Examining the Adequacy of Students' Priors and Teacher's Role in Attention to Critical Features in Designing for Productive Failure

Pee Li Leslie Toh & Manu Kapur
National Institute of Education, Nanyang Technological University, 1 Nanyang Walk, Singapore 637616
leslie.toh@nie.edu.sg, manu.kapur@nie.edu.sg

Abstract: This paper describes two classroom-based studies that examine the productive failure learning design from the following perspectives: (i) the adequacy of students' prior knowledge, and (ii) the necessity of teacher-led discernability of critical conceptual features. The entire investigation, set in the domain of monohybrid inheritance, involved 138 ninth and tenth grade science students. In the first study, students were either provided or not provided with monohybrid inheritance pre-requisites for their complex problem-solving activity. The second study compared a teacher-led with a student-led comparison and contrast during the consolidation phase. Findings of the first study suggested that students may not have adequate priors to generate representations and solution methods that explore multi-level conceptions. Findings of the second study suggested that teacher involvement in the compare-and-contrast may be beneficial for better knowledge assembly.

Introduction

There is now a growing body of research that suggests an efficacy of engaging students in exploratory or invention activities, prior to receiving an appropriate form of instructional intervention (e.g. Schwartz & Bransford, 1998; Schwartz & Martin, 2004). Kapur (2008) proposed the notion of Productive Failure (PF) to examine this efficacy, and more recently developed it into a learning design (Kapur & Bielaczyc, 2012).

The PF learning design 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 and knowledge assembly 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 sub-optimal RSMs and the assembly of canonical RSMs (mechanisms b – d).

Over a series of studies (e.g., Kapur, 2010, 2011, 2012), Kapur and colleagues have demonstrated the efficacy of the PF learning design over a direct instruction design in terms of significantly better conceptual and transfer gains without compromising procedural fluency. More importantly, findings suggest that the more RSMs students are able to generate, the more they learn from PF (Kapur, 2012; Kapur & Bielaczyc, 2012). Kapur and Bielaczyc (2012) argue that a greater RSM diversity implies greater prior knowledge activation and differentiation, which in turn helps students to better attend to and understand the critical features of the targeted concepts from subsequent instruction (Schwartz & Bransford, 1998; Schwartz & Martin, 2004).

The purpose of this paper is to more closely examine an assumption and a key mechanism of the PF learning design through two classroom-based studies in Singapore biology classrooms. In the PF learning design, there is an assumption that even though students may not have learnt the target concepts, they may still be able to leverage their informal and intuitive priors along with their formal priors to generate RSMs to complex, novel problems. Our first aim in this paper is to show that this assumption does not hold true more generally, especially when the targeted concepts are not entirely experientially grounded. Our second concern deals with one of the key mechanisms in learning design, which is, getting students to attend to the critical features of the targeted concept by comparing and contrasting student-generated RSMs and the canonical RSMs. To achieve this, PF emphasizes the role of the teacher in leading such a comparison and contrast (C&C) activity. However, the need for the teacher (as opposed to the student) doing the C&C has not been tested. Testing this mechanism forms our second aim.

Assumption of Students' Priors for Generating RSMs

Given the importance of student-generated RSMs and their correlation with learning from PF, it is important to examine more closely the assumption that students have the necessary priors to generate RSMs especially without having had any formal instruction on the targeted concepts. The complex problems of past PF studies (e.g., Kapur, 2008; Kapur & Bielaczyc, 2012) had contexts that were grounded in students' own experiences, for instance, a speeding offence (involved concepts of speed, friction, forces, acceleration, etc.), finding average speeds in a biking scenario (involved concepts of speeds, distance, time, etc.), and finding consistency of a soccer player given his goal-scoring record (involved concepts of central tendencies, graphing, deviation, etc.). Invariant across all these problem-solving contexts is that although students had not formally learned the targeted concepts, it is reasonable to suggest that they already had informal and experientially-grounded conceptions of these concepts (Palmer, 1999). If so, this only begs the question: What about problem-solving contexts based on phenomena that are not entirely experientially-grounded?

An example of such a phenomenon is monohybrid inheritance, which is a multi-level complex phenomenon. This is a specific area in biology, which comprises concepts from inheritance (macro-level) and genetics (micro-level). Essentially, inheritance is concerned with the gain of characteristics from one's parents, while genetics entails inheritance patterns, transmission of genes through generations and genetic expression (Kibuka-Sebitosi, 2007). Whereas the macro-level phenomenon of inheritance can be seen and observed, the underlying micro-level causal mechanisms cannot. Consequently, if students are placed in problem-solving contexts that that target monohybrid inheritance, they may well be able to generate RSMs at the macro-level, but may find it difficult to generate RSMs at both the macro and micro levels. We designed Study 1 to test this conjecture.

Teacher's Role in Drawing Attention to Critical Conceptual Features

A critical mechanism in the PF learning design is to draw students' attention to critical features of the targeted concepts. This is done in the consolidation phase by the teacher leading a C&C between student-generated RSMs and the canonical RSMs (Kapur & Bielaczyc, 2012). Thus, the PF learning design conjectures that the efficacy of RSM diversity is contingent upon a teacher-led C&C activity. This is in contrast with past research where students themselves were able to compare and contrast and attend to the critical features so long as the generation and exploration phase prepares the students to "see the significance of expert solutions and potential resources for learning" from the later formal consolidations (Schwartz & Martin, 2004, p. 168). Even earlier studies of PF suggested the same (Kapur, 2008; Kapur & Kinzer, 2009), as do others (e.g., Sánchez, García-Rodicio & Acuña, 2009). If so, why do more recent studies of PF (e.g., Kapur, 2010, 2011, 2012) emphasize the teachers' involvement in the consolidation phase? Is there a marginal gain of teachers leading the C&C as opposed to students doing it by themselves?

Expert-novice research does suggest that experts tend to attend to more critical features of a problem situation than novices, and are better at discerning features and significant informational patterns (Bransford, Brown, Cocking, Donovan & Pellegrino, 2000; Chi, Feltovich, & Glaser, 1981). If so, we can conjecture that having teachers (as relative experts in subject matter) may be better in drawing attention to the critical features during the C&C than if students were do it themselves. This in turn, may also result in better organizing and assembling these features into the targeted conceptions. However, this remains a conjecture, and Study 2 was designed to test this conjecture.

Sections 2 and 3 present the two studies. We conclude by discussing the findings from the two studies, and deriving implications for theory and future research.

Study 1: Testing the Assumption of Students' Priors for Generating RSMs

Participants and Design

Participants (*N* = 80; 53 female, 27 male) were ninth grade students from three intact science classes at a mainstream, co-educational secondary school in Singapore. They had not been instructed on monohybrid inheritance and its pre-requisites – specifically, nuclear division (included chromosomes and meiosis) and molecular genetics (included DNA and genes).

The 80 students were randomly divided into two conditions: a) High Micro-level Conception (HMiC) condition (*N* = 40, 28 female, 12 male) was provided with three one-hour lessons on nuclear division and molecular genetics by the first author with six years of experience teaching eleventh and twelfth grade biology, and b) Low Micro-level Conception (LMiC) condition (*N* = 40, 25 female, 15 male) was provided with three one-hour lessons about the eye, a topic unrelated to monohybrid inheritance or its pre-requisites. While both conditions were assumed to have similar levels of the macro-level conceptions, the experimental manipulation was designed to make them different in terms of the micro-level conceptions. Students in each condition were then randomly divided into dyads/triads, and were instructed to generate RSMs for a complex problem (see following section) on their own, except for the occasional affective support provided by the first author to ensure that groups remained on task. The students were given blank sheets of A4 size paper for their group work.

Design of the Complex Problem

The complex problem targeted concepts of monohybrid inheritance. It was adapted from Gregor Mendel's classical pea plants experimental data as well as the outcomes from the various genetic crosses (see below for an example).

Parental Generation	1st Generation Offspring	Further Genetic Cross	2nd Generation Offspring
Genetic cross between pea plants with purple flowers & pea plants with white flowers	Purple flowers: All; White flowers: 0	Genetic cross between offspring with purple flowers	Purple flowers: 700; White flowers: 230

The complex problem involved "effects-to-cause" (Stewart, 1988, p. 244) reasoning, and required students to generate multi-level (macro and micro) explanations for the data provided. More specifically, students needed to generate an explanation at the macro-level about the predominance of the purple flower characteristic, and concomitantly, generate micro-level causal explanations to account for how the predominance effect is possibly genetically controlled.

Data Source and Analysis

The diversity of student-generated RSMs formed the outcome measure. In previous PF studies, RSM diversity was defined as the number of different RSMs generated by a group (Kapur & Bielaczyc, 2012). In this study, RSM diversity was defined in terms of the macro and micro levels, that is, whether the RSMs generated by a group constituted explanations at the macro-level only or at both the macro and micro levels. Based on this operationalization of RSM diversity, a rubric (see Table 1) was developed to categorize the generated RSMs into either "macro-level conceptions" or "macro- and micro-level conceptions" based RSMs. RSMs in each category were further categorized according to the following critical conceptual features they targeted: "Dominant-Recessive", "Gametes Formation and Union" and "Genetic Crosses". Two experienced biology teachers independently scored the RSMs with an inter-rater reliability of 0.67.

Table 1. Rubric to analyze and categorize student-generated RSMs

Type of RSM	Conception (C)	Descriptions of Conception
Macro-level conceptions based RSM (See Figure 1a)	C1: Dominant-Recessive	Genetic materials interacted to generate the characteristics for each character
	C2: Gametes Formation and Union	Separation and union of genetic materials
Macro- and micro-level conceptions based RSM (See Figure 1b)		
	C3: Genetic Crosses	Rules for genetic crosses and the matching of characteristics of each character to their respective genetic makeup

Results

Table 2 presents the descriptive statistics. There was a significant association between the students' levels of taught pre-requisites and RSM Diversity (x^2 = 14.15, *df* = 1, *p* < 0.001). For example, the students from the HMiC condition represented genes in the form of DNA or chromosomal structures—an attempt at using the taught micro-level concepts for their RSM (See Figure 1b). However, students from the LMiC condition, *M* = 1.00, *SD* = 0.00, relied only on the macro-level conceptions. For example, they developed conceptions of "dominant" with macro-level conceptions, for example, "stronger" (See Figure 1a).

Table 2. Frequency counts of student groups from the LMiC and HMiC conditions

	Number of Student Groups	
	Macro-level RSMs	Macro- & Micro-level RSMs
LMiC (20 groups)	20	0
HMiC (19 groups)	9	10

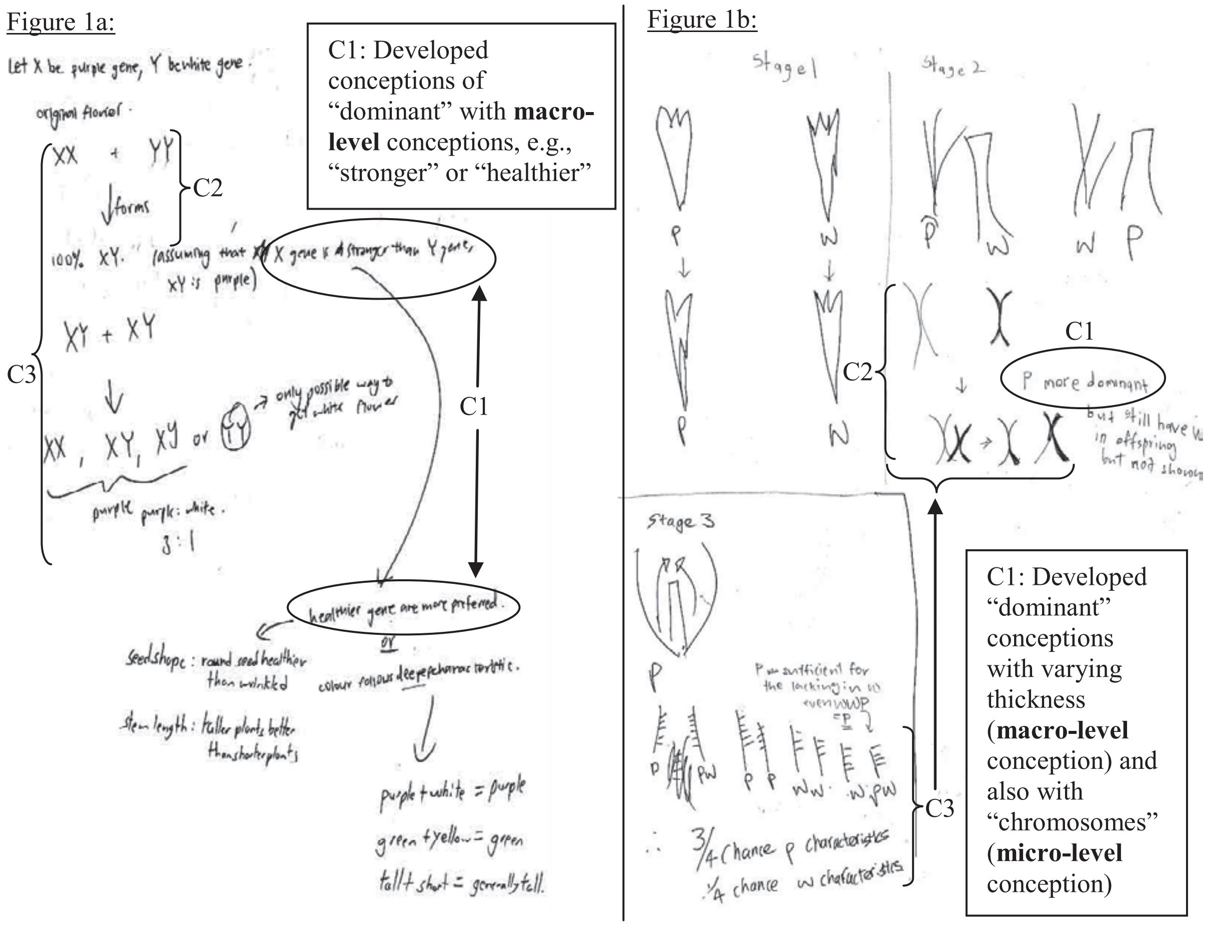

Figure 1a. LMiC: Student-generated RSM with macro-level conceptions; Figure 1b. HMiC: Similar student-generated RSM but with conceptions at both the macro and micro levels.

Study 2: Teacher's Role in Drawing Attention to Critical Features

Study 2 was designed to address the following question: How does having a teacher-led C&C versus a student-led C&C influence how much students learn from PF?

Participants

Participants (N = 58; 47 female, 11 male) were tenth grade students from two intact science classes at the same mainstream, co-educational secondary school in Singapore, and taught by the same biology teacher. Tenth grade students were chosen because at the time of the study, these students had not been taught monohybrid inheritance in the curriculum but had received instruction on its pre-requisites. Choosing students with pre-requisite knowledge was necessary as the findings of the first study indicated that students do not have adequate priors for the generation of RSMs that explored conceptions beyond the macro-level.

Design

Table 3 presents the design of the study. First, all students took a 50-minute single item pretest to ascertain their prior knowledge of monohybrid inheritance. This item pertained to a particular biological character (that is, humped versus smooth shoulders) of a mouse. It required students to first consider the first generation offspring outcomes that were obtained from a cross between two parental mice. Thereafter, they had to determine the genotypes (that is, the genetic makeup) and phenotypes (that is, the inherited characteristics) of the second generation offspring.

Each class had four 60-minute periods over three weeks. In the first period, all students in dyads/triads generated RSMs to the complex problem (the same as in Study 1) on their own, except for the occasional affective support by the teacher. Students were given blank A4-sized papers for their work. In the second period, the teacher carried out a lecture of the targeted concepts of monohybrid inheritance and explained the canonical solution.

The experimental manipulation happened in the third period after students had learnt the canonical concepts in the second period. In one class, the teacher led a whole-class discussion, where he brought students' attention to the critical conceptual features of monohybrid inheritance, that is, the conceptions of "dominant-recessive", "gametes formation and union" and "genetic crosses". He then explained and elaborated on these features. To highlight the affordances and constraints of the student-generated RSMs, he further compared and contrasted these RSMs with the critical features. Thereafter, he showed how the various critical conceptual features of monohybrid inheritance were organized and assembled into the canonical solution of the complex problem. We call this the Generation and Teacher-led (GTL) C&C condition. In the other class, students worked individually to compare and contrast on their own their generated RSMs with the canonical concepts taught in the previous period. They also had to suggest improvements for their generated RSMs. The students were given blank sheets of A4 size paper for their individual work. We call this the Generation and Student-led (GSL) C&C condition. The students also rated their level of engagement for this period with a nine items Likert scale survey (modified from Kapur, 2011) that had an alpha reliability of 0.94. The engagement rating was designed to capture differences, if any, between the GTL and GSL students on their self-reported engagement during the key manipulation, that is, the teacher- or student-led C&C activity.

The fourth period was designed for problem-solving practice, where students in both the conditions individually solved four questions on a worksheet, followed by a teacher-led discussion of the answers.

After the implementation, all students took a 50-minute, two-part, posttest, with the first part testing for procedural knowledge (one item), and the second part (four items) testing for conceptual knowledge. The procedural knowledge item was isomorphic with the pre-test item and the four worksheet questions. The conceptual knowledge items targeted the critical features such as: (i) What is the relationship between DNA, chromosomes, genes and alleles? (ii) What is monohybrid inheritance? (iii) What is the relationship of meiosis to monohybrid inheritance? (iv) What is the relationship between dominance and recessive? Ideally, one would want several items for each of the four conceptual questions for greater scale reliability. However, curriculum time constrained what we could test for, and we decided to look for greater coverage of the conceptual features rather than having more items for one particular feature. Consequently, however, the overall scale reliability of the four conceptual questions was low at 0.19. After the completion of each posttest item, the students rated the confidence of their answers using a 5-point Likert scale from 0 (no confidence) to 4 (100 % confidence).

Table 3. The overall design of Study 2

	Sequence of Events in Study 2					
		Phases of Productive Failure Learning Design				
Condition	Pretest	Generation & Exploration (1^{st} Period)	Direct Instruction (2^{nd} Period)	Comparison & Contrast (C&C) (3^{rd} Period)	Problem-Solving Practice (4^{th} Period)	Posttest
GSL	For all conditions	Student groups generated RSMs for complex problem	Teacher lectured on targeted concepts	Students individually compared and contrasted own group generated RSMs with canonical concepts	Students solved four problems; teacher discussed solutions with the whole class	For all conditions
GTL				Teacher led whole-class C&C of generated RSMs with canonical concepts		

Data Sources and Analyses

For process measures, we analyzed the student-generated RSMs with the same approach as in the first study. Two independent coders coded the student-generated RSMs with inter-rater reliability of 1.00. Other process measures included students' self-reported lesson engagement ratings during the key experimental manipulation period, where the C&C activity was either led by the teacher or the students themselves. The posttest scores on the procedural and conceptual knowledge and the item confidence ratings formed the outcomes measures. Two independent raters scored the posttest items with an inter-rater reliability of 0.92. All the instruments were validated by two experienced biology teachers from the participating school. The instruments were piloted on students who were not included in the main study. The teachers' feedback and the trial results were used to refine all instruments.

Results

Table 4 presents the descriptive as well as the inferential statistics, based on which the study's findings can be summarized as following:

1. Pretest. As expected, students in both the conditions demonstrated no knowledge of the targeted concepts, and there was statistically no significant difference between the two conditions.
2. RSM Diversity. As expected, there was variation within each condition in terms of student-generated RSMs. In each condition, some dyads or triads generated RSMs at the macro-level only, whereas others generated at both the macro and micro levels. However, a non-parametric chi-squared test (not shown in Table 4) showed there was statistically no significant difference between the two conditions on the generation of macro-level or a combined, macro- and micro-level RSMs, $x^2(1) = 2.23$, $p = 0.136$. In other words, there was statistically no difference between the two conditions in terms of what students were able to generate.
3. Self-Reported Engagement during the C&C Activity. There was generally high level of self-reported engagement in both the conditions, but there was statistically no significant difference between the two.
4. Posttest.
    a. *Procedural Knowledge*. Controlling for the effect of prior knowledge (as determined by the pretest score), an ANCOVA showed that there was statistically no significant difference between the conditions on procedural knowledge. Students from both the conditions scored close to the maximum score. This was expected given the isomorphic problem-solving practice in the final lesson for both the conditions.
    b. *Conceptual Knowledge*. Controlling for the effect of prior knowledge, an ANCOVA showed that students from the GTL condition performed significantly better than GSL condition on conceptual knowledge. This effect had a high effect size of .14.
    c. *Confidence*. As expected, students reported greater confidence for the procedural knowledge item than conceptual knowledge items. However, controlling for the effect of prior knowledge, ANCOVAs showed that there was statistically no significant difference between the conditions on their self-reported confidence for the procedural knowledge item. For conceptual knowledge items, GSL students reported slightly greater confidence than GTL students, but this effect was only marginally significant with a low-medium effect size of .06.

Table 4. Summary of descriptive and inferential statistics for Study 2

		GTL: Teacher-Led C&C		GSL: Student-Led C&C				
	Max	*M*	*SD*	*M*	*SD*	*F*	*p*	η_p^2
Pretest	5	0.13	0.34	0.04	0.19	F(1,56) = 1.54	.220	.03
Self-Reported Engagement (during C&C activity)	5	3.78	0.67	3.72	0.65	F(1,56) = 0.12	.726	<.01
Posttest								
Procedural	2	2.00	0.00	1.93	0.27	F(1,56) = 2.20	.144	.04
Conceptual	13	8.39	2.01	6.81	2.02	F(1,56) = 8.98	.004*	.14
Confidence (Procedural)	4	3.26	0.82	3.52	0.58	F(1,56) = 1.91	.172	.03
Confidence (Conceptual)	4	1.34	0.71	1.69	0.74	F(1,56) = 3.30	.075	.06

In the light of previous PF findings that RSM diversity is a significant predictor of conceptual understanding and transfer (Kapur, 2012; Kapur & Bielaczyc, 2012), we further examined the relationship between RSM diversity and posttest performance on conceptual knowledge. Within each condition, we compared the conceptual knowledge scores of students who had generated RSMs at macro-level with those who had generated at both macro and micro levels. Table 5 presents the descriptive and inferential statistics, based on which we can surmise that the descriptive trend in both conditions was that dyads or triads who generated macro- and micro-level RSMs (that is, had greater RSM diversity) performed better on average on conceptual knowledge than those who only produced macro-level RSMs. For GSL students, this trend did not reach significance and had a low effect size of .03. For GTL students, however, this trend reached marginal significance (perhaps due to low *n*) with a moderate-high effect size of .10. The moderate-high effect size for the GTL condition but not for the GSL condition tentatively confirms previous findings that RSM diversity plays a significant role in how much students learn from PF. More interestingly, it also shows that the effect of RSM diversity is mediated by whether the teacher or a student carries out the compare and contrast between the student-generated RSMs and the canonical RSMs.

Table 5. RSM diversity and conceptual knowledge scores by condition

	Posttest Conceptual Knowledge Score of Dyads/Triads who Produced:						
	Macro-Level RSMs only		Macro- & Micro-Level RSMs				
Experimental Condition	*M*	*SD*	*M*	*SD*	*F*	*p*	η_p^2
GSL	6.40	2.32	7.00	1.90	$F(1,24) = 0.61$	.445	.03
GTL	7.95	2.04	9.30	1.70	$F(1,29) = 3.09$	.090	.10

Discussion

We had two aims. First, we wanted to show that the assumption of students' priors for generating RSMs to complex problems does not hold true more generally, especially when the problem targets concepts that are not entirely experientially grounded. Our findings from Study 1 provided evidence for this. Second, we wanted to test the marginal gain of having teacher-led C&C activity over a student-led one. Findings from Study 2 suggested that a teacher-led C&C activity is significantly better than a student-led one in terms of conceptual gains. This was consistent with our hypothesis grounded in expert-novice literature that teachers as subject matter experts are perhaps better in knowing what and how to draw attention to the critical features, whereas students may not have been able to do this as well by themselves.

In Study 1, students were either provided (HMiC condition) or not provided (LMiC condition) with pre-requisites on the micro-level conceptions of monohybrid inheritance. Findings suggested that the HMiC condition students generated RSMs at both the micro and macro levels. However, the fact that these students had not received any formal instruction on monohybrid inheritance made it difficult for them to generate canonical conceptual connections on their own. Therefore, the overall interactions between the micro and macro levels gave rise to failed or suboptimal RSMs. However, having "failed" in generating canonical explanations of the phenomenon, such attempts at generating diverse RSMs are precisely the locus of powerful learning from PF (Kapur & Bielaczyc, 2012). This was not the case for the LMiC condition students who relied only on macro-level conceptions. These results suggest that one cannot always assume that students have the necessary prior resources to generate RSMs, especially for phenomenon with multi-level conceptions that are not entirely experientially-grounded. Some form of formal prior resources maybe necessary prior to getting students to generate RSMs.

In Study 2, after receiving instruction on the targeted concepts, students experienced either a teacher-led comparison and contrast (C&C) activity or a self-led C&C. Findings revealed that both the GSL and GTL conditions were effective in equipping the students with procedural fluency. However, students from the GTL condition had significantly better conceptual gains than their peers in the GSL condition. The better conceptual gains in the GTL condition may be attributed directly to the experimental manipulation of having a teacher rather than a student carrying out the C&C. For the GTL condition, the teacher, being a subject matter expert and better at discerning critical conceptual features (Bransford *et al*., 2000), might not only discuss the affordances and constraints of the student-generated RSMs, but also reinforce the critical conceptual features (Kapur & Bielaczyc, 2012). With his rich content knowledge and in-depth understanding of monohybrid inheritance (Bransford *et al*., 2000), he might also explain and elaborate on these various conceptual features, and further help in organizing and assembling them into the targeted conceptions (Kapur & Bielaczyc, 2012). GSL condition students, on the other hand, were probably aware of the critical conceptual features, especially after the complex problem solving activity. However, these students, being novices, might not have been able to effectively explain and elaborate on these features, and fail to organize and assemble them further into the targeted conceptions.

It is rather premature to generalize the above findings beyond the conditions and settings of these two studies. The low reliability of the posttest further constrains the reliability of our inference. Hence, larger samples and more reliable instruments are needed. Nonetheless, our findings can be seen as initial warrants that can drive further work to examine the boundary conditions of PF, and unpack its design elements more systematically. For example, our work going forward will be to examine how the provision of micro-level conceptions influences how much students learn from a subsequent consolidation phase. Additionally, we will also seek to understand if the collaborative participation structure of the PF learning design is a critical and necessary design decision to support student generation and exploration of RSMs (Kapur & Bielaczyc, 2012). This is essential because other studies (e.g. Sánchez, García-Rodicio & Acuña, 2009; Schwartz & Bransford, 1998) demonstrated learning gains even when they allowed students to explore or organize their RSMs individually. We believe that such efforts that unpack the design elements further stand to contribute substantively to on-going debates on how best to support students in their initial learning (Collins, 2012; Kapur & Rummel, 2009, 2012)

References

Bransford, J.D., Brown, A.L., Cocking, R.R., Donovan, M.S. & Pellegrino, J.W. (2000). *How People Learn. Brain, Mind, Experience, and School* (p. 31). Expanded Edition. Washington: National Academy Press.

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

Collins, A. (2012). What is the most effective way to teach problem solving? A commentary on Productive Failure as a method of teaching. *Instructional Science*.

Howitt, D. & Cramer, D. (2008). *Introduction to SPSS in Psychology*. *For Version 16 and Earlier*. Fourth Edition. England: Pearson Education Limited.

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

Kapur, M. & Kinzer, C.K. (2009). Productive failure in CSCL groups. *International Journal of Computer-Supported Collaborative Learning*, *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, Volume 2* (pp. 37 – 42). International Society of the Learning Sciences.

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

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*, DOI: 10.1007/s11251-012-9209-6

Kibuka-Sebitosi, E. (2007). Understanding genetics and inheritance in rural schools. *Journal of Biological Education*, *41* (2), 56 – 61.

Palmer, D.H. (1999). Exploring the link between students' scientific and nonscientific conceptions. *Science Education*, *83* (6), 639 – 653.

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

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.

Stewart, J. (1988). Potential learning outcomes from solving genetics problems: A typology of problems. *Science Education*, *72* (2), 237 – 254.

Acknowledgements

The study was funded by a Ministry of Education grant to the second author through the Office of Education Research of the National Institute of Education of Singapore. The authors would like to thank the principal, teachers and students of the participating school for their support for this research study.