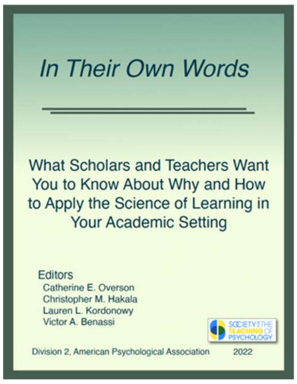

Productive Failure

Manu Kapur
ETH Zurich

Ido Roll
Technion - Israel Institute of Technology

Introduction

Suppose one wants to teach students a math concept that is novel to them, say variability (and the procedure of calculating standard deviation, *SD*). The traditional, prevailing method is to first teach students the concept and procedures of *SD* and then have them solve problems requiring the concept and procedures. This sequence of instruction followed by problem solving is commonly referred to as *direct instruction* (Kirschner et al., 2006). Advocates argue that students should be given the formal target knowledge (such as equations or concepts) before being expected to apply it (Sweller, 2010).

A contrasting method is one that reverses the sequence, that is, one that engages students in problem-solving first, and only then teaches them the concept and procedures (problem-solving followed by instruction, or PS-I, Loibl et al., 2017). One successful approach for problem-solving followed by instruction is called *Productive Failure* (Kapur, 2010, 2012, 2014; Chowrira et al., 2019). By 'failure,' we simply mean that students will typically *not be able* to generate or discover the desired solution(s) by themselves. However, to the extent that students are able to use their prior knowledge to generate suboptimal or even incorrect solutions to the problem, the process can prepare students to learn better from the subsequent instruction (Kapur & Bielaczyc, 2012; Schwartz & Martin, 2004). Not every PS-I design is Productive Failure. Productive Failure is a subset of PS-I designs where the PS and I phases are designed based on the principles of Productive Failure we outline in this chapter. Productive Failure facilitates students' understanding as they explore the task at hand and its structure, features, and requirements, before being handed the tools to solve it. In this manner, productive failure combines the benefits of exploratory problem-solving with those of explicit instruction, thereby mitigating the risk that students do not discover the correct concepts and procedures on their own (Kapur & Rummel, 2012; Loibl et al., 2017).

In this chapter, we report on a program of research on Productive Failure and its implications for designing instruction (see also Kapur, 2016). We start with a brief description of what Productive Failure is, describing its underlying design and principles. We then discuss the research basis examining the effectiveness of Productive Failure. Here we illustrate the effect by describing one of our classroom-based experimental studies, followed by summarizing key findings from our meta-analysis of fifteen years of research on productive failure. We end by discussing the educational implications of our research.

Productive Failure (PF)

Productive Failure (PF) is a learning design that affords students opportunities to generate representations and solutions to a novel problem that targets a concept they have not yet learned, which is followed by consolidation and knowledge assembly, where they receive instruction on the targeted concept. Because learners have not learned the concept, the problem-solving process usually leads to failure. By failure, we mean that while students make progress, they are typically unable to generate or discover the target solution(s) by themselves. However, as they struggle to construct valid solutions, students gain understanding of the challenge, as well as its relationship with their existing relevant knowledge, in a manner that prepares them to learn better from the subsequent instruction (Kapur & Bielaczyc, 2012; Schwartz & Martin, 2004; Loibl et al., 2017).

By encouraging students to analyze and attempt to solve relevant problems ahead of instruction, PF addresses two challenges of direct instruction. First, in direct instruction, students often gain limited understanding of the target concepts. This is because instruction uses language and representations that are foreign to them. Instead, PF lets students apply their current understandings and thus differentiate the required knowledge (e.g., Kapur & Bielaczyc, 2012; Schwartz & Bransford, 1998; Schwartz & Martin, 2004; Holmes et al., 2014). Second, direct instruction presents the target concept as a monolithic entity, the rationale of which students often fail to understand (Chi et al., 1988; Schwartz & Bransford, 1998). By recognizing the requirements of the target solution during the initial problem-solving phase, students are more able to understand the rationale behind the target procedures (Loibl et al., 2017).

It is important to emphasize that we do not advocate that any type of ill-designed problem-solving activities prior to instruction suffices. Rather, we put forward an effective set of principles for design of such activities. The PF process is designed around four key principles:

1. Activation and differentiation of prior knowledge in relation to the targeted concepts (that is, understanding what existing knowledge is relevant, what it achieves, and what are its shortcomings).
2. Attention to critical features of the targeted concepts (e.g., understanding the idea that variability needs to account for the number of points, or is a property of distance).
3. Explanation and elaboration of these features, including their mathematical properties (e.g., for the concept of variability, division by *N* helps control for sample size; subtraction is a measure of distance).
4. Organization and assembly of the critical conceptual features into the targeted concepts.

PF has two main phases. During the initial problem solving (or **Generation** phase), students are given opportunities to explore the affordances and constraints of different approaches towards the solution. These approaches are based on familiar procedures and representations that students may already know. During the subsequent **Consolidation** phase, students are given opportunities to assemble these features and update them based on explicit instruction about the typical solutions which use the target knowledge to be taught. As stated above, it is important to emphasize that not every failure is productive, and not every task is suitable for exploration. The following design features support the above-mentioned principles:

1. The problem-solving context should accommodate various solution approaches. Students should be able to make progress in these, even if they do not solve them entirely. While students should be challenged, they should not be frustrated (see Metcalfe, this volume, on the region of proximal learning).
2. These challenges should invite students to explain and elaborate on their existing knowledge as well as task requirements.
3. While no explicit, external feedback is provided, students should be able to extract situational feedback by evaluating their solution approaches using their intuition and interpretation of their outcomes (Nathan, 1998; Roll et al., 2014). Such self-feedback helps students better understand

the properties of their approaches. Students later should contrast their own solutions with the taught target procedure.

Evidentiary Basis of PF

Here we illustrate with an example of one such study (Kapur, 2012) in a real classroom setting. Kapur (2012) compared learning from PF and Direct Instruction (DI) through a pre-posttest, quasi-experimental study with 133 ninth-grade mathematics students (14–15-year-olds) from a public school in Singapore. Students had no instructional experience with the targeted concept prior to the study. All students, in their intact classes, participated in four, 50-minute periods of instruction on the concept as appropriate to their assigned condition. The same teacher taught both the PF and DI conditions.

In the PF condition, students spent forty minutes to solve the following data analysis problem on their own:

The PF problem solving task

Mr. Fergusson, Mr. Merino, and Mr. Eriksson are the mangers of the Supreme Football Club. They are on the lookout for a new striker, and after a long search, they short-listed three potential players: *Mike Arwen*, *Dave Backhand*, and *Ivan Right*. All strikers asked for the same salary, so the managers agreed that they should base their decisions on the players' performance in the Premier League for the last 20 years. Table 1 shows the number of goals that each striker had scored between 1988 and 2007.

Table 1: Number of goals scored by three strikers in the Premier League.

Year	Mike Arwen	Dave Backhand	Ivan Right
1988	14	13	13
1989	9	9	18
1990	14	16	15
1991	10	14	10
1992	15	10	16
1993	11	11	10
1994	15	13	17
1995	11	14	10
1996	16	15	12
1997	12	19	14
1998	16	14	19
1999	12	12	14
2000	17	15	18
2001	13	14	9
2002	17	17	10

2003	13	13	18
2004	18	14	11
2005	14	18	10
2006	19	14	18
2007	14	15	18
The managers agreed that the player they hire should be a consistent performer. They decided that they should approach this decision mathematically and would want a formula for calculating the consistency of performance for each player. This formula should apply to all players and help provide a fair comparison. The managers decided to get your help. Please come up with a formula for consistency and show which player is the most consistent striker. Show all working and calculations on the paper provided.			

The data analysis problem presented a distribution of goals scored each year by three soccer players over a twenty-year period. Students were asked to design a quantitative index to determine the most *consistent* player. The choice of consistency was to help students harness their intuitive understanding of the concept to their mathematical generation of indexes. During this generation phase, no explicit guidance was provided. After this, the teacher first consolidated by comparing and contrasting student-generated solutions with each other, and then modeled and worked through the canonical solution. Finally, students solved three data analysis problems for practice, and the teacher discussed the solutions with the class.

In the DI condition, the teacher explained the canonical formulation of the concept of variance using a series of worked examples. After each worked example, students solved a comparable problem, then their errors, misconceptions, and critical features of the concept were discussed with the class as a whole. Having learned the concept, students then solved the same problem that the PF students solved in their generation phase, following which the teacher discussed the solutions with the class. The DI cycle ended with a final set of three data analysis problems for practice (the same problems given to the PF students), and the teacher discussed the solutions with the class.

Process findings suggested that PF groups generated on average six solutions to the problem. These student-generated solutions have been described in greater detail elsewhere (see Kapur, 2012, 2013, 2014). Notably, none of the PF groups were able to generate the canonical formulation of *SD*. Furthermore, none of the PF groups generated a consistent and effective mathematical expression that captures consistency. In contrast, analysis of DI students' classroom work revealed that students relied *only* on the canonical formulation to solve data analysis problems. The solutions generated by PF students suggested that not only were students' prior knowledge activated (central tendencies, graphing, differences, etc.) but that students were able to assemble them into different ways of measuring consistency. Therefore, the more solutions students can generate, the more it can be argued that they are able to conceptualize the targeted concept in different ways; that is, their prior knowledge is not only activated but also differentiated in the process of generation. In other words, these solutions can be seen as a measure, albeit indirect, of knowledge activation and differentiation; the greater the number of such solutions, the greater the knowledge activation and differentiation.

On the day immediately following the intervention, all students took a posttest comprising three types of items: procedural fluency, conceptual understanding, and transfer (for the items, see Kapur, 2012). Items that evaluated procedural fluency required students to calculate the value of *SD* and interpret it in a given context. Items that evaluated conceptual understanding required students to notice missing features in suboptimal solutions and correct them, articulate why *SD* is formulated the way it is, and apply its mathematical properties. Transfer items required students to flexibly adapt their knowledge of *SD* to solve problems on normalization, a concept that was not taught nor prompted during instruction.

Analysis of pre-post performance suggested that PF students significantly outperformed their DI counterparts on conceptual understanding and transfer without compromising procedural fluency. Further analyses revealed that the number of solutions generated by PF students was a significant predictor of how much they learned from PF. That is, the more solutions they generated, the better they performed on the procedural fluency, conceptual understanding, and transfer items on the posttest.

These findings are consistent with the earlier studies on productive failure (Kapur, 2008; Kapur & Kinzer, 2009), as well as with additional studies that show the benefits of having students invent (often faulty) solutions to novel problems as a way of preparing them for future instruction (Chowrira et al., 2019; Schwartz & Bransford, 1998; Schwartz & Martin, 2004). Notably, evidence for learning from PF comes not only from quasi-experimental studies conducted in the real ecologies of classrooms across topics and age groups (e.g., Kapur, 2012, 2013; Schwartz & Bransford, 1998; Schwartz & Martin, 2004), but also from controlled experimental studies (e.g., DeCaro & Rittle-Johnson, 2012; Kapur, 2014; Loibl & Rummel, 2014; Roll et al., 2011; Schmidt & Bjork, 1992; Schwartz et al., 2011; Chowrira et al., 2019).

These findings are also consistent with the math education literature that emphasizes the role of struggle in learning (e.g., Hiebert & Grouws, 2007). More broadly, these findings can also be seen to be consistent with some forms of Problem-Based learning (PBL) environments in which students are given just-in-time instruction after they have engaged in problem-solving first (Capon & Kuhn, 2004; Hmelo-Silver et al., 2007). More broadly, these findings demonstrate the value of constructivist instruction (Tobias & Duffy, 2009). Proponents of DI often point out sparse data to support instruction that is minimally guided (Kirschner et al., 2006; Klahr, 2010; Sweller, 2010). The PF approach consistently demonstrates that, done right, students' sense making based on their prior knowledge can be effective vehicles to promote more robust learning. However, it is important to emphasize that these are effective learning tools when followed by DI. That is, rather than a single-phase instruction that either provides or withholds information, the PF effect demonstrates the value of combining them by delaying information.

To determine the effectiveness of PF across all the studies, we carried out a meta-analysis of relevant research from the past two decades. Our meta-analysis of more than 12,000 participants in 166 experimental comparisons (Sinha & Kapur, 2021) found that:

1. Productive Failure students significantly outperformed their counterparts in the traditional instruction-first classrooms in conceptual understanding and transfer (Cohen's $d = 0.36$ [95% confidence interval (*CI*) 0.20, 0.51]), without compromising procedural knowledge; simply put, students exposed to the PF design developed a better understanding of domain-specific ideas, why these ideas work for certain problems, and whether and how these ideas can be applied to solve novel (but related) problems in the future.
2. The higher the fidelity to the design principles of PF, the stronger the effect (with Cohen's d up to 0.58). To put this in context, this effect is about three times the effect a good teacher has on student learning in one year.

Educational Implications

The PF effect is far from magic. It is the result of carefully designed instruction that invokes evidence-based learning mechanisms. It is worth emphasizing again that not every problem-solving activity followed by instruction sequence equals PF. The meta-analysis clearly showed that fidelity to PF principles is key. We first discuss the fidelity implications of PF, followed by more general implications for orchestrating PF in academic courses.

Fidelity to PF Principles

Fidelity to PF comes from making sure the design of the problem-solving tasks and activities is in line with the PF principles, as is the context and the social surround or classroom culture within which PF is used.

1. *Features of PF tasks.* Task development is key. To be effective, tasks should include the following features. These often take 2-3 iterations to develop:
    a. *Contrasts*. Tasks should include contrasts that can be intuitively, and even better if perceptually, evaluated. These allow learners to evaluate their own solutions. While no explicit feedback is given, learners should be able to evaluate their own generated methods.
    b. *Variant-invariant relations*. Varying certain features of the task and keeping others invariant affords students the opportunities to focus on the critical features. For example, in the PF task earlier, we kept the central tendencies and the range the same. Once students noticed this, it afforded them and even implicitly nudged them to focus on the distribution of the data. There is no one way of doing this. Alternatively, variant-invariant relations can be designed on structure and stories, e.g., keeping structure constant but changing the stories, or vice versa. The same could be done with multiple representations. All of these must be carefully designed to keep some features constant and others variable.
    c. *Affective draw*. To encourage learners to persist in these challenging problems the tasks should be engaging. Using intuitive language and inviting intuitive and informal ways of reasoning motivates students to engage with the tasks.
    d. *Room for baby steps*. Correct solutions should not require insights. Instead, learners should be able to make gradual progress towards meaningful approaches to solving the challenges. While overall students' solutions are typically erroneous, they often include valid features and ideas.
    e. *Invite multiple solutions and representations*. To facilitate students' understanding of the target concept and its relationship with neighboring concepts, tasks should invite a variety of solutions or representations.
    f. *Low computational load*. Students should focus on designing ideas and exploring their properties, rather than on calculating or applying solutions.
    g. *Be problem oriented*. While the goal and benefits of PF are conceptual, PF tasks ask students to compute different attributes by designing methods. By contextualizing the challenge, learners explore more aspects of the problem and are given tools to evaluate their progress.
    h. *Working alone or together*. Although the initial design of PF advocated for collaborative problem solving, thus far the evidence shows that PF is effective in both configurations, whether learners work individually or in small groups (e.g., Mazziotti et al., 2019).
2. *Classroom context and culture:* Managing student expectations and norms in line with PF design is key. Students should be encouraged to explore and take risks. The expectation should be on the effort and generating ideas even if they are incorrect or suboptimal. The guiding principle is

that every answer that can be explained is a good answer. Answers that provide fuller explanations are better. Students should be told that they may struggle a little and feel frustrated, yet it is this struggle and dealing with frustration that is normal and even seen as an indicator of learning. These expectations and norms need to be emphasized repeatedly over the course of the academic term so that students understand the new didactic contract being put in place. In many ways, one could conceive the unit of design and change not to be a pedagogy or teaching method, but classroom culture.

General Considerations and Implications

Having worked with hundreds of teachers, schools, and even education systems at large in a variety of applied contexts, we now outline some general considerations for using PF.

1. *Let students grapple with new concepts*: In our daily lives, most effective learning comes by first encountering new topics. Also in schools, novices should be invited to meaningfully explore new concepts before being taught target procedures. This exploration helps them build relevant experiences with which they can later encode instruction.
2. *Type of knowledge*: PF is effective when learning new concepts, and it shows effects on conceptual learning and preparation for future learning (PFL). For example, it helps students understand how certain concepts should be applied, what they mean, and how they can be used to assemble more complex ideas. Alas, compared to DI, PF is not more effective for procedural fluency, for example, calculating mean, manipulating algebraic variables, and so on.
3. *Budget your use of PF*: It follows from point 2 that we do not advocate the use of PF for all types of learning goals. Do not use PF in every lesson. Instead, target 3-5 key concepts and big ideas over the course of a semester. Big ideas are those that are foundational to multiple topics, such as ratios (Koedinger & Roll, 2012). We have found that designing for PF does not require a major overhaul in the curriculum content or time. Even a surgical addition prior to key concept lectures can result in positive effects.
4. *Use PF in your STEM lessons*: Our meta-analysis showed that PF was found to work well across STEM topics (Loibl et al., 2017; Sinha & Kapur, 2021). However, so far, it was not found to be as effective for domain-general skills, and there is only scarce evidence for non-STEM fields. While absence of evidence is not evidence of absence, much more research is needed in domain-general skills and non-STEM fields before we can make an evidence-based claim about using PF.
5. *Learner profile*: PF is an effective tool in diverse classrooms. It has shown to work with all ability groups provided the design fidelity is good. Age-wise, meta-analysis shows better effects for older students (secondary school onwards), possibly because they have better problem-solving skills. Older students are arguably better at analyzing their own solutions and extract situational feedback that allows them to learn from their challenges. The fidelity of PF studies with the younger children also tended to be low. Once again, more research is needed here.
6. *Technology*: Some evidence suggests that technology can be used to support PF (Roll et al., 2010; Chase et al., 2019; Shalala et al., 2021). Blended learning, gaming, and Virtual- and Mixed-Reality (VR and MR) environments are all relevant contexts for PF. For example, we are engaging undergraduate students in playing complex ethical scenario games in their research contexts prior to learning ethics in their traditional lectures. Likewise, we are designing embodied interaction activities in VR and MR environments in calculus, graph theory, and linear algebra as pre-instructional activities. Similar ideas are being applied in medical education using clinical diagnosis and reasoning simulations.

7. *Assessment of effectiveness*: PF shows benefits on a variety of measures, such as student production, classroom interaction, performance on tests provided they measure conceptual understanding and transfer. As the PF process requires complex problem solving and independent learning, it provides an opportunity to assess key competencies and attributes, such as persistence, creativity, collaboration, and interest. Alas, with few exceptions, these have not been studied sufficiently in the literature (c.f. Roll et al., 2012; Massey-Allard et al., 2019; Belenky & Nokes-Malach 2013)
8. *Teacher Professional Development*: Teachers should be supported in learning to design PF instruction. There is a need for a development runway for iterations and experimentation. Failure to design for this often results in premature failure of PF.

Conclusion

Contrary to the commonly held belief that there is little efficacy in having learners solve novel problems that target concepts they have not yet learned, our work suggests that there is indeed such an efficacy even if learners do not formally know the underlying concepts needed to solve the problems, and even if such problem solving leads to failure initially. A wealth of research has identified several design guidelines that support learning by attempting to generate solutions to novel problems. Applying these principles to key concepts can achieve a large effect on students' knowledge and motivation. Further, PF is not a monolithic design and other designs may facilitate similar effects. Notably, PF is only one way to orchestrate known mechanisms that are effective also in other forms of instruction. For example, creating a safe classroom culture, encouraging students to activate their prior knowledge, or providing room for students to make sense of their solutions have all been shown to be effective.

Author Note

This work was supported by several Ministry of Education, Singapore grants to the first author, and also by ETH Zurich, Switzerland. This work was also supported by the Gordon and Betty Moore Foundation and by the Israel Science Foundation (ISF) Grant No 1573/21 to the second author. We thank Dr. Janan Saba for her inputs. Any opinions are those of the authors and do not represent the views of the funding agencies.

References

Belenky, D. M., & Nokes-Malach, T. J. (2013). Mastery-approach goals and knowledge transfer: An investigation into the effects of task structure and framing instructions. *Learning and Individual Differences*, *25*, 21–34. https://doi.org/10.1016/j.lindif.2013.02.004

Capon, N., & Kuhn, D. (2004). What's so good about problem-based learning? *Cognition and Instruction*, *22*(1), 61–79. https://doi-org.unh.idm.oclc.org/10.1207/s1532690Xci2201_3

Chase, C. C., Connolly, H., Lamnina, M., & Aleven, V. (2019). Problematizing helps! A classroom study of computer-based guidance for invention activities. *International Journal of Artificial Intelligence in Education*, *29*(2), 283-316.

Chi, M. T. H., Glaser, R., & Farr, M. J. (1988). *The nature of expertise*. Erlbaum.

Chowrira, S. G., Smith, K. M., Dubois, P. J., & Roll, I. (2019). DIY productive failure: boosting performance in a large undergraduate biology course. *NPJ Science of Learning*, *4*(1), 1-8.

DeCaro, M. S., & Rittle-Johnson, B. (2012). Exploring mathematics problems prepares children to learn from instruction. *Journal of Experimental Child Psychology*, *113*(4), 552-568.

Hiebert, J., & Grouws, D. A. (2007). The effects of classroom mathematics teaching on students' learning. In F. Lester (Ed.), *Second Handbook of Research on Mathematics Teaching and Learning* (pp. 371-404). Information Age.

Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: a response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist*, *42*(2), 99-107.

Holmes, N. G., Day, J., Park, A. H. K., Bonn, D. A., & Roll, I. (2014). Making the failure more productive: scaffolding the invention process to improve inquiry behaviors and outcomes in invention activities. *Instructional Science*, 42(4), 523–538. https://doi.org/10.1007/s11251-013-9300-7

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. (2014). Productive failure in learning math. *Cognitive Science*, *38*(5), 1008-1022. https://doi.org/10.1111/cogs.12107

Kapur, M. (2013). Comparing learning from productive failure and vicarious failure. *The Journal of the Learning Sciences*, *23*(4), 651-677. https://doi.org/10.1080/10508406.2013.819000

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., & 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

Kapur, M., & Rummel, N. (2012). Productive failure in learning from generation and invention activities. *Instructional Science*, 40(4), 645-650. https://doi.org/10.1007/s11251-012-9235-4

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., & Kinzer, C. (2009). Productive failure in CSCL groups. *International Journal of Computer-Supported Collaborative Learning* (*ij*CSCL), *4*(1), 21-46. https://doi.org/10.1007/s11412-008-9059-z

Kapur, M. (2008). Productive failure. *Cognition and Instruction*, *26*(3), 379-424. https://doi.org/10.1080/07370000802212669

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work. *Educational Psychologist*, *41*(2), 75-86.

Klahr, David (2018): Coming Up for Air: But Is It Oxygen or Phlogiston? A Response to Taber's Review of Constructivist Instruction: Success Or Failure?. Carnegie Mellon University. Journal contribution. https://doi.org/10.1184/R1/6614000.v1

Koedinger, K. R., & Roll, I. (2012). Learning to think: Cognitive mechanisms of knowledge transfer. In K. J. Holyoak & R. G. Morrison (Eds.), *The Oxford handbook of thinking and reasoning* (pp. 789–806). Oxford University Press.

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.

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

Massey-Allard, J., Roll, I., & Ives, J. (2019). Supporting productive exploration in invention activities: are simulations too challenging? Paper presented at Physics Education Research Conference 2019, Provo, UT. Retrieved from https://www.compadre.org/Repository/document/ServeFile.cfm?ID=15302&DocID=5214

Mazziotti, C., Rummel, N., Deiglmayr, A., & Loibl, K. (2019). Probing boundary conditions of Productive Failure and analyzing the role of young students' collaboration. *NPJ science of learning*, *4*(1), 1-9.

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

Roll, I., Aleven, V., & Koedinger, K. R. (2010). The invention lab: Using a hybrid of model tracing and constraint-based modeling to offer intelligent support in inquiry environments. In V. Aleven, J. Kay, & J. Mostow (Eds.), *Proceedings of the International Conference on Intelligent tutoring systems* (pp. 115-124). Springer Verlag.

Roll, I., Aleven, V., & Koedinger, K. (2011). Outcomes and mechanisms of transfer in invention activities. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 33. Retrieved from https://escholarship.org/uc/item/4rx036z0

Roll, I., Holmes, N. G., Day, J., & Bonn, D. A. (2012). Evaluating metacognitive scaffolding in Guided Invention Activities. *Instructional Science*, *40*(4), 691–710. https://doi.org/10.1007/s11251-012-9208-7

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.

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

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., 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.

Shalala, R., Amir, O., & Roll, I. (2021). Towards asynchronous data science invention activities at scale. In Hmelo-Silver, C. E., De Wever, B., & Oshima, J. (Eds.), *Proceedings of the 14th International Conference on Computer-Supported Collaborative Learning - CSCL 2021* (pp. 43-50). Bochum, Germany: International Society of the Learning Sciences.

Sinha, T., & Kapur, M. (2021). When problem solving followed by instruction works: Evidence for productive failure. *Review of Educational Research*, *91*(5), 761–798. https://doi.org/10.3102/00346543211019105

Sweller, J. (2009). What human cognitive architecture tells us about constructivism. In S. Tobias & T. M. Duffy (Eds.), *Constructivist instruction: Success or failure?* (pp. 127–143). Routledge/Taylor & Francis Group.

Tobias, S., & Duffy, T. M. (Eds.). (2009). *Constructivist instruction: Success or failure?* Routledge/Taylor & Francis Group.