Chapter 12
Learning from Productive Failure

Manu Kapur and Leslie Toh

Abstract Situating our work within the constructivist debate about effective ways of designing for learning, we describe our program of research on productive failure (PF). The PF learning design affords students opportunities to engage in authentic mathematical practice where they start by generating and exploring solutions to a novel design problem followed by consolidation and knowledge assembly. In doing so, PF affords students opportunities to activate and differentiate their prior knowledge, so that they are better prepared to attend to and learn the critical conceptual features of the targeted concepts during the subsequent instruction. Our 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 in brief wherein key aspects of the productive failure design were tested over multiple classroom-based studies in Singapore public schools and how these studies helped us interrogate and understand the criticality of key mechanisms embodied in the PF design.

Keywords Productive failure • Authentic practice • Mathematics

Introduction

Proponents of direct instruction bring to bear substantive empirical evidence against unguided 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 (Sweller 2010; Kirschner et al. 2006). Kirschner et al. (2006) argued 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). Commonly cited problems with unguided or minimally guided instruction include increased working

M. Kapur (✉) • L. Toh
Learning Sciences Laboratory, National Institute of Education, Nanyang Technological University, Singapore, Singapore
e-mail: manu.kapur@nie.edu.sg

Y.H. Cho et al. (eds.), *Authentic Problem Solving and Learning in the 21st Century*, Education Innovation Series, DOI 10.1007/978-981-287-521-1_12

memory load that interferes with schema formation (Sweller 1988), encoding of errors and misconceptions (Brown and Campione 1994), lack of adequate practice and elaboration (Klahr and Nigam 2004), as well as affective problems of frustration and de-motivation (Hardiman et al. 1986).

Consequently, this has led to a commonly held belief that there is little efficacy in having learners solve novel problems that target concepts they have not learned yet. Perhaps this belief 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 belief 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 (Kirschner et al. 2006). 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 conclusion 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—does not follow.

To determine if there is such an efficacy, a stricter comparison for direct instruction would be to compare it with an approach where students first generate representations and methods to novel problems on their own followed by direct instruction. It can be expected that the generation process will likely lead to failure. By failure, I simply mean that students will not be able to develop or discover the canonical solutions by themselves. Yet, what is critical is not the failure to develop the canonical solution per se but the very process of generating and exploring multiple representations and solution methods, which can be productive for learning *provided* that direct instruction on the targeted concepts is subsequently provided (Kapur and Bielaczyc 2012; Kapur and Rummel 2009; Schwartz and Martin 2004).

This chapter reports on a program of research that explores the possibility of affording learners the opportunity to engage in a process of generating solutions to novel problems and shows how this process invariably leads to suboptimal solutions (i.e., failure to generate the canonical solutions) but can still be a productive exercise in failure provided that some form of direct instruction follows (Kapur 2010, 2011, 2012, 2014, 2015). Thus argued, instead of reporting experiments comparing discovery learning with direct instruction, the work presented herein seeks to understand whether combining the two—as instantiated in the learning design called *productive failure* (Kapur and Bielaczyc 2012)—can be more effective than direct instruction alone.

We start with a brief review of research that supports the case for productive failure and points to an efficacy of learner-generated solutions provided that an appropriate form of direct instruction builds upon it. Next, we provide a brief description of the mechanisms embodied in the design principles of productive failure. Following this, we describe a program of design research wherein key aspects of the productive failure design were tested over multiple classroom-based studies in Singapore public schools. Our aim is not to describe each study in detail. Instead, it is to articulate the underlying logic of how the various studies help us test and understand some of the critical design decisions of PF.

The Case of Failure in Learning and Problem-Solving

Research on *impasse-driven learning* (Van Lehn et al. 2003) with college students in coached problem-solving situations provides strong evidence for the role of failure in learning. Successful learning of a principle (e.g., a concept, a physical law) was associated with events when students reached an impasse during problem-solving. Conversely, when students did not reach an impasse, learning was rare despite explicit tutor explanations of the target principle. Instead of providing immediate or direct instruction upfront, e.g., in the form of feedback, questions, or explanations, when the learner demonstrably makes an error or is "stuck," Van Lehn et al.'s (2003) findings suggest that it may well be more productive to delay that instruction up until the student reaches an impasse—a form of failure—and is subsequently unable to generate an adequate way forward.

Building on this, Mathan and Koedinger (2003) compared learning under two different feedback conditions on student errors. In the immediate feedback condition, a tutor gave immediate feedback on student errors. In the delayed feedback condition, the tutor allowed the student to detect their own error first before providing feedback. Their findings suggested that students in the delayed feedback condition demonstrated a faster rate of learning from and on all the subsequent problems. Delayed feedback on errors seemed to have resulted in better retention and better preparation to learn from subsequent problems (Mathan and Koedinger 2003).

Further evidence for such *preparation for future learning* (PFL; Schwartz and Bransford 1998) can be found in the *inventing to prepare for learning* (IPL) research by Schwartz and Martin (2004). In a sequence of design experiments on the teaching of descriptive statistics with intellectually gifted students, Schwartz and Martin (2004) demonstrated an existence proof for the hidden efficacy of invention activities when such activities preceded direct instruction, despite such activities failing to produce canonical conceptions and solutions during the invention phase. However, the proponents of direct instruction have criticized PFL and IPL studies because of a lack of adequate control and experimental manipulation of one variable at a time, which makes it difficult to make causal attributions of the effects (Kirschner et al. 2006).

Earlier experiments in *productive failure* (Kapur 2008) provide evidence from randomized-controlled experiments for the role of failure in learning and problem by delaying structure. Kapur (2008) examined students solving complex problems without the provision on any external support structures or scaffolds. 11th-grade student triads from seven high schools in India were randomly assigned to solve either ill- or well-structured physics problems in an online, chat environment. After group problem-solving, all students individually solved well-structured problems followed by ill-structured problems. Ill-structured groups generated a greater diversity of representations and methods for solving the ill-structured problems. However, ill-structured group discussions were found to be more complex and divergent than those of their well-structured counterparts, leading to poor group performance (Kapur et al. 2005, 2006, 2007). Notwithstanding, findings suggested a hidden efficacy in the complex, divergent interactional process even though it seemingly led to failure. Kapur argued that delaying the structure received by students from the ill-structured groups (who solved ill-structured problems collaboratively followed by well-structured problems individually) helped them discern how to structure an ill-structured problem, thereby facilitating a spontaneous transfer of problem-solving skills. Findings from this study have since been replicated (Kapur and Kinzer 2009).

These findings are consistent with other research programs that suggest that conditions that maximize performance in the shorter term are not necessarily the ones that maximize learning in the longer term (Clifford 1984; Schmidt and Bjork 1992). Collectively, it is reasonable to reinterpret their central findings as all of them point to the efficacy of learner-generated processing, conceptions, representations, and understandings, even though such conceptions and understandings may not be correct initially and the process of arriving at them not as efficient. The above findings, while preliminary, underscore the implication that by delaying instructional support—be it explanations, feedback, direct instruction, or well-structured problems—in learning and problem-solving activities so as to allow learners to generate solutions to novel problems can be a productive exercise in failure (Kapur 2008).

More than simply indicating a delay of instructional structure, these studies also underscore the presence of desirable difficulties and productive learner activity in solving problems. It is this interest in what is present, that is, the features of productive learner activity (even if it results in "failure"), that forms the core of our work. Based on the literature and our own studies in PF, we have begun to develop a design theory of what needs to be present in student problem-solving contexts in which instructional structure is delayed. We are interested in testing our theoretical conjectures by investigating their embodiment in the design of problem-solving experiences that, although leading to short-term performance failure, are efficacious in the longer term. We briefly describe these design principles and the theoretical conjectures they embody next (for a fuller description, see Kapur and Bielaczyc 2012).

Designing for Productive Failure (PF)

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 and Bielaczyc 2012; Schwartz and Bransford 1998; Schwartz and 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 et al. 1988; Schwartz and Bransford 1998).

Cognizant of these two problems, PF engages students in a learning design (for a fuller explication of the design principles, see Kapur and 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 (RSMs). Phase 2 affords opportunities for organizing and assembling the relevant student-generated RSMs into canonical RSMs. 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 RSMs (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 suboptimal RSMs and the assembly of canonical RSMs (mechanisms b–d).

The PF design also undertakes a commitment that there is more to learning mathematics than just *learning about* mathematics, which is necessary but not sufficient. Part of learning mathematics, and arguably the more important part perhaps, is to engage in the *authentic* practice of mathematics akin to that of mathematicians. This involves *learning to be* like a member of the mathematical community (Thomas and Brown 2007). But what does authentic mathematical practice entail? Inventing representational forms, developing domain-general and specific methods, flexibly adapting and refining or inventing new representations and methods when others do not work, critiquing, elaborating, explaining to each other, and persisting in solving problems define the epistemic repertoire of authentic mathematical practice (Bielaczyc and Kapur 2010; Bielaczyc, Kapur and Collins 2013; diSessa and Sherin 2000). Learning to be like a mathematician is to learn and do what mathematicians do; it involves a "mathematical" way looking at the world, understanding the constructed nature of mathematical knowledge, and persisting in participating in the construction and refinement of mathematical knowledge. Learning to be, therefore, clearly foregrounds the epistemological aspects of authentic mathematical practice. Needless to say, both learning about and learning to be are important commitments, but the latter remains much neglected in comparison to the former. The epistemological commitments of PF aim to redress this imbalance and thus engage the learner in authentic learning and practice of mathematics.

Examining the PF Design in the Real Ecologies of Singapore Classrooms

Having articulated the mechanisms embodied in the design principles of PF, we now describe the implementation in a series of classroom-based experiments. To bring about change in classroom practice and pedagogy, especially in a system of high-stakes testing such as Singapore, it was important to compare a new learning design (e.g., PF) with a design most prevalent in practice (e.g., DI). Thus, we started by comparing learning from PF with DI.

Comparing PF with DI

We illustrate a comparison of learning from PF and DI through a pre-posttest, quasi-experimental study (hereinafter referred to as Study 1) with 133, ninth-grade mathematics students (14–15-year-olds) from a public school in Singapore (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-min 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 20-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 were 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), we have described these student-generated solutions in greater detail. For the present purposes, we only briefly describe the four categories of solutions:

(a) *Central tendencies* (e.g., using mean, median, mode)
(b) *Qualitative methods* (e.g., organizing data using dot diagrams, frequency polygons, line graphs to examine clustering and fluctuations' patterns)
(c) *Frequency methods* (e.g., counting the frequency with which a player scored above, below, and at the mean to argue that the greater the frequency at the mean relative to away from the mean, the better the consistency)
(d) *Deviation methods* (e.g., range; calculating the sum of year-on-year deviations to argue that the greater the sum, the lower the consistency; calculating absolute deviations to avoid deviations of opposite signs canceling each other; calculating the average instead of the sum of the deviations).

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. This was not surprising given that they had been taught the canonical formulation of SD, which is also easy to compute and apply. All DI students were accurately able to apply the concept of SD to solve the very problem that the PF students tried to generate a solution to.

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). 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 the students generated, the better they performed on the procedural fluency, conceptual understanding, and transfer items on the posttest. We refer to this effect as the *solution generation effect*.

Discussion

These findings are consistent with the seminal studies on productive failure (Kapur 2008; Kapur and Kinzer 2009) and also with other studies described earlier (e.g., Schwartz and Bransford 1998; Schwartz and Martin 2004). These findings suggest that there is in fact a utility in having students solve novel problems first. To explain these findings, we argued 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. Whereas PF students were afforded opportunities to work with not only the solutions that they generated but also the canonical solutions that they received during direct instruction, DI students worked with only the canonical ones. Hence, DI students worked with a smaller number of solutions, and consequently, their knowledge was arguably not as differentiated as their PF counterparts.

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 following critical features of the targeted concept that are necessary to develop a deep understanding of the concept. Granted that student-generated solutions are at best an indirect measure of prior knowledge activation and differentiation, it was nonetheless a critical difference between the two conditions by design. Importantly, this difference needs to be situated in the argument made by the proponents of DI in their questioning of the utility of getting students to generate solutions to solve novel problems on their own. They argue that students should be given the canonical solutions (either through worked examples or direct instruction) before getting them to apply these to solve problems on their own (Sweller 2010).

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, we describe four such lines of inquiry, each testing a critical aspect of the PF design. Once again, fuller descriptions of these studies can be found in our published work, and therefore, our 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, coeducational 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 and demonstrated the tractability of PF across a range of math ability provided that 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), we addressed this question. Participants were 109, secondary 1 (grade 7) students from a coeducational 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 Study 1 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 toward productive solution steps.

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 > guided-generation > LP seemed consistent across the different types of items. We argued that giving 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.

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? We refer to learning from 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), we compared the effectiveness of learning from PF and VF. Participants were one hundred and thirty six ($N = 136$) grade eight mathematics students (14–15-year-olds) from two coeducational public schools in Singapore. Sixty four students from School A and seventy two students from School B participated in the study. In both 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 evaluation 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; see Kapur 2013 for examples of solutions). 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, after controlling for prior knowledge, school, and ability differences, PF students significantly outperformed VF students on conceptual understanding and transfer, without compromising procedural fluency. These findings underscored the primacy of generation over mere exposure, thereby evidencing a key mechanism of the PF design. In more recent work (Kapur 2014), we have compared PF, VF, and DI and shown the findings to be consistent with Kapur (2013).

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 for instruction than simply telling them those features.

In Kapur and Bielaczyc (2011), we addressed this question. 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 Study 1. The Strong-DI condition was the same as the DI condition in Study 1 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. These findings suggested that although telling students that novel information can be effective, the generation and exploration phase is nonetheless better in preparing students to receive these features.

Conclusion

Contrary to the commonly held belief that there is little efficacy in having learners solve novel problems that target concepts they have not learned yet, 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. Our work also demonstrates how engaging students in the process of generating, exploring, critiquing, and refining solutions affords them with the opportunity to engage in authentic practice. Authenticity refers not so much to the actual task or problem but the context and culture within which such problem-solving occurred that afforded students opportunities to not only learn about mathematics but also be like a mathematician (Thomas and Brown 2007).

In this chapter, we traced the developmental trajectory of PF from its inception to a learning design. We started by describing the mechanisms embodied in the PF design, as well as the principles guiding the design. Our initial work in the schools compared the PF design with the most prevalent design in classroom instruction, that is, DI. Findings from an initial comparison between PF and DI were encouraging yet raised further lines of inquiry that necessitated a closer examination of some critical aspects of the PF design, namely, (a) the role of math ability, (b) the role of guidance during the generation, (c) the role of learning from vicarious failure, and (d) the role of attention to critical features. Each of these lines of inquiry was pursued through classroom-based quasi-experimental studies.

Thus far, our work has focused on a closer interrogation of the design to more systematically unpack and examine its design assumption and decisions. Through such an "iterative" examination in real ecologies, our goal for the PF learning design is to become more "ecologically valid and practice-oriented" (Confrey 2006, p. 144). More importantly, the iterative examination of the design further generates theoretical conjectures that in turn drive future work. In other words, the continuous examination of the design enables the development of possible design principles that direct, apprise, and advance educational research and practice (Anderson and Shattuck 2012). Therefore, our future work would continue to interrogate the PF design and all its constituent mechanisms, design principles, and design decisions, while at the same time iterate and refine the PF design.

Acknowledgments The work reported in this chapter was funded by grants from the Ministry of Education of Singapore. This chapter has been reproduced with some adaptation and update, and

in accordance with the publishing agreement, from a chapter (Kapur and Toh 2013) that was contributed to a handbook of educational design cases.

Appendix A: The Complex Problem Scenario

Mr. Ferguson, Mr. Merino, and Mr. Eriksson are the managers 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 12.1 shows the number of goals that each striker had scored between 1988 and 2007.

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.

Table 12.1 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

References

Anderson, T., & Shattuck, J. (2012). Design-based research: A decade of progress in education research? *Educational Researcher, 41*(1), 16–25.

Brown, A., & Campione, J. (1994). Guided discovery in a community of learners. In K. McGilly (Ed.), *Classroom lessons: Integrating cognitive theory and classroom practice* (pp. 229–270). Cambridge, MA: MIT Press.

Bielaczyc, K., & Kapur, M. (2010). Playing epistemic games in science and mathematics classrooms. *Educational Technology, 50*(5), 19–25.

Bielaczyc, K., Kapur, M., & Collins, A. (2013). Building communities of learners. In C. E. Hmelo-Silver, A. M. O'Donnell, C. Chan, & C. A. Chinn (Eds.), *International handbook of collaborative learning* (pp. 233–249). New York: Routledge.

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

Clifford, M. M. (1984). Thoughts on a theory of constructive failure. *Educational Psychologist, 19*(2), 108–120.

Confrey, J. (2006). The evolution of design studies as methodology. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 135–151). New York: Cambridge University Press.

diSessa, A. A., & Sherin, B. L. (2000). Meta-representation: An introduction. *The Journal of Mathematical Behavior, 19*, 385–398.

Hardiman, P., Pollatsek, A., & Weil, A. (1986). Learning to understand the balance beam. *Cognition and Instruction, 3*, 1–30.

Kapur, M. (2008). Productive failure. *Cognition and Instruction, 26*(3), 379–424.

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. (2013). Comparing learning from productive failure and vicarious failure. *The Journal of the Learning Sciences, 23*(4), 651–677.

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

Kapur, M. (2015). The preparatory effects of problem solving versus problem posing on learning from instruction. *Learning and Instruction, 39*, 23–31.

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. 2812–2817). Austin: Cognitive Science Society.

Kapur, M., & Bielaczyc, K. (2012). Designing for productive failure. *The Journal of the Learning Sciences, 21*(1), 45–83.

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. (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). Sydney: International Society of the Learning Sciences.

Kapur, M., & Toh, P. L. L. (2013). Productive failure: From an experimental effect to a learning design. In T. Plomp & N. Nieveen (Eds.), *Educational design research – Part B: Illustrative cases* (pp. 341–355). Enschede: SLO.

Kapur, M., Voiklis, J., & Kinzer, C. (2005, June). Problem solving as a complex, evolutionary activity: A methodological framework for analyzing problem-solving processes in a computer-supported collaborative environment. In *Proceedings the Computer Supported Collaborative Learning (CSCL) conference* (pp. 252–261). Mahwah: Erlbaum.

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: Erlbaum.

Kapur, M., Hung, D., Jacobson, M., Voiklis, J., Kinzer, C., & Chen, D.-T. (2007). Emergence of learning in computer-supported, large-scale collective dynamics: A research agenda. In C. A. Clark, G. Erkens, & S. Puntambekar (Eds.), *Proceedings of the international conference of computer-supported collaborative learning* (pp. 323–332). Mahwah: Erlbaum.

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

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.

Mathan, S., & Koedinger, K. (2003). Recasting the feedback debate: Benefits of tutoring error detection and correction skills. In U. Hoppe, F. Verdejo, & J. Kay (Eds.), *Artificial intelligence in education: Shaping the future of education through intelligent technologies* (pp. 13–20). Amsterdam: Ios Press.

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.

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

Sweller, J. (2010). What human cognitive architecture tells us about constructivism. In S. Tobias & T. M. Duffy (Eds.), *Constructivist instruction: Success or failure* (pp. 127–143). New York: Routledge.

Thomas, D., & Brown, J. S. (2007). The play of imagination: Extending the literary mind. *Games and Culture, 2*(2), 149–172.

Van Lehn, 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.