Designing for Productive Failure in Mathematical Problem Solving

Manu Kapur & June Lee
National Institute of Education, Singapore

Abstract

This paper describes two quasi-experimental studies of *productive failure* (Kapur, 2008) in Singapore public schools for the curricular unit on average speed. In the first study, seventh-grade mathematics students from intact classes experienced one of two conditions: a) productive failure, where students solved complex, ill-structured problems on average speed *without* any instructional support or scaffolds up until a teacher-led consolidation, or b) traditional lecture and practice. Despite seemingly failing in their collective and individual problem-solving efforts, students from the productive failure condition significantly outperformed their counterparts from the other two conditions on both the well-structured and higher-order application problems on the post-test. The second study, conducted in another school with significantly lower academic ability students, replicated the findings of the first study. Findings and implications of productive failure for theory, design of learning, and future research are discussed.

Introduction

When and how to design structure in learning and problem-solving activities is a fundamental theoretical and design issue in education and the learning sciences. Structure can be operationalized in a variety of forms such as structuring the problem itself, scaffolding, instructional facilitation, provision of tools, content support, expert help, and so on (Puntambekar & Hübscher, 2005). Thus conceived, structure is designed to constrain or reduce the degrees of freedom in learning and problem solving activities; the lower the degree of freedom, the greater the structure (Woods, Bruner, & Ross, 1976). By doing so, structure increases the likelihood of novices achieving performance success during problem solving, which they might not otherwise be able to in the absence of support structures. Indeed, a vast body of research supports the efficacy of such an approach. This has led some to argue that instruction should be heavily guided especially at the start, for without it, little if any learning will take place (e.g., Kirschner, Sweller, & Clark, 2006). Further support for starting with greater structure in learning and problem solving activities with a gradual reduction (or fading) over time as learners gain expertise comes from several quarters (e.g., Puntambekar & Hübscher, 2005; Vygotsky, 1978; Woods et al., 1976).

More often than not therefore, researchers have tended to focus on ways of structuring learning and problem-solving activities so as to achieve performance success, whereas the role of failure in learning and problem solving remains largely underdetermined and under-researched by comparison (Clifford, 1984; Schmidt & Bjork, 1992). What is perhaps more problematic is that an emphasis on achieving performance success has in turn led to a commonly-held belief that there is little efficacy in novices solving problems without the provision of support structures. While this belief may well be grounded in empirical evidence, it is also possible that by engaging novices to persist and even fail at tasks that are beyond their skills and abilities can be a productive exercise in failure. Research reported in this paper explores this very possibility of designing for *productive failure* (Kapur, 2008).

Failure and Structure

The role of failure in learning and problem solving is no doubt intuitively compelling. For example, research on *impasse-driven learning* (VanLehn 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 structure, e.g., in the form of feedback, questions, or explanations, when the learner demonstrably makes an error or is "stuck," VanLehn et al's (2003) findings suggest that it may well be more productive to delay that structure up until the student reaches an impasse—a form of failure—and is subsequently unable to generate an adequate way forward. Echoing this delaying of structure in the context of text comprehension (also with college students), McNamara (2001) found that whereas low-knowledge undergraduate learners tended to benefit from high-coherence texts, high-knowledge undergraduate learners benefited from low-coherence texts, and especially more so when a low-coherence text preceded a high-coherence one. This, McNamara argues, suggests that reading low-coherence texts may force learners to engage in compensatory processing by using their prior knowledge to fill in the conceptual gaps in the target text, in turn, preparing them better to leverage a high-coherence text subsequently. Further evidence for such *preparation for future learning* (Schwartz & Bransford, 1998) can be found in the inventing to prepare for learning 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.

Clearly, the relationship between failure and structure forms a common thread through the abovementioned studies. It is reasonable to reinterpret their central findings collectively as an argument for a *delay of structure* in learning and problem-solving situations, be it in the form of feedback and explanations, coherence in texts, or direct instruction. Indeed, 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. However, the abovementioned studies deal with students solving well-structured problems as is typically the case in schools (Spiro et al., 1992).

While there exists a substantive amount of research examining students solving ill-structured problems *with* the provision of various support structures and scaffolds, Kapur's (2008) work on *productive failure* examined students solving complex, ill-structured problems *without* the provision on any external support structures or scaffolds. Kapur (2008) asked 11th-grade student triads from seven high schools India 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 group discussions were found to more complex and divergent than those of their well-structured counterparts, leading to poor group performance. However, 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* (Marton, 2007) how to structure an ill-structured problem, thereby facilitating a spontaneous transfer of problem-solving skills (Kapur & Kinzer, 2009).

The above findings, while preliminary, underscore the implication that by delaying structure in the learning and problem-solving activities so as to allow learners to persist in and possibly even fail while solving complex, ill-structured problems can be a productive exercise in failure.

The purpose of this paper is to report findings from an on-going, classroom-based research program with grade seven students on productive failure in mathematical problem solving at two mainstream public schools in Singapore. The two schools, hereinafter referred to as School A and School B, were selected based on the academic ability profile of their student intake as evidenced by the Primary School Leaving Examination (PSLE[1]: the 6th-grade national standardized tests used to gain entry into secondary schools, i.e., grades 7 through 10). A MANCOVA, $F(2, 191) = 660.52$, $p < .001$, revealed that students from School A achieved a significantly higher PSLE score, $M = 235.6$, $SD = 2.45$, than those from School B, $M = 209.3$, $SD = 5.95$. Likewise, students from School A achieved significantly better PSLE Math grade[2], $M = 1.60$, $SD = .55$, than those from School B, $M = 2.50$, $SD = .62$. Details of the two studies follow next; School A followed by School B.

Study 1: School A

Participants

Participants were 75, Secondary 1 (7th-grade) students (43 male, 33 female) at a co-educational, secondary school in Singapore. Students were from two math classes (37 and 38 students respectively) taught by the *same* teacher. Students had limited or no experience with the targeted curricular unit—average speed—prior to the study.

Research Design

A pre-post, quasi-experimental design was used with one class ($n = 37$) assigned to the 'Productive Failure' (PF) condition and other class ($n = 38$) assigned to the 'Lecture and Practice' (LP) condition. Both classes participated in the seven lessons totaling six hours of instructional time over two weeks. Before the unit, all students wrote a 30-minute, 9-item pre-test ($\alpha = .72$) as a measure of prior knowledge of the targeted concepts. There was no significant difference between the two conditions on the pre-test, $F(1,73) = .177$, $p = .675$. After the unit, all students took a post-test.

Productive Failure (PF) The 37 students in the PF class were assigned to groups by the teacher, resulting in 13 groups (11 triads, 2 dyads). In the PF instructional design cycle, student groups took two periods to work face-to-face on the first ill-structured problem. Following this, students took one period to solve two extension problems individually. The extension problems required students to consider the impact of changing one or more parameters in the group ill-structured problem. No extra support or scaffolds were provided during the group or individual problem-solving nor was any homework assigned at any stage. The PF cycle—group followed by individual problem solving—was then repeated for the next three periods using another ill-structured problem scenario and its corresponding what-if extension problems. Only during the seventh (and last) period was a consolidation lecture held where the teacher led a discussion of the targeted concepts.

Two ill-structured problem scenarios were developed for the unit on average speed. The design of the ill-structured problem scenarios was closely aligned to the design typology for problems espoused by several scholars (e.g., Voss, 1988). An additional design element was that of *persistence*, i.e., the focus was more on students being able to persist in problem solving than on actually being able to solve the problem successfully. A focus on ensuring that students solve a problem which they may not otherwise be able to in the absence of support structures necessitates the provision of relevant support structures and scaffolds during problem solving. However, a focus on persistence does not necessitate such a provision as long as the design of the problem allows students to make some inroads into exploring the problem and solution space without necessarily solving the problem successfully. Validation of the problem scenarios was carried out through multiple iterations of design and pilot-testing with a small group of students from the previous cohort of students from the school. The validation exercise informed the time allocation

[1] More information on the PSLE can be found at www.moe.edu.sg.

[2] The lower the mean, the better the grade; grade A* is equivalent to 1 point, A to 2 points, B to 3, and so on.

for group and individual tasks as well as the design element of persistence.

Lecture & Practice (LP) The 38 students in the LP class were involved in teacher-led lectures guided by the course workbook. The teacher introduced a concept (e.g., average speed) to the class, worked out some examples, encouraged students to ask questions, following which students solved problems for practice. The teacher then discussed the solutions with the class. For homework, students were asked to continue with the workbook problems. This cycle of lecture, practice/homework, and feedback then repeated itself over the course of seven periods. Note that the worked-out examples and practice problems were typically well-structured problems with fully-specified parameters, prescriptive representations, predictive sets of solution strategies and solution paths, often leading to a single correct answer. Students worked independently most of the time although some problems were solved collaboratively.

In short, the LP condition represented a design that was highly structured from the very beginning with the teacher leading the students through a set of well-structured problems with proximal feedback and regular practice. The PF condition represented a design that delayed structure (in the form of the consolidation lecture) up until students had completed two ill-structured problem scenarios and the corresponding what-if extension problems *without* any instructional facilitation, support structures, or scaffolds.

It is important to note that the research design allows for a comparison between instructional designs as wholes, not their constituent elements. Unlike laboratory experiments, the reality of classroom-based research is that one is rarely able to isolate individual elements of an instructional design in a single study because it is the complexity of how the individual elements combine that gives rise to the efficacy of a particular design (Brown, 1992). Thus, we put greater emphasis on an ecological comparison of designs vis-à-vis causal attribution of effects to design elements.

Hypothesis Based on past research on productive failure, we hypothesized that compared to the LP condition, designing for persistence and delaying structure in the PF condition may result in students attempting to assemble key ideas and concepts underlying average speed, as well as exploring various representations and methods for solving the ill-structured problems (Schwartz & Martin, 2004; Spiro et al., 1992). We did not expect students who were novices to the targeted concept of average speed to use the most effective representations and domain-specific methods for solving the problems, nor did we expect them to be successful in their problem-solving efforts (Chi et al., 1981; Kirschner et al., 2006). However, such a process may be integral to engendering the necessary knowledge differentiation which may help students better *discern* and understand those very concepts, representations, and methods when presented in a well-assembled, structured form during the consolidation lecture (Marton, 2007; Schwartz & Bransford, 1998).

Process Results

Problem representation Group-work artifacts provided a rich source of data about the nature of problem representations produced by the groups in the process of solving the problem. A qualitative analysis revealed that groups produced a *diversity* of linked representations. We illustrate this using a paradigmatic example (see Figure 1).

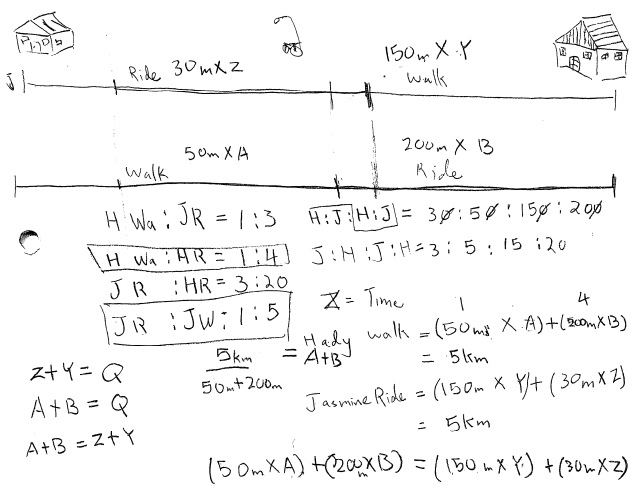

Figure 1: Example of Linked, Representational Diversity.

First, we describe the essence of the ill-structured problem scenario that Figure 1 refers to (the actual problem scenario was two pages long). The problem presented a scenario where two friends, Jasmine (J) and Hady (H), were on their way to an audition on their bikes, when Jasmine's bike broke down. Given their different walking and biking speeds, groups had to determine the change over point, i.e., the optimal distance that Jasmine and Hady should bike and walk, in order to reach the audition at the same time.

Figure 1 reveals that the group used a diverse but linked set of *iconic* (e.g., house, bicycle), *graphical* (e.g., straight lines for J and H), *proportional* (e.g., ratios between J's and H's speeds and distances for walking and riding), and *letter-symbolic algebraic* representations (e.g., using numbers and unknowns such as X, Y, A, B to link with other representations). Additionally, the group set up systems of algebraic equations. The use of letter-symbolic algebraic representations is significant because the introduction of algebra in the formal curriculum does not happen until after the unit on rate and speed. As hypothesized, however, despite producing various inter-connected representations, the group was not able to solve the problem successfully. For example, neither the proportions nor the algebraic equations were manipulated further to solve the problem[3]. To confirm if this inability to develop a solution held true more generally, we analyzed group solutions as well as individual solutions to the extension problems.

[3] See Kapur (2009) for a fuller qualitative analysis of the nature of group representations, methods and strategies used by the groups as evidenced in their discussions and solutions.

Group & Individual Solutions All groups were able to identify relevant parameters such as the various distances, speeds, and time, and perform basic calculations involving these parameters (e.g., calculating time, given speed and distance). However, they were unable to build on their representations to devise either domain-general and/or domain-specific strategies, develop at least one solution, and support it with quantitative and qualitative arguments (Chi et al., 1981; Spiro et al., 1992). For the first and second ill-structured problems, only 11% and 21% of the groups respectively managed to solve the problem; an average success rate was evidently low at only 16%. This was not surprising because the problem scenarios were carefully designed and validated for students to persist in problem solving without necessarily doing it successfully.

Likewise for the first and second individual extension problems, 3% and 20% of the students respectively managed to solve the problem; an average success rate of only 11.5%.

Confidence ratings PF students rated their confidence their extension problem solutions on a 5-point Likert scale from 0 (0% confidence) to 4 (100% confidence). The average confidence was low, $M = 1.22$, $SD = .82$.

Student Performance in the LP condition Students in the LP condition, by design, repeatedly experienced performance success in solving well-structured problems under the close monitoring, scaffolding, and feedback provided by the teacher. Data from homework assignments provided a proxy measure for student performance in the LP condition. Based on the teacher's report, the average percentage score on the homework assignments ranged between 88% and 100%.

Summary Our analysis revealed that in spite of attempting various representations and methods for solving the problem, PF students seemed unsuccessful in their problem-solving efforts, be it in groups or individually. Their self-reported confidence in their own solutions was also reportedly low. These process findings also double up as a manipulation check demonstrating that students in the PF condition experienced failure at least in the conventional sense. Thus, on conventional measures of efficiency, accuracy, and performance success, students in the PF condition seemed to have failed relative to their counterparts in the LP condition.

Outcome Results

Post-test All students took a 40-minute, 6-item post-test (α = .86) comprising five well-structured problem items similar to those on the pre-test, and one complex problem item (for the items, see Kapur, 2009). Score on the well-structured items and the complex item formed the two dependent variables in our analysis. Controlling for the effect of prior knowledge as measured by the pretest, a MANCOVA revealed a statistically significant effect of condition (PF vs. LP) on posttest scores, $F(2, 71) = 5.64$, $p = .005$, partial eta-squared, $\eta^2 = .14$[4].

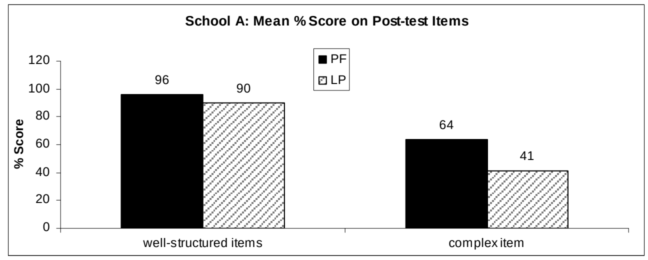

Figure 1. School A's breakdown of post-test performance as a *percentage* of the maximum score for the four types of items.

Univariate analysis further suggested that (see Figure 1):

i. On the five well-structured items (maximum score = 32), students from the PF class scored higher, $M = 30.61$, $SD = 4.09$, than those from the LP class, $M = 28.89$, $SD = 4.26$. This effect was statistically significant, $F(1,72) = 4.78$, $p = .032$, partial $\eta^2 = .06$. Notwithstanding the moderate effect size, it was remarkable that PF students who were not given any homework or practice assignments still managed to outperform LP students who did receive such practice and feedback on well-structured type of items.

ii. On the complex item (maximum score = 7), students from the PF class scored higher, $M = 4.35$, $SD = 2.13$, than those from the LP class, $M = 2.77$, $SD = 2.12$. This effect was statistically significant, $F(1,72) = 10.11$, $p = .002$, partial $\eta^2 = .12$.

Thus, students from the PF class outperformed those from the LP class on both the well-structured and complex items on the posttest thereby suggesting that the productive failure hypothesis held up to empirical evidence.

Study 2: School B

Participants

Participants were 114, Secondary 1 (grade 7) students (63 male, 51 female; 12-13 years old) from a secondary school in Singapore. Students were from three intact math classes (36, 38, and 40 students). Students had limited or no experience with the targeted curricular unit—average speed—prior to the study. Recall that these students were of significantly lower academic ability, in both general and math ability compared to students from School A.

Research Design

The research design and procedures were same as in School A with the following variations:

i. Two of the three classes were taught by the same teacher, teacher A. Of the two classes taught by teacher A, one class ($n = 38$) was assigned to the PF condition (referred to as PF-A) and the other ($n = 40$) to the LP condition. The third class ($n = 36$), taught by the second teacher, teacher B, was assigned to the PF condition (PF-B). There was no significant difference between the three classes on the pretest, $F(2,114) = .536$, $p = .586$.

[4] As a rule of thumb, partial $\eta^2 = .01$ is considered a small, .06 medium, and .14 a large effect size (Cohen, 1977).

The same teacher for the LP and PF-A condition allowed us to test for replication of findings from School A. At the same time, the PF-B class under a different teacher allowed us to investigate variation between teachers A and B in their enactment of the same PF design.

ii. A pilot study with a small group of students prior to the actual study revealed that School B students' thresholds for persistence were lower than that in School A. Therefore, the individual extension problems (which immediately follow group problem solving) were removed from the design. The time saved was spent on an extended consolidation lecture, as proposed by the teachers, given the significantly lower math ability of their students compared to those from School A.

iii. Two well-structured problem items from the posttest were dropped because of low reliability. Based on the findings from School A wherein PF students generated and explored a variety of representations, we conjectured that they might also demonstrate better *representational flexibility* in problem solving (Goldin, 1998). Thus, two additional items were added to the posttest to measure *representational flexibility*, that is, the extent to which students are able to flexibly adapt their understanding of the concepts of average speed to solve problems that involve tabular and graphical representations.

Process Results

Representations and Group Solutions Our analysis (similar to that for School A but not reported here due to space constraints) suggested that despite producing various inter-connected representations and methods for solving the problems, groups were ultimately unable to solve the problems successfully. None of the PF groups from the PF-A class and only 4 of the 13 groups in PF-B class were successful at solving one of the problems. Therefore, the average success rate was 7%.

Individual Extension Problems & **Confidence Ratings** As described above, individual extension problems and the associated confidence ratings were not part of the PF design for School B.

Student Performance in the LP condition As in School A, based on the teacher's report, the average percentage score on the homework assignments ranged between 84% and 100%.

Summary In sum, the analysis revealed that in spite of attempting various representations and methods for solving the problem, PF students seemed unsuccessful in their problem-solving efforts. Once again, on conventional measures of efficiency, accuracy, and performance success, students in the PF conditions seemed to have failed relative to their counterparts in the LP condition.

Outcome Results

Post-test Score on the well-structured items and the complex item formed the two dependent variables in our analysis. Controlling for the effect of prior knowledge, a MANCOVA revealed a statistically significant multivariate effect of class (PF-A, PF-B, or LP) on posttest scores, $F(8, 212) = 6.48$, $p < .001$, partial $\eta^2 = .20$.

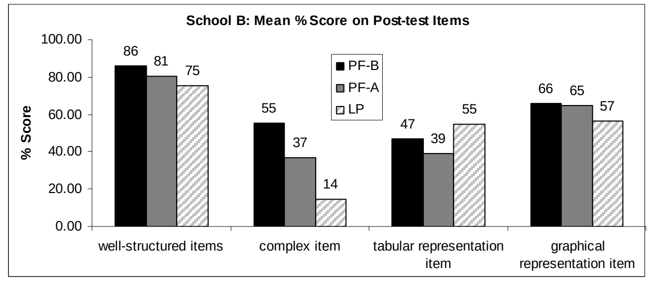

Figure 2. School B's breakdown of post-test performance as a *percentage* of the maximum score for the four types of items.

Univariate analysis further suggested that (see Figure 2):

i. On the three <u>well-structured items (maximum score = 19)</u>, students from the PF-B class scored the highest, $M = 16.30$, $SD = 2.29$, followed by those from the PF-A class, $M = 15.30$, $SD = 2.45$, and then by the LP class, $M = 14.30$, $SD = 3.11$. This effect was statistically significant, $F(2, 110) = 5.26$, $p = .007$, partial $\eta^2 = .09$.

ii. On the <u>complex item (maximum score= 7)</u>, students from the PF-B class scored the highest, $M = 3.86$, $SD = 2.44$, followed by those from the PF-A class, $M = 2.58$, $SD = 2.65$, and then by the LP class, $M = 1.01$, $SD = 2.32$. This effect was significant, $F(2, 110) = 17.98$, $p < .001$, partial $\eta^2 = .25$.

iii. On the <u>tabular representation item (maximum score = 6)</u>, students from the LP class scored the highest, $M = 3.30$, $SD = 2.23$, followed by those from the PF-B class, $M = 2.80$, $SD = 1.73$, and then by the PF-A class, $M = 2.34$, $SD = 1.87$. However, this effect was statistically not significant, $F(2, 110) = 1.09$, $p = .339$, partial $\eta^2 = .02$.

iv. On the <u>graphical representation item (maximum score = 3)</u>, students from the PF-B class scored the highest, $M = 1.97$, $SD = 1.05$, followed by those from the PF-A class, $M = 1.95$, $SD = 1.01$, and then by the LP class, $M = 1.70$, $SD = .97$. However, this effect was not significant, $F(2, 110) = 2.06$, $p = .133$, partial $\eta^2 = .04$.

Discussion

This study was designed to explore the hidden efficacies, if any, in delaying structure in the learning and performance space of students by having them engage in unscaffolded problem-solving of complex, ill-structured problems prior to direct instruction. In both studies, students from the PF conditions outperformed those from the LP conditions on both the well-structured and complex problem items on the posttest, thereby suggesting that the productive failure hypothesis held up to empirical evidence. Furthermore, in the second study (in School B), there were no significant differences between the conditions on the tabular representation item. This could be because of the relative concreteness of a tabular representation, which might have been easier for students to work with than a more abstract representation. On the graphical representation item, students from the PF conditions performed better than their counterparts from the LP condition. However, this effect did not reach significance. Future studies with larger samples should help unpack this effect. Finally, the second study also suggested some variance within teachers' enactments of the PF design. This was expected, and we are currently analyzing this variance using classroom observation notes and video data that we collected in School B.

As hypothesized, explanation for the above findings comes from the notions that perhaps what was happening in the productive failure condition was that students were seeking to assemble or structure key ideas and concepts while attempting to represent and solve the ill-structured problems. Indeed, qualitative analyses revealed that students tried different concepts, representations, and methods for solving the problems. They were evidently not successful, but the process of exploring the problem and solution spaces for representations and methods for solving the problem may have engendered sufficient knowledge differentiation that prepared them to better discern and understand those very concepts, representation, and methods when presented in a well-assembled, structured form during the consolidation lecture (Marton, 2007; Schwartz & Bransford, 1998). Furthermore, it is plausible that having explored various representations and methods for solving the complex ill-structured problems, they perhaps better understood the affordances of the representations and methods when delivered by the teacher during the consolidation lecture. In other words, when the teacher explained the "canonical" representations and methods for solving the problem, they perhaps better understood not only why the canonical representations and methods work but also the reasons why the non-canonical ones—the ones they tried—did not work.

It is of course much too early to attempt any generalization of the claims from two studies; the scope of inference technically holds only under the conditions and settings of the two studies and is thus circumscribed by the content domain, communication modality, age-group, and various socio-cultural factors. The work reported herein represents a preliminary but important step towards developing a theory of conditions under which failure can be productive. Going forward, therefore, future research would do well to extend this study to larger samples across schools and subjects. At the same time, further analyses of group discussions should unpack learning mechanisms underpinning the productive failure effect. Examining learner characteristics (e.g., motivation, goal orientation, ability, etc.) as well as the nature and content of interactional behaviors (e.g., explanations, critiquing, elaborating, up-taking, etc.) and relating them to eventual gains in group and individual performance would be most insightful in developing an explanatory base for productive failure.

Acknowledgements

The research reported in this paper was funded by grants to the first author from the Learning Sciences Lab of the National Institute of Education of Singapore.

References

Brown, A. L. (1992). Design experiments. *Journal of the Learning Sciences, 2*(2), 141-178.

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, *18*(1), 32-42.

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

Goldin, G. A. (1998). Representational systems, learning, and problem solving in mathematics. *Journal of Mathematical Behavior*, *17*(2), 137-165.

Kapur, M. (2009). Productive failure in mathematical problem solving. *Instructional Science*. doi: 10.1007/s11251-009-9093-x.

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

Kapur, M., & Kinzer, C. (2009). Productive failure in CSCL groups. *International Journal of Computer-Supported Collaborative Learning* (*ij*CSCL), *4*(1), 21-46.

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

Marton, F. (2007). Sameness and difference in transfer. *The Journal of the Learning Sciences*, *15*(4), 499-535.

McNamara, D. S. (2001). Reading both high-coherence and low-coherence texts: Effects of text sequence and prior knowledge. *Canadian Journal of Experimental Psychology*, *55*(1), 51-62.

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.

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.

Spiro, R. J., Feltovich, R. P., Jacobson, M. J., & Coulson, R. L. (1992). Cognitive flexibility, constructivism, and hypertext. In T. M. Duffy, & D. H. Jonassen (Eds.), *Constructivism and the technology of instruction: A conversation*. NJ : Erlbaum.

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

Voss, J. F. (1988). Problem solving and reasoning in ill-structured domains. In C. Antaki (Ed.), *Analyzing everyday explanation: A casebook of methods*. London: Sage Publications.

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