
Title	Learning from productive failure
Author(s)	Manu Kapur
Source	<i>Learning: Research and Practice</i> , 1(1), 51-65
Published by	Taylor & Francis (Routledge)

Copyright © 2015 Taylor & Francis

This is an Accepted Manuscript of an article published by Taylor & Francis in *Learning: Research and Practice* on 19/12/2014, available online:

<http://www.tandfonline.com/10.1080/23735082.2015.1002195>

Notice: Changes introduced as a result of publishing processes such as copy-editing and formatting may not be reflected in this document. For a definitive version of this work, please refer to the published source.

Citation: Kapur, M. (2015). Learning from productive failure. *Learning: Research and Practice*, 1(1), 51-65. <http://dx.doi.org/10.1080/23735082.2015.1002195>

Manuscript ID: LRP201416

RUNNING HEAD: Productive Failure

Learning from Productive Failure

Abstract

I advance a theoretically and empirically-grounded case for designing for and learning from failure, and instantiate it in a learning design called Productive Failure (PF). I describe the key mechanisms and the design principles of PF. The PF learning design comprises a generation and exploration phase followed by a consolidation and knowledge assembly phase. Findings show that the PF learning design is more effective in developing conceptual understanding and transfer than a direct instruction design. Follow-up studies are described wherein key aspects of the productive failure design were tested over multiple classroom-based studies as well as controlled experiments, and how these studies helped us interrogate and understand the criticality of key mechanisms embodied in the PF design. Implications for the learning theory and the design of instruction are discussed by situating findings in the long-standing instructivist-constructivist debate.

Designing and examining the effectiveness of appropriate support structures to help students learn is an important area in the educational and the learning sciences (Wood, Bruner & Ross, 1976). By help I mean the various support structures such as structuring of the problem itself, scaffolding, instructional facilitation, provision of tools, content support, expert help, and so on (e.g., Hmelo-Silver, Duncan & Chinn, 2007; Kapur & Rummel, 2009; Puntambekar & Hubscher, 2005). Regardless of the type of structure, it is provided to support learners while they are engaged in learning something novel to them. The essential argument is that structure helps learners accomplish what they might not otherwise be able to in the absence of structure. However, there are two fundamental questions that need to be addressed before one can design appropriate support structures for learning:

- a) how does one know that the learner actually needs help?
- b) how does one know what kinds of help to provide them with?

A frequent response to these questions is simply that we—researchers and designers—already have a reasonably good idea from past research what novices are unable to accomplish when learning something new. For example, they may not be able to solve problems requiring concepts they have not learned yet (Sweller, 1988), or they may not be able to collaborate productively (Dillenbourg, 2002), or they may have misconceptions of natural phenomenon that need to be addressed (Chi, Glaser & Farr, 1988), and so on. Much research has documented these difficulties and the kinds of help that learners need. Therefore, the argument goes that if we already know the kinds of help learners need, then we may as well anticipate and design the appropriate structures to help them learn. In other words, if we know the areas students typically need help with, then we can build proper support structures in the learning environment to foster the desired learning outcomes.

Much empirical work and analysis supports the case for helping learners and has consistently shown that minimally structured problem-based learning rarely leads to anything

meaningful (Mayer, 2004; Kirschner, Sweller & Clark, 2006). Hence, there is a deeply ingrained maxim that support structures need to be provided for learners when learning something new, for without such support, they may fail to learn anything. However, this maxim does not necessarily imply that there is little or no efficacy embedded in the resulting “failure” because there is a possibility that failure, if designed for well, may actually be beneficial for learning. It is this possibility that I advance in this paper.

I argue that in spite of knowing the kinds of help students need when learning something new, it may well be productive to withhold such help initially even if it leads to failure in the process. I am not alone in making such an argument. Several programs of research collectively point to the efficacy of designing for difficulty, uncertainty, and struggle in the initial learning (e.g., English, 2013; Schmidt & Bjork, 1992; Schwartz & Martin, 2004). My research on *Productive Failure* (Kapur, 2008, 2010, 2011, 2012, 2013, 2014; Kapur & Bielaczyc, 2012) has also been grounded in the belief that engaging novices to try, and even fail, at tasks that are beyond their skills and abilities can, under certain conditions, be productive for developing deeper understanding and learning provided one can build upon and learn from this failure.

This paper is organized into four sections. I start by explicating the design principles of Productive Failure (PF) and the theoretical mechanisms they embody. Next, I describe key findings from a program of research on PF in learning mathematics. I conclude by situating my findings in the broader research literature, as well as deriving implications for the long-standing instructivist-constructivist debate on the design of learning (Kirschner, et al., 2006; Tobias & Duffy, 2012).

What is Productive Failure?

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

learned yet, followed by consolidation and knowledge assembly where they learn the targeted concept. Because learners have not learned the concept, and further, are asked to generate solutions without any cognitive support or scaffolds, the problem solving process invariably leads to failure. By failure, I simply mean that students will typically be unable to generate or discover the correct solution(s) by themselves. However, to the extent that students are able to use their prior knowledge to generate sub-optimal or even incorrect solutions to the problem, the process can be productive in preparing them to learn better from the subsequent instruction that follows (Kapur, 2012, 2014; Schwartz & Martin, 2004).

Thus conceived, PF can be contrasted with a teaching method that is commonly advocated for both in theory and practice, that is, direct instruction (e.g., Klahr & Nigam, 2004). However, there are at least two problems with direct instruction in the initial phase of learning something new or solving a novel problem. First, students often do not have the necessary prior knowledge differentiation to be able to discern and understand the affordances of the domain-specific representations and methods underpinning the targeted concepts given during direct instruction (e.g., Kapur, 2012; Schwartz & Bransford, 1998; Schwartz & Martin, 2004). Second, when concepts are presented in a well-assembled, structured manner during direct instruction, students may not understand why those concepts, together with their representations, and methods, are assembled or structured in the way that they are (Chi, Glaser, & Farr, 1988; Schwartz & Bransford, 1998).

Cognizant of these two problems, PF engages students in a learning design (for a fuller explication of the design principles, see Kapur & Bielaczyc, 2012) that embodies four core, interdependent mechanisms: a) activation and differentiation of prior knowledge in relation to the targeted concepts, b) attention to critical conceptual features of the targeted concepts, c) explanation and elaboration of these features, and d) organization and assembly of the critical conceptual features into the targeted concepts. These mechanisms are embodied

in a two phase design: a generation and exploration phase (Phase 1) followed by a consolidation phase (Phase 2). Phase 1 affords opportunities for students to generate and explore the affordances and constraints of multiple representations and solution methods. Phase 2 affords opportunities for organizing and assembling the relevant student-generated solutions into canonical solutions. The designs of both phases were guided by the following core design principles that embody the abovementioned mechanisms:

- (1) create problem-solving contexts that involve working on complex problems that challenge but do not frustrate, rely on prior mathematical resources, and admit multiple solutions (mechanisms a and b);
- (2) provide opportunities for explanation and elaboration (mechanisms b and c);
- (3) provide opportunities to compare and contrast the affordances and constraints of failed or sub-optimal solutions and the assembly of canonical solutions (mechanisms b-d).

The design of the tasks, activities and social surround is developmentally-calibrated with each group of students, through multiple iterations of pilot-testing, refinement, and implementation. Tasks and activities that are found to not sufficiently activate and differentiate student prior knowledge are iteratively refined so that they afford students the opportunities to leverage both their formal (content learnt in the formal school curriculum) and informal (intuitive ways of thinking and reasoning developed through experience in the world, and alongside formal mathematics knowledge) mathematical knowledge.

Examining the PF Design

Having articulated the mechanisms embodied in the design principles of PF, the test of the effectiveness of PF through a series of studies will now be described. The description will be initiated by comparing learning from PF with Direct Instruction (DI).

Comparing PF with DI

A comparison of learning from PF and DI through a pre-posttest, quasi-experimental study with 133, ninth-grade mathematics students (14-15 year olds) from a public school in Singapore was illustrated (for fuller details, see Kapur, 2012). The targeted concept was Standard Deviation (SD), which is typically taught in the tenth grade, and therefore, 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 the first two periods working face-to-face in triads to solve a complex data analysis problem on their own (see Appendix A). 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. During this generation phase, no cognitive guidance or support was provided. In the third period, the teacher first consolidated by comparing and contrasting student-generated solutions with each other, and then modeled and worked through the canonical solution. In the fourth and final period, students solved three data analysis problems for practice, and the teacher discussed the solutions with the class.

In the DI condition, the teacher used the first period to explain the canonical formulation of the concept of variance using two sets of “worked-example followed by problem-solving” pairs. The data analysis problems required students to compare the variability in 2-3 given data sets, for example, comparing the variability in rainfall in two different months of a year. After each worked example, students solved an isomorphic problem, following which, their errors, misconceptions, and critical features of the concept were discussed with the class as a whole. To motivate students to pay attention and remain engaged, they were told that they will be asked to solve isomorphic problems after the

teacher-led worked examples. In the second period, students were given three isomorphic data analysis problems to solve, and the solutions were discussed by the teacher. In the third period, students worked in triads to solve the same problem that the PF students solved in the first two periods, following which the teacher discussed the solutions with the class. DI students did not need two periods to solve the problem because they had already learned the concept. The DI cycle ended with a final set of three data analysis problems for practice (the same problems given to the PF students), which the students solved individually, and the teacher discussed the solutions with the class.

Process findings suggested that PF groups generated on average six solutions to the problem. Elsewhere (see Kapur, 2012, 2013, 2014), these student-generated solutions are described in greater detail. None of the PF groups were able to generate the canonical formulation of SD. In contrast, analysis of DI students' classroom work revealed that students relied *only* on the canonical formulation to solve data analysis problems.

Furthermore, the solutions generated by PF students suggested that not only were students' priors activated (central tendencies, graphing, differences, etc.) but that students were able to assemble them into different ways of measuring consistency. After all, PF students could only rely on their priors—formal and intuitive—to generate these solutions. Therefore, the more they can generate, the more it can be argued that they are able to conceptualize the targeted concept in different ways, that is, their priors are 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 after the intervention, all students took a posttest comprising three types of items: procedural fluency, conceptual understanding, and transfer (for the items, see Kapur, 2012). Procedural fluency items required students to calculate the value of

SD and interpret it in a given context. Conceptual understanding items targeted knowledge of the critical features of the concept of SD, and required students to notice missing features in sub-optimal 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 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 students generated, the better they performed on the procedural fluency, conceptual understanding, and transfer items on the posttest. I refer to this effect as the *solution generation effect*.

Discussion

These findings are consistent with the seminal studies on productive failure (Kapur, 2008, 2012), and also with the PFL and IPL studies described earlier (Schwartz & Bransford, 1998; Schwartz & Martin, 2004). These findings are 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 where students are given just-in-time instruction after they have engaged in problem-solving first (Capon & Kuhn, 2004; Edelson, 2001; Hmelo-Silver et al., 2007). It is instructive to note that not all PBL environments are designed for in ways that are similar to PF.

As noted earlier, the efficacy of learning from PF over DI has also been independently replicated in randomized controlled settings (Kapur, 2014; Schwartz et al., 2011; DeCaro &

Rittle-Johnson, 2012). To explain these findings, this paper posits the argument that the PF design invoked learning processes that not only activated but also differentiated students' prior knowledge as evidenced by the number of student-generated solutions. The solution generation effect evidenced the mechanism of activation and differentiation of knowledge. According to the design theory of PF, what prior knowledge differentiation affords in part is a comparison and contrast between the various solutions—among the student-generated solutions as well as between the student-generated and canonical solutions. Specifically, these contrasts afford opportunities to attend to the critical features of the targeted concept that are necessary to develop a deep understanding of the concept, which is precisely what conceptual understanding items on the posttest measured.

Further Studies Examining the PF Design

On the one hand, the finding that the more solutions students generate, the more they learn from PF on average—the solution generation effect—evidenced one of the key mechanisms of the PF design of prior knowledge activation and differentiation. On the other hand, the solution generation effect also raised important questions for further inquiry. In this section, four such lines of inquiry will be described; each testing a critical aspect of the PF design. Once again, fuller descriptions of these studies can be found in published work elsewhere, and therefore, the intention here is to briefly describe and summarize the findings and their implications for the PF design.

The role of math ability

A key assumption in the PF design is that students have the formal and intuitive resources for generation and exploration prior to learning a new concept. In the light of the solution generation effect, an obvious and immediate question given was to examine the role of math ability. After all, one could expect math ability to influence what and how much students generate, and consequently how much students learn from PF.

Testing the efficacy of PF over DI across different math ability profiles was precisely the aim of the studies reported in Kapur and Bielaczyc (2012). Students were purposefully sampled from three public, co-educational schools with significantly different math ability profiles—75 high ability, 114 medium ability, and 113 low ability—on the national standardized examinations in Singapore. In each school, students in their intact classes were assigned to the PF or the DI condition taught by the same teacher.

Several key findings were demonstrated: a) the relative efficacy of PF over DI was replicated, b) the solution generation effect was replicated, and c) students with significantly different math ability were not as different in terms of their capacity to generate solutions during the generation and exploration phase. Consequently, students across different ability profiles were able to learn better from PF than DI. Taken together, these findings provided a strong evidence for the design principles of PF, as well as the mechanisms of activation and differentiation of prior knowledge and attention to critical features, and demonstrated the tractability of PF across a range of math ability provided one is able to design according to the design principles of PF.

The role of guided versus unguided generation

A critical design decision for PF is to not provide cognitive guidance or support during the generation and exploration phase. The solution generation effect showed that students of different math abilities are in fact able to leverage their formal and intuitive resources to generate solutions even in the absence of any cognitive guidance or support. However, this only begged the question: might not guiding students during the generation and exploration phase result in an even better production of solutions, which in turn may help students learning even more from PF? In other words, what is the marginal gain of providing students with guidance during the generation and exploration phase?

In Kapur (2011), this question was addressed. Participants were 109, Secondary 1 (grade 7) students from a co-educational public school in Singapore. Students were from three mathematics classes taught by the same teacher. The participating school was a mainstream school comprising average-ability students on the grade six national standardized tests. The same study design as in Kapur (2010) was used except that in addition to the PF and DI conditions, a third condition—the guided generation condition—was added. One class was assigned to each condition. The guided-generation condition was exactly the same as the PF condition but with one important exception. Whereas students in the PF condition did not receive any form of cognitive guidance or support during the generation and exploration phase, students in the guided-generation condition were provided with cognitive support and facilitation throughout that process. Such guidance was typically in the form of teacher clarifications, focusing attention on significant issues or parameters in the problem, question prompts that engendered student elaboration and explanations, and hints towards productive solution steps (Hmelo-Silver et al., 2007; Puntambekar & Hübscher, 2005).

Findings suggested that students from the PF condition outperformed those from the DI and guided-generation conditions on procedural fluency, conceptual understanding and transfer. The differences between guided-generation and DI conditions were not significant, though students from the guided-generation condition performed marginally better than those from the DI condition. Overall, the descriptive trend $PF > \text{guided-generation} > DI$ seemed consistent across the different types of items.

It is noteworthy that Loibl and Rummel (2013) independently replicated this effect in a study with three similar conditions: unguided problem-solving prior to instruction, guided problem-solving prior to instruction in which students were supported with cognitive prompts during the problem-solving phase designed for them to notice critical features of the solutions they generated (e.g. “maybe there are situations where your solution does not work, have a

look at this counter-example”), and direct instruction. They found that in spite of guidance helping students generate better quality solutions, there was no significant difference between the guided and unguided problem-solving conditions, that is, cognitive guidance during the generation phase did not result in better learning on the posttest. A plausible reason consistent with the design theory of PF could be that both low and high quality solutions present opportunities to learn during the subsequent instruction, especially through a comparison and contrast between the student-generated solutions and the canonical solution.

Taken together, these findings perhaps suggest that giving cognitive guidance too early or in the process of generation does not add to the preparatory benefits of generation in part because students may not be ready to receive and make use of the guidance provided. This, of course, does not imply that all guidance is unnecessary or unproductive but that more research is needed to understand the types of guidance necessary to support the generation phase.

The role of generating versus studying and evaluating solutions

A critical mechanism embodied in the PF design is one of generation and exploration of solutions relying only on students’ formal and intuitive resources. However, it was not clear from the solution generation effect whether what was critical is the generation of solutions or simply an exposure to these solutions. Simply put, is it really necessary for students to generate the solutions or can these solutions be given to students to study and evaluate, that is, the opportunity to learn from the failed problem-solving efforts of their peers? In this paper, learning is referred to the failed problem-solving efforts of others as learning from *Vicarious Failure* (VF). If productive failure is a design where students have an opportunity to learn from their own failed solutions, then vicarious failure is a design where students have an opportunity to learn from the failed solutions of their peers.

In Kapur (2013), the effectiveness of learning from PF and VF was compared. Participants were one hundred and thirty six ($N = 136$) grade eight mathematics students (14-15 year olds) from two co-educational public schools in Singapore. Sixty-four students from School A and seventy-two students from School B participated in the study. In both the schools, students came from two intact classes taught by the same teacher. As per the PF design, PF students experienced the generation and exploration phase followed by the consolidation and knowledge assembly phase. VF students differed from the PF condition only in the first phase: The generation and exploration phase was replaced with a study and evaluate phase, where instead of generating and exploring solutions, students worked in small groups to study and evaluate student-generated solutions (available from earlier work, e.g., Kapur, 2012). VF students then received the same consolidation and knowledge assembly as PF students. In the study and evaluation phase for VF students, students first read the complex problem (see Appendix A) and were then presented with the student-generated solutions one-by-one counterbalanced for order with the prompt: "Evaluate whether this solution is a good measure of consistency. Explain and give reasons to support your evaluation." The number of solutions was pegged to the average number of solutions produced by PF groups, that is, six. The most frequently-generated solutions by the PF students were chosen for VF condition.

Findings suggested that PF students significantly outperformed VF students on conceptual understanding and transfer, without compromising procedural fluency. It is noteworthy that an earlier study by Roll and colleagues (Roll, Alevan, & Koedinger, 2011) found that students who generated and evaluated solutions in an intelligent tutoring environment performed better on conceptual understanding than students who evaluated given solutions. However, Roll and colleagues did not find any significant differences on transfer measures. In more recent work (Kapur, 2014), learning from PF, VF, and DI in a

randomized controlled experimental study with students working alone (as opposed to working in groups) was compared, findings show not only that learning from PF is better than VF (that is, consistent with Kapur (2013) and Roll et al. (2011)) but also that learning from VF is better than DI.

These findings provide evidence for the design theory of PF that generation of solutions produces better activation and differentiation of knowledge than an exposure to and evaluation of those solutions, and further that activation and differentiation afforded by VF is still better than that afforded by direct instruction. In other words, these findings simply suggest that when learning a new concept, we seem to learn better from our own failed solutions than those of others' although, absent the opportunity to learn from our own failures, we are better off trying to learn from others' failed solutions than from direct instruction. Taken together, these findings collectively underscore the primacy of generation over mere exposure, thereby evidencing a key mechanism of the PF design.

The role of attention to critical features

As discussed earlier, the contrasts among and between the student-generated solutions and the canonical solutions afford students the opportunities to attend to the critical features of the targeted concept. However, if what is essential is that students attend to the ten critical features, then why not simply tell students these critical features? Why bother having them generate, and compare and contrast the solutions? Simply put, do students really need to generate before receiving the critical features, or would telling the critical features without any generation work just as well? Addressing this question would help understand a critical mechanism of PF that the generation and exploration of solutions better prepares students to understand the critical features during instruction than simply telling them those features.

In Kapur and Bielaczyc (2011), this question was addressed. Participants were 57, ninth-grade mathematics students (14-15 year olds) from two intact classes in an all-boys

public school in Singapore. One class was assigned to the PF condition, and the other class to the ‘Strong-DI’ condition. Both classes were taught by the same teacher. The PF condition was exactly the same as in Kapur (2012). The Strong-DI condition was the same as the DI condition in Kapur (2012) except that the teacher drew attention to the ten critical features during instruction (e.g., why deviations need to be taken from the mean, why they must be positive, why divide by n , etc.). While explaining each step of formulating and calculating SD, the teacher explained the appropriate critical features relevant for that step. For example, when explaining the concept of “deviation of a point from the mean”, the teacher discussed why deviations need to be from a fixed point, why the fixed point should be the mean, and why deviations must be positive. During subsequent problem solving and feedback, the teacher repeatedly reinforced these critical features throughout the lessons.

Findings suggested that PF students significantly outperformed their Strong-DI counterparts on conceptual understanding without compromising on procedural fluency. There were no differences in terms of transfer although a recent, randomized controlled experiment (Kapur, 2014) demonstrated significant effects on transfer as well. What students seemed to be learning from PF is an understanding of the critical features of the concept, and why it is formulated the way it is. Simply telling and explaining these features does not seem to be effective if students have not had the opportunity to activate and differentiate their prior knowledge, and engage in the comparison and contrasting of the solutions to attend to the critical features of the concept.

Future Research Directions

Going forward, there are several lines of research for future work:

- i. Generating solutions to novel problems can be seen as one way of preparing to learn from subsequent instruction (e.g., DeCaro & Rittle-Johnson, 2012; Kapur, 2014). Engaging in anchored instruction (e.g., Cognition and Technology Group at Vanderbilt, 1992) or

Problem-based Learning (PBL) before a lecture on the requisite concepts is another (e.g., Capon & Kuhn, 2004). Comparing and contrasting cases is still another kind of a preparatory activity (e.g., Schwartz et al., 2011). What might some other types of preparatory activities be? One such preparatory activity is vicarious failure wherein students generate evaluations of solutions, and previous work has suggested that vicarious failure was better than direct instruction but not as effective as productive failure (Kapur, 2013, 2014). Might engaging students in generating questions with or without solutions—commonly known as problem finding or problem posing—be also a productive preparatory activity? Are all these preparatory activities equally effective? Are there conditions under which their effectiveness varies relative to each other? These are some of the questions that are open for future research.

- ii. A related line of work concerns the design of support for generative activities. In PF, cognitive or domain-specific support is not provided to guide students. However, this does not preclude other kinds of support that may be useful during the generation phase. For example, domain-general support in the form of metacognitive scaffolds has been shown to be effective in the generation phase (Roll et al., 2012). Likewise, Westermann and Rummel (2012) provided a peer-interaction role-play script (2012) to support collaborative interaction during the generation phase, and Kapur (2012) provided affective support to encourage students to persist in the generative task. All of these support measures were designed to help students, but without giving or telling them the targeted content knowledge. In other words, an open area for research is to design and examine the efficacy of support structures during the generation phase to keep students from unproductive failure experiences, while leaving a central portion of student generation unstructured, thus allowing for productive failure (Kapur, 2012).

- iii. There is also much work needed in the design of the consolidation phase (Collins, 2012). The goal of such research would be to explore the most effective way of carrying out the consolidation phase, especially ways of organizing the comparison and contrasting of solutions, and assembling them into the canonical structures.
- iv. The variation within PF groups also needs to be unpacked. While, on average, PF groups exhibited productive failure, naturally some “failed” more than others in the shorter term inasmuch as some gained more than others in the longer term. What explains this variation *within* the PF condition? Examining learner characteristics as well as the nature of interactional behaviors and relating them to eventual gains in group and individual performance would be most insightful (Kapur et al., 2005, 2006).
- v. Related to the point above is the role of affective mechanisms in PF. In addition to the cognitive benefits of PF, future work should also examine the how PF affords students the opportunities to develop greater agency, motivation, and interest, among other affective constructs. Moreover, the development of epistemic knowledge and resources also merits further analytical work (Bielaczyc & Kapur, 2010).
- vi. Another avenue for future work is to examine the relative effect of different types of student-generated solutions on learning. If, as indeed the solution generation effect suggested, the ability to design multiple solutions is associated with learning from PF, are qualitative solutions (e.g., graphs) more strongly associated with learning than quantitative solutions (e.g., computing averages and differences), or vice versa? If so, does the sequence in which students generate such solutions matter? Examining these questions forms the focus of my continuing analyses and experimental work.
- vii. Thus far, the efficacy of PF has been demonstrated in the domains of math and science. Might PF also work for less structured domains such as the languages, humanities, and

arts? While one could conjecture that principles embodied in the PF design are not domain-dependent, empirical evidence is needed to test the conjecture.

- viii. Finally, examining productive failure from the teachers' perspective, their motivation, content knowledge, and beliefs about learning and knowing would also be important. Concomitantly, future work could also examine how pre-service and in-service programs could be designed to develop teachers' capacity in using PF. For example, one professional development model could be to design such programs using the PF method. Doing so could not only help teachers experience PF but also better understand and learn how to design using PF.

In sum, future work and analyses at multiple grain sizes at the level of the task, the activity structures, and the social surround with both students and teachers might speak to these concerns and add further explanatory power to productive failure.

General Discussion

By combining the benefits of generation and exploration with that of (and followed by) direct instruction, findings from the program of research on PF contribute to the ongoing and persistent debate comparing the effectiveness of direct instruction with discovery learning approaches (see Tobias & Duffy, 2010) by pointing to a move away from simplistic comparisons between pure discovery learning and direct instruction.

Proponents of direct instruction continue to epitomize discovery learning or pure problem solving as the constructivist ideal (e.g., Kirschner et al., 2006; Klahr & Nigam, 2004). They bring to bear substantive empirical evidence against un-guided or minimally-guided instruction to claim that there is little efficacy in having learners solve problems that target novel concepts, and that learners should receive direct instruction on the concepts before any problem solving. Perhaps this view is best captured by Sweller (2010), "What can conceivably be gained by leaving the learner to search for a solution when the search is

usually very time consuming, may result in a suboptimal solution, or even no solution at all?” (p. 128). The basis for this view comes from a large body of empirical evidence that has compared some form of heavily-guided direct instruction (e.g., worked examples) favorably with unguided or minimally-guided discovery learning instruction. This led Kirschner et al. (2006) to argue that “Controlled experiments almost uniformly indicate that when dealing with novel information, learners should be explicitly shown what to do and how to do it” (p. 79). It is of course not surprising that learners do not learn from unguided or minimally-guided discovery learning when compared with a heavily-guided direct instruction. However, the following two inferences do not logically follow.

The first inference that does not logically follow is that direct instruction is the most effective method of instruction for teaching new concepts. The superiority of direct instruction over pure problem solving or unguided discovery learning does not mean that there do not exist other methods that are just as or even more effective than direct instruction. One has to look no further than Mayer’s (2004) argument against discovery learning. In the process of reviewing studies that compared unguided discovery with guided discovery and direct instruction, Mayer (2004) ended up showing that both unguided discovery and direct instruction were not the most effective for learning; guided discovery outperformed both unguided discovery and direct instruction conditions. For example, Kittel (1957) compared learning of logic problems (e.g., finding the rule for excluding a word from a group of words) through pure discovery, guided discovery (where a hint was given) and direct instruction (where the rule was told first), and found guided discovery to be the best for retention and transfer. Gagne and Brown’s (1961) study echoed similar findings in the learning of math where students had to learn how to derive formulas to sum a number series. There are other methods too such as PBL whose effectiveness over direct instruction has been demonstrated though these effects are not always consistent, in part due to how PBL is designed for

differently in different studies (see Hmelo-Silver et al., 2007 for a relevant review of such studies). Finally, research on IPL and PF also represent teaching methods that are more effective than direct instruction.

The second inference that does not logically follow is that there is little efficacy in having learners solve problems that target concepts they have not learned yet—something that they have to do in unguided discovery learning or pure problem solving. To determine if there such an efficacy, a stricter comparison for direct instruction is needed in which one would compare it with an approach where students first engage in pure problem solving on their own followed by direct instruction. As the findings on PF have shown, the problem solving process initially leads to a failure in generating correct solutions. Yet, this very process can be productive for learning provided direct instruction on the targeted concepts is subsequently provided. The solution generation effect only further bolsters the efficacy of learners solving problems that target concepts they have not learned yet because findings showed that the more solutions they generate, the more they seem to learn from the subsequent direct instruction.

At a theoretical level, it is perhaps worth clarifying that a commitment to a constructivist epistemology does not necessarily imply a commitment to discovery learning (Hmelo-Silver et al., 2007; Mayer, 2004). Simply leaving learners to generate and explore without consolidating is unlikely to lead to learning, or at least learners cannot be expected to “discover” the canonical representations by themselves as indeed my findings suggest. Instead, a commitment to a constructivist epistemology requires that we *build* upon learners’ prior knowledge. However, one cannot build upon prior knowledge if one does not know what this prior knowledge is in the first place.

Importantly, this emphasis on learners’ prior knowledge is critical even from the perspective of cognitive load theory (CLT) (Sweller, 1988)—the theoretical basis that

grounds the argument for direct instruction. According to CLT, the processing of novel information depends upon an interaction between a limited working memory capacity and relevant information stored in the long term memory. Therefore, Kirschner et al. (2006) argued that, “Any instructional theory that ignores the limits of working memory when dealing with novel information or ignores the disappearance of those limits when dealing with familiar information is unlikely to be effective” (p. 77).

If what a learner already knows—prior knowledge—about a concept is a critical determinant of either limiting or expanding the working memory (WM) capacity as conceptualized by CLT, then does not a commitment to CLT entail a commitment to understanding whether and to what extent the targeted concept is novel to the learner? And herein lies the problem: How does one determine what is novel to the learner when any inference from learners’ performance is inextricably tied to the nature of the task and context? If one defines novelty via the canonical lens, that is, the correct established normative way of thinking and performing in a domain, then it is easy to assume that learners do not have any prior knowledge. Consequently, one is constrained to work within the limiting aspects of the WM, which is what the proponents of direct instruction largely seem to have done (e.g., Carroll, 1994; Sweller & Chandler, 1991; Paas, 1992).

However, if we allow for the concomitant possibility that learners may have some prior knowledge and resources about a concept they have yet to learn, could we not design tasks and activity structures to elicit this knowledge? Many of the studies cited in this paper demonstrate that it is possible to design generation tasks based on the assumption that even though students may not have canonical knowledge about the targeted concept, this does not mean that they do not have or cannot generate some ideas, albeit suboptimal or even incorrect, about that concept. The challenge really is to design such problem-solving contexts that can activate and elicit students’ thinking about the concept even if they have not formally

learned it yet. To the extent that one can accomplish this, it follows that by activating and working with these priors in the long-term memory, one can leverage the expandable aspects of the WM capacity. At the very least, this is a theoretical possibility that the CLT allows for.

Therefore, whether one makes commitment to the information processing perspective undergirding the CLT or the constructivist perspective, the primacy of what a learner already knows—formally or intuitively—seems to be common and important to both. It follows that at the very least the burden on the designer is to first understand the nature of learners' prior knowledge structures; the very structures upon which the claimed “building” will be done. Designing for productive failure presents one way of doing so, wherein students first generate and explore solutions, and in the process externalize their prior knowledge structures, before direct instruction.

References

- Bielaczyc, K., & Kapur, M. (2010). Playing epistemic games in science and mathematics classrooms. *Educational Technology*.
- Bruner, J. S. (1986). *Actual minds, possible worlds*. Cambridge MA: Harvard University Press.
- Chi, M. T. H., Glaser, R., & Farr, M. J. (1988). *The nature of expertise*. Hillsdale, NJ: Erlbaum.
- 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.
- Dillenbourg, P. (2002). Over-scripting CSCL: The risks of blending collaborative learning with instructional design. *Three worlds of CSCL. Can we support CSCL?*, 61-91.
- English, A. R. (2013). *Discontinuity in Learning: Dewey, Herbart and Education as Transformation*. Cambridge University Press
- Gagne, R. M., & Brown, L. T. (1961). Some factors in the programming of conceptual learning. *Journal of Experimental Psychology*, 62, 313–321.
- Hiebert, J., & Grouws, D. A. (2007). The effects of classroom mathematics teaching on students' learning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 371-404). Charlotte, NC: 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.
- Kapur, M. (2014). Productive failure in learning math. *Cognitive Science*.
DOI: 10.1111/cogs.12107

- Kapur, M. (2013). Comparing learning from productive failure and vicarious failure. *The Journal of the Learning Sciences*. DOI: 10.1080/10508406.2013.819000
- Kapur, M. (2012). Productive failure in learning the concept of variance. *Instructional Science*, 40(4), 651-672.
- 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. (2010). Productive failure in mathematical problem solving. *Instructional Science*, 38(6), 523-550.
- Kapur, M. (2008). Productive failure. *Cognition and Instruction*, 26(3), 379-424.
- Kapur, M., & Bielaczyc, K. (2012). Designing for productive failure. *The Journal of the Learning Sciences*, 21(1), 45-83.
- Kapur, M., & Bielaczyc, K. (2011). *Classroom-based experiments in productive failure*. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp.x-y). Austin, TX: Cognitive Science Society.
- Kapur, M., & Kinzer, C. (2009). Productive failure in CSCL groups. *International Journal of Computer-Supported Collaborative Learning (ijCSCL)*, 4(1), 21-46.
- Kapur, M., & Rummel, N. (2012). Productive failure in learning and problem solving. *Instructional Science*. 40 (4), 645-650.
- Kapur, M., & Rummel, N. (2009). The assistance dilemma in CSCL. In A. Dimitracopoulou, C. O'Malley, D. Suthers, & P. Reimann (Eds.), *Computer Supported Collaborative Learning Practices- CSCL2009 Community Events Proceedings, Vol 2* (pp. 37-42). International Society of the Learning Sciences.
- Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work. *Educational psychologist*, 41(2), 75-86.

- Kapur, M., Voiklis, J., Kinzer, C., & Black, J. (2006). Insights into the emergence of convergence in group discussions. In S. Barab, K. Hay, & D. Hickey (Eds.), *Proceedings of the International Conference on the Learning Sciences* (pp. 300-306). Mahwah, NJ: Erlbaum.
- Kapur, M., Voiklis, J., & Kinzer, C. (2005). Problem solving as a complex, evolutionary activity: A methodological framework for analyzing problem-solving processes in a computer-supported collaborative environment. *Proceedings the Computer Supported Collaborative Learning (CSCL) Conference* (pp. 252-261). Mahwah, NJ: Erlbaum.
- Kittel, J. E. (1957). An experimental study of the effect of external direction during learning on transfer and retention of principles. *Journal of Educational Psychology*, 48, 391–405.
- Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction: Effects of direct instruction and discovery learning. *Psychological Science*, 15 (10), 661–667.
- Loibl, K., & Rummel, N. (2013). The impact of guidance during problem-solving prior to instruction on students' inventions and learning outcomes. *Instructional Science*. Advance online publication. doi: 10.1007/s11251-013-9282-5
- Mayer, R. (2004). Should there be a three-strikes rule against pure discovery learning? The case for guided methods of instruction. *American Psychologist*, 59, 14–19.
- Paas, F. (1992). Training strategies for achieving transfer of problem-solving skill in statistics: A cognitive load approach. *Journal of Educational Psychology*, 84 (4), 429-434.
- Puntambekar, S., & Hübscher, R. (2005). Tools for scaffolding students in a complex learning environment: What have we gained and what have we missed? *Educational Psychologist*, 40(1), 1-12.

- Roll, I., Alevan, V., & Koedinger, K. (2011). Outcomes and mechanisms of transfer in invention Activities. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 2824-2829). Austin, TX: Cognitive Science Society.
- 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., 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.
- Sweller, J. (2010). What human cognitive architecture tells us about constructivism. In S. Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12, 257–285.
- Sweller, J., & Chandler, P. (1991). Evidence for cognitive load theory. *Cognition and Instruction*, 8 (4), 351-362.
- Tobias, S., & Duffy, T. M. (2010), *Constructivist instruction: Success or failure*. New York, NJ: Routledge.
- Westermann, K. & Rummel, N. (2012). Delaying Instruction – Evidence from a Study in a University Relearning Setting. *Instructional Science*, 40(4), 673-689.

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, 17(2), 89-100.

APPENDIX A

The Complex Problem Scenario

The organizers of the Premier League Federation have to decide which one of the three players - Mike Arwen, Dave Backhand and Ivan Right – should receive the “The Most Consistent Player for the Past 20 Years” award. Table 1 shows the number of goals that each striker scored between 1992 and 2011.

Table 1: Number of goals scored by the three strikers in the Premier League between 1992 & 2011

Year	Mike Arwen	Dave Backhand	Ivan Right
1992	14	13	13
1993	9	9	18
1994	14	16	15
1995	10	14	10
1996	15	10	16
1997	11	11	10
1998	15	13	17
1999	11	14	10
2000	16	15	12
2001	12	19	14
2002	16	14	19
2003	12	12	14
2004	17	15	18
2005	13	14	9
2006	17	17	10
2007	13	13	18
2008	18	14	11
2009	14	18	10
2010	19	14	18
2011	14	15	18

The organizers agreed to approach this decision *mathematically* by designing a measure of consistency. They decided to get your group’s help. Here is what you must do:

- (1) Design as many different measures of consistency as you can.
- (2) Your measure of consistency should make use of all data points in the table.
- (3) Show all working and calculations on the papers provided.