the same mean and range but different distribution of the data points.[1] Regardless of potential differences in the conceptualizations used in the papers, in our overview, contrasting cases are merely classified as such, if the cases differ in only one feature at a time to make the deep features salient and if they are introduced during the problem-solving phase to guide students' thinking.

During the explicit instruction phase, the canonical solution is explained. This phase has two main variants, which we introduce in column 9 of Table 1 (*building instruction on student solutions*). In some implementations (e.g., Belenky and Nokes-Malach 2012; Glogger-Frey et al. 2015; Roll et al. 2009; Schwartz and Martin 2004), students are given the canonical solution without referring back to student solutions. In other studies (e.g., Kapur 2010, 2011, 2012; Loibl and Rummel 2014b; Roll et al. 2011), instruction includes an additional component in that the teacher builds on typical student-generated solutions *before* explaining the structurally relevant components of the canonical solution. Thus, students have the opportunity to compare and contrast failed or suboptimal solutions against the canonical one (Kapur and Bielaczyc 2012). A brief example of how two student solutions can be compared is provided in Table 2. The comparison in the example aims to help students realize that taking absolute values is a valid approach for measuring distances between data points.

In summary, each of the two phases of PS-I has two main variants, resulting in four different instantiations of PS-I. As far as we can tell from our analysis of the papers, the differences described above are rarely addressed explicitly and have not yet been investigated systematically. The lack of attention to these variations seems surprising given that seemingly small modifications to an instructional design can trigger different cognitive mechanisms and may thereby have a major impact on learning outcomes.

Table 3 summarizes the patterns of learning outcomes that we found in the analyzed papers (see Table 1) sorted by differences in the specific PS-I design. It is important to emphasize that we were aiming to identify patterns in the data based on our descriptive analyses of the papers. We were, however, not aiming to perform a statistical meta-analysis due to the rather small number of studies.

The first row in Table 3 includes studies in which a rich problem (i.e., without contrasting cases) is followed by instruction that builds on student solutions. In all studies in this class, the PS-I conditions outperformed the control conditions on conceptual knowledge and/or measures of transfer, whereas the results regarding procedural knowledge were mixed.

The second row includes studies that introduce contrasting cases in the problem-solving phase, followed by instruction without student solutions. Also in most of these studies, except

[1] For a discussion on the different processes triggered by contrasting cases in comparison to rich problems, see Loibl and Rummel (2014b).

the ones in no. 4, the PS-I conditions outperformed the control conditions on transfer. In the studies in no. 4 (Glogger-Frey et al. 2015), the PS-I condition was outperformed by the control condition on transfer items. No study in this row found an effect of PS-I on procedural knowledge.

A third class of studies, described in the third row, includes studies that combine the use of contrasting cases during the problem-solving phase with instruction that builds on student solutions. These studies found beneficial effects on conceptual knowledge. One of these studies (no. 14: Loibl and Rummel 2014b) also found that the combination of both elements (contrasting cases during the problem-solving phase and instruction building on student solutions) is not necessarily better than building upon student solutions without the use of contrasting cases.

While the results of these three classes of PS-I designs show a strong overall trend in favor of PS-I with regard to conceptual knowledge and transfer (cf. Table 3, rows 1–3: with one exception all studies in these classes show positive effects in at least one of the two assessments), it is important to emphasize that other factors beyond the sequence of problem solving and instruction may account for this effect: It is possible that contrasting cases during problem solving and/or instruction building on student solutions were implemented in the PS-I condition but not in the control condition. This potential confound does not allow to conclude whether learning effects are due to the PS-I structure or whether they stem from the additional design elements (contrasting cases and/or building instruction on student solutions). To account for the confound (different design elements plus different sequence), we analyzed the materials and procedures reported in the papers to evaluate whether the studies provided the same problem description and data during the problem-solving phase (with or without contrasting cases) and the same content and prompts during the explicit instruction phase (with or without student solutions) across conditions.[2] Indeed, eight studies confounded design elements with instructional sequence (nos. 6, 7, 8, 10, 11, 14, 18, 20: Kapur 2010, 2011, 2012; Kapur and Bielaczyc 2011, 2012; Loibl and Rummel 2014b; Schwartz and Bransford 1998; Schwartz and Martin 2004). Six studies used identical materials across conditions (nos. 1, 4, 13, 16, 17, 19: Belenky and Nokes-Malach 2012; Glogger et al. 2015; Loibl and Rummel 2014a; Roll et al. 2009, 2011; Schwartz et al. 2011[3]). Given the high number of confounded studies, it is possible that the effect reflected in Table 3 is due to better learning resources (providing contrasting cases and/or building instruction on student solutions) being available to PS-I participants. However, five of the six papers that used the same materials across conditions show the same pattern of beneficial effects on conceptual knowledge and/or transfer. This (albeit limited) set of studies suggests that the advantage of PS-I designs over I-PS conditions probably does not stem from providing contrasting cases during problem solving or building instruction on student solutions *alone* and that the PS-I sequence at least contributes to the beneficial outcome.

Finally, row 4 includes studies *without* contrasting cases during the problem-solving phase and with instruction that does *not* build on student solutions. Notably, all of these studies used the same materials in both conditions. There is no clear trend for studies in this category. The

[2] Naturally, some wording is expected to differ by condition (e.g., “invent a formula/strategy for the following problem” vs. “solve the problem using this formula/strategy”), which is not counted as confound.

[3] In the studies in no. 19, the instruction and the worksheets with contrasting cases were held constant across conditions. However, students in the I-PS condition received an additional reminder of the formula and a worked example prior to solving each worksheet, whereas students in the PS-I condition were told what invention means and what they need to invent prior to their first problem-solving attempt.

mixed results in this class of studies could be due to statistical noise. Alternatively, issues with the implementation of conditions may account for the lack of trend. In one study (no. 15: Matlen and Klahr 2013), the pretest included an invention activity for all conditions. This diffusion of treatment may explain the missing effect (however, Toth et al. 2000 suggest that there was no pretest × treatment interaction; for a more detailed discussion of this issue, see Kapur 2012). In the studies that found beneficial effects for PS-I, students worked on rather short problems (no. 2: DeCaro and Rittle-Johnson 2012) or contrasting cases were discussed during the explicit instruction phase (no. 9: Kapur 2014b). In contrast, studies with rich problems found no or negative effects when neither contrasting cases during the problem-solving phase nor instruction that builds on student solutions were implemented (nos. 3, 5, 12, 13a, and 15: Fyfe et al. 2014; Hsu et al. 2015; Loehr et al. 2014; Loibl and Rummel 2014a; Matlen and Klahr 2013). This result sheds some light on the preceding discussion of design elements: Without contrasting cases during problem solving and without instruction building on student solutions, there seem to be no clear benefits for the PS-I structure.

This finding matches the outcome of a large number of worked example studies, which often do not include contrasting cases or instruction building on student solutions: Many studies of worked examples find beneficial effects for example-problem pairs (cf. I-PS) in comparison to problem-example pairs (cf., PS-I; e.g., Hsu et al. 2015; Leppink et al. 2014; Reisslein et al. 2006; Van Gog et al. 2011). Reviewing the worked example literature in detail is beyond the scope of this paper, but recent review papers on worked examples (e.g., Renkl 2014; Van Gog and Rummel 2010) draw a clear picture, favoring example-problem pairs over problem-example pairs. Taking evidence from the two bodies of research (PS-I and worked examples literature) together, it seems that I-PS may outperform PS-I when the PS-I design is such that students neither work with contrasting cases during problem solving nor are student solutions discussed during instruction.

Given that the majority of PS-I studies were conducted in mathematics, it is of interest to look at whether results extend to other domains as well. One study that found negative effects, despite using contrasting cases, targeted a psychology topic (no. 4: Glogger-Frey et al. 2015). Thus, one could argue that the effectiveness of PS-I approaches might depend on the domain. However, another study that also targeted a psychology topic found beneficial effects in comparison to several control conditions (no. 18: Schwartz and Bransford 1998). Therefore, the domain alone does not seem to tell the whole story. Due to the limited number of studies with topics outside math-related domains, the effect of the domain needs to be further investigated in future studies.

In summary, in our review, we found evidence suggesting that PS-I is better suited to promote conceptual knowledge and transfer but has no clear benefits for procedural knowledge. This outcome supports the notion that different learning goals (e.g., procedural fluency vs. conceptual understanding) require different instructional means (cf. Kalyuga and Singh 2015; Koedinger et al. 2012). In addition, the positive effects of PS-I on conceptual knowledge and transfer are limited to PS-I designs where students work with contrasting cases during problem solving and/or that involve instruction building on student solutions (cf. rows 1–3 in Table 3). Without these design elements, the effects of PS-I are less favorable or can even be negative for learning (cf. row 4 from Table 3). In the following section, we discuss how this pattern of results may be explained with reference to cognitive mechanisms triggered by the different designs of PS-I.

Identifying Mechanisms that Make PS-I Productive for Learning

The above review demonstrated that not all implementations of PS-I are equally beneficial. Against this background, we aim to explain how specific features of the problem solving phase and the instruction phase in PS-I foster learning. Kapur and Bielaczyc (2012) proposed several mechanisms to explain the effectiveness of one particular type of PS-I implementation, Productive Failure. In this section we build on their work and expand it to include additional forms of PS-I. We ground the discussion of the mechanisms in general learning theories and support each hypothesized mechanism using available data from PS-I studies. Specifically, for each mechanism, we (a) define the relevant mechanism, (b) describe how the design of PS-I inherently facilitates the mechanism, and (c) discuss how the mechanism explains the abovementioned patterns in learning outcomes. Finally, we make the case that the synergy between these mechanisms, as facilitated by specific PS-I manipulations, further supports learning.

As most studies included in our review (see Table 1) were carried out in the domain of mathematics, our conclusions should be taken in light of this limitation. Still, when possible, we included examples from other disciplines. By suggesting domain-general mechanisms and highlighting their role in PS-I, we hope to lay the foundations for additional work on PS-I in a variety of domains.

Prior Knowledge

The first mechanism accounting for the effectiveness of PS-I for learning proposed by Kapur and Bielaczyc (2012) is "activation and differentiation of prior knowledge in relation to the targeted concepts" (p. 49). Prior knowledge is stored in schemas (e.g., Linn 1995; for an overview, see Anderson 1983). It has been well established that learning takes place when these schemas are modified to support integration of new knowledge (e.g., Sweller et al. 1998). Modification and integration requires that existing schemas are activated (e.g., Sweller 1988). Encouraging students to activate prior knowledge has been found to be effective for learning in a variety of instructional designs (i.e., not only within PS-I). For example, Schmidt et al. showed that prior knowledge activation by means of small-group discussion enhances subsequent text processing (Schmidt et al. 1989). They explain their results by suggesting that new information is more easily integrated with prior knowledge components once prior knowledge is activated.

The activation of prior knowledge is inherent in PS-I: Students typically access their prior knowledge and ideas in order to invent solutions to the new problem (Kapur and Bielaczyc 2012; Schwartz and Bransford 1998; Schwartz and Martin 2004; Schwartz et al. 2007). In order to be effective, the problem provided in the problem-solving phase should be one for which students have some prior knowledge and ideas. Thus, knowledge activation in this setting depends on a careful design of the problem. Notably, because students have not yet received formal instruction on the solution, they are likely to activate their intuitive ideas. Intuitive ideas are informal understandings that students gain based on their real-life experiences (cf. Nathan et al. 2010). While prior knowledge usually refers to canonical knowledge, intuitive ideas are not yet consolidated and may therefore be partial or erroneous. Nevertheless, it has been argued that students' intuitive ideas provide additional resources for future learning (Kapur and Bielaczyc 2011).

The PS-I literature itself provides direct support for the prior knowledge mechanism. Several studies (Kapur 2014a, 2014b; Kapur and Bielaczyc 2011; Roll et al. 2011) compared conditions where students attempt to solve problems by themselves to conditions where students evaluated pre-designed solutions. In these studies, students in the PS-I condition *and* the evaluation (control) condition worked with erroneous or incomplete solution approaches (either self-generated or given) prior to instruction. All studies found benefits for students generating their own solution ideas in comparison to evaluating ideas proposed by others. As generating own ideas focuses on activating one's very own prior knowledge, this finding provides support for the proposed mechanism.

Can this mechanism alone explain the PS-I findings? Students activate prior knowledge in all PS-I manipulations, including the less effective ones (cf. row 4 in Table 3). Also, prior knowledge activation should support all types of learning, including acquisition of procedural knowledge (which is not typically benefited by PS-I). Thus, this mechanism does not provide a sufficient explanation in and by itself when explaining the differentiated effects of PS-I (i.e., positive effect on conceptual knowledge and transfer but not necessarily on procedural knowledge; only effective with contrasting cases and/or building instruction on student solutions).

Indeed, not all forms of prior knowledge activation are productive for learning as pointed out by Schwartz et al. (2007). In one of their examples, during instruction on sampling methods, students activated prior knowledge of the concept of fairness. After activating this concept, students perceived fairness as a sufficient criterion to judge sampling methods and they perceived all fair samples to be good ones, even when they were biased (e.g., self-selection). Literature on conceptual change further discusses how certain knowledge can hamper the acquisition of new knowledge, if existing and new knowledge components are not compatible (e.g., Chi et al. 1994; Prediger 2008; Vosniadou and Verschaffel 2004). To give just one example, knowing that multiplication of natural numbers leads to a result which is larger than each number hampers the understanding of multiplication with fractions (Prediger 2008).

To conclude, prior knowledge activation allows students to better integrate the following instruction with existing schemas. However, these schemas can be partial, naïve, or even erroneous. As will be addressed in the next sub-section, the activation of prior knowledge is a necessary prerequisite for becoming aware of these gaps in the existing schemas, because students can only notice where their knowledge is insufficient, if they bring their existing knowledge to bear on a problem.

Awareness of Knowledge Gaps

Research has shown that students process canonical solutions more deeply when they are aware of their own impasses and knowledge gaps (VanLehn et al. 2003). According to VanLehn's (1999) impasse-repair-reflect process, once the learner reaches an impasse, she or he applies strategies to repair the impasse. A similar process is described in Chi's (2000) work on repairing mental models. In her imperfect mental model view, Chi states that learners' initial mental models often differ from normative models. Before students can repair their models, they need to become aware of the flaws in them. The importance of awareness of gaps and limitations of one's knowledge is supported by various findings: Heckler et al. showed that exposing students to the fact that salient factors cannot explain a certain phenomenon (in our example average cannot explain variability, in their example a certain computer chip component cannot be installed in a specific kitchen appliance) prompts students to look for alternative explanations (Heckler et al. 2008). Similarly, Acuña et al. showed that warning of possible errors before presenting an instructional explanation can foster learning as the warning helps students to become aware of their own knowledge gaps (Acuña et al. 2010; Sánchez et al. 2009). Siegler (1983) argues that students become motivated to learn new information when they realize that their initial ideas make wrong predictions.

Knowledge gaps are better identified when they are experienced (Eva and Regehr 2011; Needham and Begg 1991). In PS-I studies, students report higher perceived knowledge gaps when attempting to invent prior to instruction in comparison to students who receive instruction prior to solving problems (Glogger-Frey et al. 2015; Loibl and Rummel 2014a). Note that this result was found for a mathematics topic (standard deviation, Loibl and Rummel 2014a) as well as a non-mathematics topic (evaluation of learning strategies, Glogger-Frey et al. 2015).

Failing to solve a problem is not effective if students are not aware of their failure. PS-I manipulations differ in the way and extent in which they foster this awareness. At times, students simply realize that their initial ideas do not offer complete solutions during problem solving. However, studies that provide their data in the form of rich problems may not lend themselves to easy self-evaluation. In these cases, subsequent instruction that *builds on common erroneous solutions* before explaining the canonical solution can help students realize the limitations of their knowledge and increase their curiosity to receive more information (Loibl and Rummel 2014a). Alternatively, some studies use *contrasting cases* that can be intuitively ranked. In the example presented in the right column of Table 2, students can intuitively see that player A scores more consistently than player B and player B scores more consistently than player C. Students can use their intuitive ranking of the cases to evaluate their invented solution and extract grounded feedback (Nathan 1998; Roll et al. 2014). In the example in Table 2, player B and player C have the same range. When comparing this outcome to their intuitive ranking, students may realize that range does not solve the problem. Indeed, Roll et al. showed that directing students to make predictions using contrasting cases resulted in students identifying their errors and revising their solutions more often than students who had access to the same contrasting cases but were not prompted to make these predictions (Holmes et al. 2014; Roll et al. 2012).

To summarize, identifying one's own knowledge gaps is necessary to initiate modifications of existing (partial, naïve, or erroneous) schemas. PS-I approaches can support this in two ways: either using contrasting cases during the problem-solving phase that support students in intuitively evaluating their solutions or by reviewing student solutions at the beginning of the instruction phase. Implementations of PS-I that do not include any of these design features are bound to be less effective in promoting students awareness of their knowledge gaps and thus less effective in supporting mental model repairs and learning.

Recognition of Deep Features

Another mechanism proposed by Kapur and Bielaczyc (2012) to explain the effectiveness of Productive Failure (i.e., one particular implementation of PS-I) concerns the deep features of the target knowledge. They suggest that Productive Failure helps students identify, explain, and organize deep features of the target knowledge. The impact of noticing and encoding deep features has been well established in the literature on expertise (e.g., Chi et al. 1981) and on problem solving (e.g., Duncker 1945; Quilici and Mayer 1996, 2002; Wertheimer 1959). Here we review literature that gives evidence for contribution of PS-I to the recognition and encoding of deep features. Interestingly, the variations of PS-I address deep features in different ways:

Identifying deep features is inherent to developing solutions to problems with *contrasting cases*, because the cases vary in one feature at a time (e.g., Schwartz and Martin 2004). Indeed, Loibl and Rummel (2014b) showed that students address more deep features in their solutions if provided with contrasting cases in comparison to rich problems. Similarly, Roll et al. used contrasting cases to direct students' attention to some deep features, but not others, during the problem-solving phase (Roll et al. 2012). Students who were explicitly directed to examine the contrasting cases spontaneously explained the deep features three times as often as students in the control condition (despite also having contrasting cases available), who rather self-explained features that were not relevant for the solution.

Another way of highlighting deep features in PS-I is to compare non-canonical *student solutions* to each other and to the canonical solution during subsequent instruction (e.g., Kapur 2012; Loibl and Rummel 2014a). Explaining why erroneous solutions are incorrect has been found to be beneficial for learning, in some cases even more than explaining correct solutions (Booth et al. 2013). Furthermore, the comparison supports students in detecting differences between their own prior ideas and the canonical solution (Loibl and Rummel 2014a). More precisely, through comparing them to other students' solution and to the canonical solution, students experience how their solution approaches fail to address one or more important aspects of the problem (e.g., diSessa et al. 1991). This process guides students' attention to the deep features addressed by the canonical solution (cf. Durkin and Rittle-Johnson 2012). Similarly to the use of contrasting cases, the comparison of student solutions should explain one deep feature at a time, leading to a meaningful understanding of these features (see example in Table 2 as well as the discussion in Loibl and Rummel 2014a).

Introducing the canonical solution after highlighting the deep features (by providing contrasting cases during problem solving or by building instruction on student solutions) enables students to organize the target knowledge by its deep features. In contrast, if the initial problem does not include contrasting cases and the instruction does not build on student solution, it seems unlikely that students notice and elaborate on the deep features on their own—a possible explanation for the limited effectiveness of these PS-I manipulations (cf. Loibl and Rummel 2014a).

When students encode their knowledge by its deep features, it should be easier for them to adapt it to new situations. Therefore, we would expect benefits on items requiring flexibility (including procedural flexibility items, cf. Rittle-Johnson and Star 2009) and transfer. In contrast, we would not expect effects on conventional procedural knowledge items requiring the mere application of a learned procedure as a whole. Indeed, the findings on PS-I show positive effects on transfer items, if contrasting cases were provided during problem solving or the instruction built on student solution. PS-I results on procedural knowledge are mixed (ranging from positive to negative effects). This might be due to different operationalization of procedural knowledge with procedural flexibility (which require adapting the learned procedure) or conventional items (which require applying the learned procedure as is). An analysis of the procedural posttest items administered by the different research groups could shed light on this hypothesis. However, this task is beyond the scope of the current paper.

Is the highlighting of deep features sufficient to explain the PS-I effect? Deep features can also explicitly be addressed during instruction without previous problem solving (e.g., Kapur and Bielaczyc 2011; Loibl and Rummel 2014a). However, in these two examples, instruction that addressed the deep features followed by problem solving was still less effective than

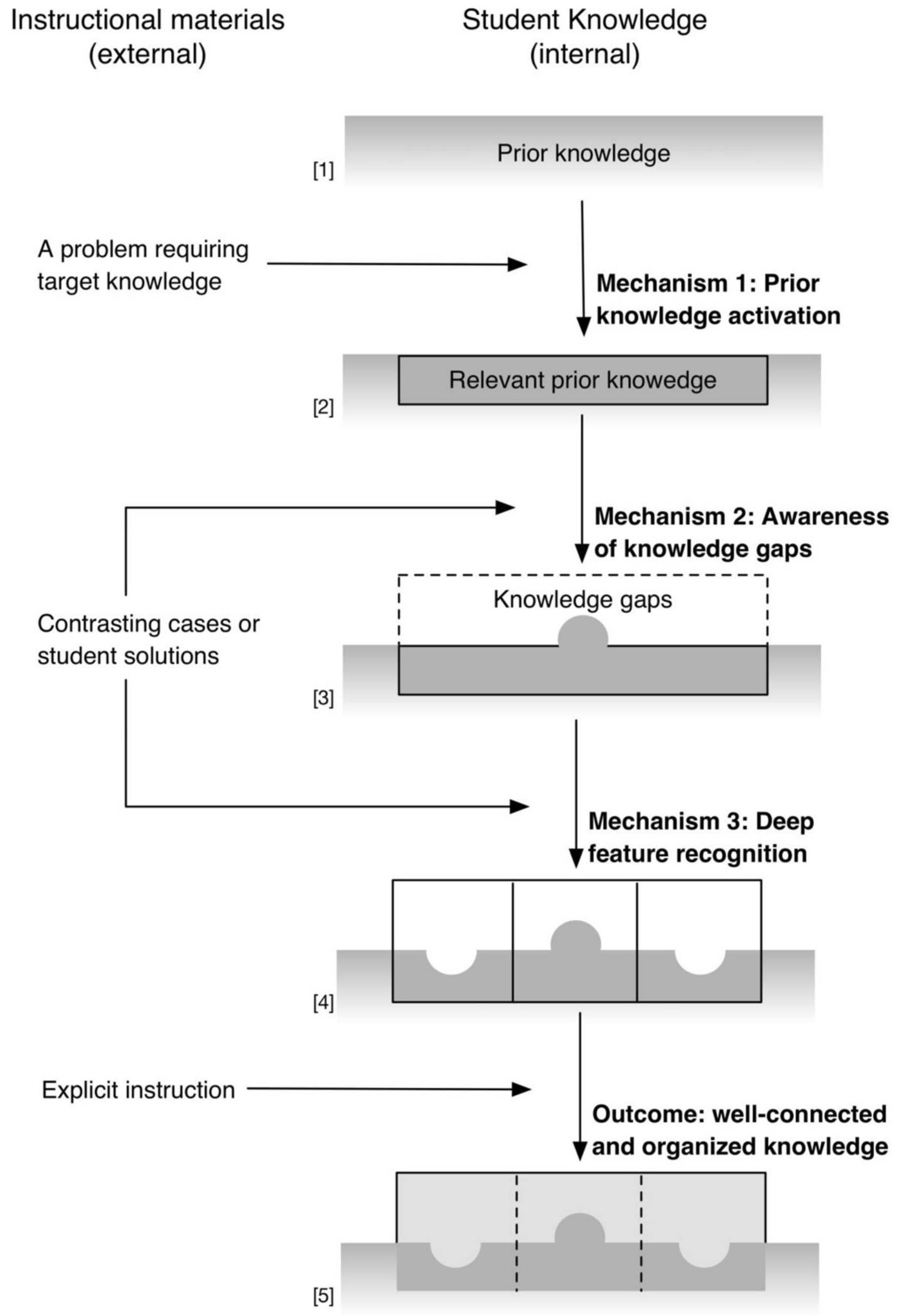

Fig. 1 Knowledge acquisition in PS-I

problem solving followed by identical instruction. Thus, it seems that highlighting the deep features requires activation of prior knowledge and raising awareness of knowledge gaps in order to be most effective.

Towards a Theory of Problem Solving Followed by Instruction

Above we have argued that PS-I approaches prompt students to activate their prior knowledge, become aware of their specific knowledge gaps, and encode knowledge in relation to deep features. Next we investigate how these mechanisms complement each other as we work towards a theory of PS-I.

Figure 1 illustrates the interaction between the three suggested mechanisms. Students enter a learning situation with prior knowledge (visualized as a shaded grey area, [1]). The students are then confronted with a problem targeting a new topic (e.g., standard deviation). They already know relevant concepts, such as average or range, which they *activate* in their attempt to solve the problem ([2]). While some students may be satisfied with their initial solution (e.g., average or range), others may realize the limitations of their existing knowledge and become aware of their *knowledge gaps* (white part of the box in [3]). This awareness of knowledge gaps is more likely if the problem includes contrasting cases or if the teacher builds on student solutions prior to introducing the canonical solution because these design elements offer detailed situational feedback on students' initial attempts. By comparing contrasting cases during the problem-solving phase or discussing student solutions during the instruction phase, students are encouraged to recognize the *deep features* of the problem (components of the box in [4]). For example, when learning standard deviation, by comparing two sets with similar extreme values but different distributions, students can realize that their solution method should take all values into account. Overall, by the time students receive explicit instruction on the canonical solution, they have activated their prior knowledge, realized its limitations, and identified a set of deep features that needs to be addressed. Subsequently, when eventually receiving explicit instruction about the canonical solution, students are likely to perceive the presented solution as a set of functional elements that fit in their knowledge gaps and address the identified deep features (grey box in [5]). For example, students understand the sequence and the goals of mathematical manipulations that are applied to determine standard deviation: calculating deviations from a fixed reference point to avoid sequencing effects, squaring and adding them to include all values and to avoid that positive and negative deviations cancel each other out, and dividing by the number of data points to control for sample size.

We argue that there is a cascading effect: To encode the target knowledge according to the deep features, students first need to identify the features. To identify the features, they first need to realize what their own knowledge fails to achieve. To do that, they first need to activate their prior knowledge. We suggest that this cascading effect leads to well-connected and well-organized knowledge. The high degree of connectedness with prior knowledge (i.e., with what students both did and did not know) and students' ability to decompose the knowledge into its deep features allow students to apply the learned knowledge to new situations and to adapt it accordingly. Therefore, the expected learning outcomes based on the explanations above match the general findings on PS-I described earlier positive effects on transfer and conceptual items but only if PS-I is implemented with contrasting cases during problem solving or instruction building on student solutions.

Consolidating Contradictory Results

The cascading effect of the three proposed mechanisms can explain most patterns that were identified in Table 3. Almost all studies that used contrasting cases during the problem-solving phase and/or built on student solutions during the instruction phase showed benefits for PS-I on conceptual or transfer assessments compared with a control condition. The only paper with differing outcomes is the paper by Glogger-Frey et al. (2015, see Table 1 no. 4). In their studies, studying worked examples that display and explain the relevant deep features was more beneficial than inventing own methods with contrasting cases. It is hard to explain why this specific paper shows different outcomes. It might be a false negative. However, there may also be specific characteristics in the design of the studies that led to the different outcomes. One potential explanation is that the worked examples were better designed to highlight and explain the deep features as the contrasting cases. An in-depth analysis of the used contrasting cases and worked examples and their potential to make the features salient would be needed to answer this question. Either way, these studies demonstrate an alternative route (using worked examples) to trigger similar mechanisms, namely, encoding the solution in relation to its deep features. Indeed, well-designed worked examples provide the information in meaningful building blocks that make the deep features salient (e.g., Renkl 2005).

We did not find beneficial effects for PS-I in studies that did not include contrasting cases in the problem-solving phase nor built on student solutions in the instruction phase (bottom row of Table 3). In these studies, learners still activated their prior knowledge during problem solving to invent their solutions. However, in these implementations, learners may have remained satisfied with their erroneous solutions and may not have identified the features required for the correct solution. Only when learners realize that their invented solution is not sufficient and that the canonical solution rest upon a set of deep features, learners may be motivated to revise their ideas and may appreciate the canonical solution taught during instruction.

However, there are two studies in the last row of Table 3 that found benefits for PS-I on conceptual assessments: First, in the study by Kapur (2014b, see Table 1 no. 9), contrasting cases were discussed during the explicit instruction phase. Second, in the study by DeCaro and Rittle-Johnson (2012, see Table 1 no. 2), students were given accuracy feedback on their solutions. It seems that in this context the feedback achieved a similar goal to that of providing contrasting cases or building instruction on student solutions. Naturally this is only a speculation, as one study alone does not allow drawing conclusions.

Limitations and Future Research

To test our theoretical assumptions, further research is needed to extend the discussed findings on the three mechanisms. More work is especially needed with regard to the deep feature mechanism. While we have direct support that PS-I promotes prior knowledge activation (e.g., Kapur 2014a; Kapur and Bielaczyc 2011; Roll et al. 2011) and awareness of knowledge gaps (Glogger-Frey et al. 2015; Loibl and Rummel 2014a), the support for the deep feature mechanism is more indirect (number of deep features addressed in the invented solutions, Loibl and Rummel 2014b; Roll et al. 2012; Wiedmann et al. 2012; number of deep features noticed in a post-test assessment, Roll et al. 2011). Future research could aim at implementing conditions directed at triggering the proposed mechanisms one by one to test our assumptions more thoroughly. In addition, further research should focus in more detail on how the different mechanisms build upon each other and how they can be supported to best foster well-connected knowledge—both in PS-I and other approaches.

The main limitation of existing PS-I studies is their domain specificity, with studies in mathematics clearly dominating. The few studies that were done in other domains (e.g., Schwartz and Bransford 1998) suggest that the mechanisms transfer, at least to some degree, to other domains. We believe that the mechanisms are especially explanatory for rather structured domains that allow for clear identification of deep features and evaluation of solution attempts. However, further studies are required to test this claim.

Notably, our proposed mechanisms focus solely on cognitive explanations. We assume that other processes, such as motivation (Belenky and Nokes-Malach 2012; Glogger-Frey et al. 2015) or collaborative structure (Sears 2006; Mazziotti et al. 2014, 2015; Westermann and Rummel 2012), contribute to learning from PS-I. However, it seems that non-cognitive mechanisms alone could not explain the overall patterns of the findings for the following reasons. Collaborative (e.g., Kapur 2012) as well as individual (e.g., Kapur 2014b) PS-I have been shown effective and results of studies that compare individual and collaborative PS-I remain inconclusive (Sears 2006; Mazziotti et al. 2014, 2015). Motivational factors certainly play a role in PS-I. However, there are mixed relations with learning outcomes: Glogger-Frey et al. (2015) found increased motivation to learn the canonical solution after invention, but this was not reflected in the learning outcomes measured after instruction. Belenky and Nokes-Malach (2012) showed that invention can foster mastery orientation which in turn increased learning in their study. Yet, this effect was only true for students who entered the study with low mastery orientation. Also, motivational accounts may not explain the differentiated effects on transfer and conceptual knowledge versus procedural knowledge.

Conclusion

To conclude, PS-I approaches that consist of a problem-solving phase followed by an instruction phase have received a growing interest as evident by a fairly large number of recent research publications. In this paper, we aimed to identify differences in the implementation of PS-I designs and to descriptively relate these differences to learning outcomes. Our review indicates that PS-I can be beneficial mainly for the acquisition of conceptual knowledge and the ability to transfer. However, whether this potential unfolds depends to a large degree on the implementation of PS-I: PS-I seems beneficial only if the problem is presented in the form of contrasting cases that differ in only one feature at a time or if student solutions are discussed during subsequent instruction. We proposed three complementary cognitive mechanisms that facilitate learning from PS-I, if contrasting cases and/or student solutions are implemented: prior knowledge activation, awareness of knowledge gaps, and recognition of deep features. We argue that the synergy between these mechanisms leads to well-connected and well-organized knowledge that explains the divergent findings of PS-I approaches—both regarding specific PS-I implementations and regarding different types of knowledge. By working towards a theory of PS-I, we hope to advance the discussion about the value of failed problem-solving attempts followed by instruction in specific variations of PS-I.

References

*Papers included in overview (Table 1)

Acuña, S. R., García-Rodicio, H., & Sánchez, E. (2010). Fostering active processing of instructional explanations of learners with high and low prior knowledge. *European Journal of Psychology of Education, 26*(4), 435–452.

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

*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*(3), 399-432.

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin, 128*(4), 612–637.

Booth, J. L., Lange, K. E., Koedinger, K. R., & Newton, K. J. (2013). Using example problems to improve student learning in algebra: differentiating between correct and incorrect examples. *Learning and Instruction, 25*, 24–34.

Chi, M. T. H. (2000). Self-explaining expository texts: the dual processes of generating inferences and repairing mental models. In R. Glaser (Ed.), *Advances in instructional psychology* (pp. 161–238). Hillsdale: Lawrence Erlbaum Associates.

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

Chi, M. T. H., Slotta, J. D., & de Leeuw, N. (1994). From things to processes: a theory of conceptual change for learning science concepts. *Learning and Instruction, 4*, 27–43.

DeCaro, D. A., DeCaro, M. S., & Rittle-Johnson, B. (2015). Achievement motivation and knowledge development during exploratory learning. *Learning and Individual Differences, 37*, 13–26.

*DeCaro, M. S., & Rittle-Johnson, B. (2012). Exploring mathematics problems prepares children to learn from instruction. *Journal of Experimental Child Psychology, 113*, 552-568

diSessa, A. A., Hammer, D., Sherin, B. L., & Kolpakowski, T. (1991). Inventing graphing: meta-representational expertise in children. *The Journal of Mathematical Behavior, 10*(2), 117–160.

Duncker, K. (1945). On problem-solving. *Psychological Monographs, 58*(Whole No. 270).

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.

Eva, K. W., & Regehr, G. (2011). Exploring the divergence between self-assessment and self-monitoring. *Advances in Health Sciences Education, 16*(3), 311–329.

*Fyfe, E. R., DeCaro, M. S. & Rittle-Johnson, B. (2014). An alternative time for telling: when conceptual instruction prior to problem solving improves mathematical knowledge. *British Journal of Educational Psychology, 84*(3), 502-519.

*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*, 72-87.

Heckler, A. F., Kaminski, J. A., & Sloutsky, V. M. (2008). Learning associations that run counter to biases in learning: overcoming overshadowing and learned inattention. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society*. (pp. 511-516). Austin, TX: Cognitive Science Society

Holmes, N. G., Day, J., Park, A. H., Bonn, D. A., & Roll, I. (2014). Making the failure more productive: scaffolding the invention process to improve inquiry behaviours and outcomes in productive failure activities. *Instructional Science, 42*(4), 523–538.

*Hsu, C.-Y., Kalyuga, S. & Sweller, J. (2015). When should guidance be presented in physics instruction? *Archives of Scientific Psychology, 3*(1), 37-53.

Kalyuga, S., & Singh, A.-M. (2015). Rethinking the boundaries of cognitive load theory in complex learning. *Educational Psychology Review. Advance Online Publication*. doi:10.1007/s10648-015-9352-0.

*Kapur, M. (2010). Productive failure in mathematical problem solving. *Instructional Science, 38*(6), 523-550.

*Kapur, M. (2011). A further study of productive failure in mathematical problem solving: unpacking the design components. *Instructional Science, 39*(4), 561-579.

*Kapur, M. (2012). Productive failure in learning the concept of variance. *Instructional Science, 40*(4), 651-672.

Kapur, M. (2014). Comparing learning from productive failure and vicarious failure. *Journal of the Learning Sciences, 23*(4), 651–677.

*Kapur, M. (2014b). Productive failure in learning math. *Cognitive Science, 38*(5), 1008-1022.

Kapur, M. (2016). Examining productive failure, productive success, unproductive failure, and unproductive success in learning. *Educational Psychologist, 51*(2), 289–299.

*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.

Koedinger, K. R., Corbett, A. T., & Perfetti, C. (2012). The knowledge-learning-instruction (KLI) framework: bridging the science-practice chasm to enhance robust student learning. *Cognitive Science, 36*, 757–798.

Leppink, J., Paas, F., Van Gog, T., Van der Vleuten, C., & Van Merrienboer, J. (2014). Effects of pairs of problems and examples on task performance and different types of cognitive load. *Learning and Instruction, 30*, 32–42.

Linn, M. C. (1995). Designing computer learning environments for engineering and computer science: the scaffolded knowledge integration framework. *Journal of Science Education and Technology, 4*, 103–126.

*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.

*Loibl, K. & Rummel, N. (2014a). Knowing what you don't know makes failure productive. *Learning and Instruction, 34*, 74-85.

*Loibl, K., & Rummel, N. (2014b). The impact of guidance during problem-solving prior to instruction on students' inventions and learning outcomes. *Instructional Science, 42*(3), 305-326.

*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.

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, K. O'Connor, T. Lee, T., & L. D'Amico, *Proceedings of the 11th international conference of the learning sciences (ICLS 2014), Vol. 3* (pp. 1184-1185). International Society of the Learning Sciences, Inc.

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 2015* (Vol. 2, pp. 570–575). Gothenburg: ISLS.

Nathan, M. J. (1998). Knowledge and situational feedback in a learning environment for algebra story problem solving. *Interactive Learning Environments, 5*(1), 135–159.

Nathan, M. J., Alibali, M. W., Masarik, D. K., Stephens, A. C., & Koedinger, K. R. (2010). Enhancing middle school students' representational fluency: a classroom-based study. *WCER Working Paper Series no, 2010-9*. Wisconsin Center for Educational Research: Madison, WI. Retrieved June 28, 2013, from http://www.wcer.wisc.edu/Publications/workingPapers/Working_Paper_No_2010_09.pdf.

Needham, D. R., & Begg, I. M. (1991). Problem-oriented training promotes spontaneous analogical transfer: memory-oriented training promotes memory for training. *Memory & Cognition, 19*(6), 543–57.

Paul, A. M. (2012) Why floundering is good. *Time Magazine*. Retrieved June 28, 2013, from http://ideas.time.com/2012/04/25/why-floundering-is-good/.

Prediger, S. (2008). The relevance of didactic categories for analysing obstacles in conceptual change: revisiting the case of multiplication of fractions. *Learning and Instruction, 18*(1), 3–17.

Quilici, J. L., & Mayer, R. E. (1996). Role of examples in how students learn to categorize statistics word problems. *Journal of Educational Psychology, 88*(1), 144–161.

Quilici, J. L., & Mayer, R. E. (2002). Teaching students to recognize structural similarities between statistics word problems. *Applied Cognitive Psychology, 16*(3), 325–342.

Reisslein, J., Atkinson, R., Seeling, P., & Reisslein, M. (2006). Encountering the expertise reversal effect with a computer-based environment on electrical circuit analysis. *Learning and Instruction, 16*, 92–103.

Renkl, A. (2005). The worked-out-example principle in multimedia learning. In R. Mayer (Ed.), *Cambridge handbook of multimedia learning* (pp. 229–246). Cambridge: Cambridge University Press.

Renkl, A. (2014). Towards an instructionally-oriented theory of example-based learning. *Cognitive Science, 38*, 1–37.

Rittle-Johnson, B., & Schneider, M. (2014). Developing conceptual and procedural knowledge of mathematics. In R. Cohen Kadosh & A. Dowker (Eds.), *Oxford handbook of numerical cognition*. Oxford: Oxford University Press.

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

*Roll, I., Aleven, 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: Cognitive Science Society.

*Roll, I., Aleven, V., & Koedinger, K. R. (2011). Outcomes and mechanisms of transfer. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 2824–2829). Austin: Cognitive Science Society.

Roll, I., Holmes, N. G., Day, J., & Bonn, D. (2012). Evaluating metacognitive scaffolding in guided invention activities. *Instructional Science, 40*(4), 691–710.

Roll, I., Wiese, E., Long, Y., Aleven, V., & Koedinger, K. R. (2014). Tutoring self- and co-regulation with intelligent tutoring systems to help students acquire better learning skills. In R. Sottilare, A. Graesser, X. Hu, & B. Goldberg (Eds.), *Design recommendations for adaptive intelligent tutoring systems: volume 2—adaptive instructional strategies* (pp. 169–182). Orlando: U.S. Army Research Laboratory.

Sánchez, E., García Rodicio, H., & Acuña, S. R. (2009). Are instructional explanations more effective in the context of an impasse? *Instructional Science, 37*(6), 537–563.

Schmidt, H. G., De Volder, M. L., De Grave, W. S., Moust, J. H. C., & Patel, V. L. (1989). Exploratory models in the processing of science texts: the role of prior knowledge activation through small-group discussion. *Journal of Educational Psychology, 81*(4), 610–619.

*Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. *Cognition and Instruction, 16*(4), 475–522.

*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.

*Schwartz, D. L., & Martin, T. (2004). Inventing to prepare for future learning: the hidden efficiency of encouraging original student production in statistics instruction. *Cognition and Instruction, 22*(2), 129–184.

Schwartz, D. L., Sears, D., & Chang, J. (2007). Reconsidering prior knowledge. In M. C. Lovett & P. Shah (Eds.), *Thinking with data* (pp. 319–344). New York: Routledge.

Sears, D. A. (2006). *Effects of innovation versus efficiency tasks on collaboration and learning*. Doctoral dissertation, Stanford University, California. Retrieved November 07, 2012, from http://www.stat.auckland.ac.nz/~iase/publications/dissertations/06.Sears.Dissertation.pdf.

Siegler, R. S. (1983). How knowledge influences learning. *American Scientist, 71*, 631–638.

Sweller, J. (1988). Cognitive load during problem solving: effects on learning. *Cognitive Science, 12*(2), 257–285.

Sweller, J., van Merrienboer, J., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251–296.

Toth, E. E., Klahr, D., & Chen, Z. (2000). Bridging research and practice: a cognitively-based classroom intervention for teaching experimentation skills to elementary school children. *Cognition and Instruction, 18*(4), 423–459.

Van Gog, T., Kester, L., & Paas, F. (2011). Effects of worked examples, example-problem, and problem example pairs on novices' learning. *Contemporary Educational Psychology, 36*, 212–218.

Van Gog, T., & Rummel, N. (2010). Example-based learning: integrating cognitive and social-cognitive research perspectives. *Educational Psychology Review, 22*(2), 155–174.

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.

VanLehn, K., Siler, S., Murray, C., Yamauchi, T., & Baggett, W. B. (2003). Why do only some events cause learning during human tutoring? *Cognition and Instruction, 21*(3), 209–249.

Vosniadou, S., & Verschaffel, L. (2004). Extending the conceptual change approach to mathematics learning and teaching. *Learning and Instruction, 14*(5), 445–451.

Wertheimer, M. (1959). *Productive thinking*. New York: Harper & Row.

Westermann, K., & Rummel, N. (2012). Delaying instruction: evidence from a study in a university relearning setting. *Instructional Science, 40*(4), 673–689.

Wiedmann, M., Leach, R. C., Rummel, N., & Wiley, J. (2012). Does group composition affect learning by invention? *Instructional Science, 40*(4), 711–730.