

Productive Failure in Mathematical Problem Solving

Manu Kapur (manu.kapur@nie.edu.sg), Leigh Dickson (leigh.dickson@nie.edu.sg)

National Institute of Education, 1 Nanyang Walk, Singapore

Toh Pui Yhing (toh_pui_yhing@moe.edu.sg)

Clementi Town Secondary School, 121 Queensway, Singapore

Abstract

This paper reports on a quasi-experimental study comparing a “productive failure” (Kapur, 2006, in press) instructional design with a traditional “lecture and practice” instructional design for a two-week curricular unit on rate and speed. Participants comprised 75, 7th-grade mathematics students from a mainstream secondary school in Singapore. Students experienced either a traditional lecture and practice teaching cycle or a productive failure cycle, where they solved complex, ill-structured problems in small groups without the provision of any support or scaffolds up until a consolidation lecture by their teacher during the last lesson for the unit. Findings suggest that students from the productive failure condition produced a diversity of linked problem representations but were unable to produce good quality solutions, be it in groups or individually. Expectedly, they reported low confidence in their solutions. Despite seemingly failing in their collective and individual problem-solving efforts, students from the productive failure condition significantly outperformed their counterparts from the lecture and practice condition on both well- and ill-structured problems on the post-tests. After the post-test, they also demonstrated significantly better performance in using structured-response scaffolds to solve problems on relative speed—a higher-level concept not even covered during instruction. Findings and implications are discussed.

Introduction

Situative, socio-constructivist theories of learning emphasize the importance of having learners engage in authentic, ill-structured problem-solving activities for meaningful learning to take place (Brown, Collins, & Duguid, 1989; Spiro et al., 1992). The complex and ill-structured nature of the problem thus demands, as is argued, that support structures be provided as learners engage in solving such problems, for without the support structures, learners may fail. Structure, broadly conceived, comes in a variety of forms such as structuring the problem itself, scaffolding, instructional facilitation, provision of tools, expert help, and so on, and a substantial amount of research examines the effect of structuring and scaffolding learners within ill-structured problem-solving activities (Reiser, 2004). Inadvertently, this has led to a commonly-held belief that there is little efficacy in students solving complex, ill-structured problems *without* the provision of external structures to support them (Kapur, 2006, in press). Of course, believing in the efficacy of structuring what might otherwise be a complex, divergent, and unproductive process is well-placed (Kirschner, Sweller, & Clark, 2006). However, allowing for the concomitant possibility that under certain conditions letting learners persist, struggle,

and even fail at tasks that are ill-structured and beyond their skills and abilities may in fact be a productive exercise in failure requires a paradigm shift. We explore this very possibility.

Failure and Structure

Several research programs increasingly point to the role of failure in learning and problem solving. For example, research on *impasse-driven learning* (VanLehn et al., 2003) in coached problem-solving situations suggests that 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, McNamara (2001) found that whereas low-knowledge learners tended to benefit from high-coherence texts, high-knowledge 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, Schwartz and Martin (2004) demonstrated an existence proof for the hidden efficacy of invention activities when such activities preceded direct instruction (e.g., lectures), 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 involve students solving well-defined, well-structured problems, which 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, Kapur's (2006, in press) work on *productive failure* examined students solving complex, ill-structured problems *without* the provision on any external support structures or scaffolds. Kapur (2006, in press) 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 be 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 (2006, in press) 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.

An important instructional implication one could derive from the abovementioned research programs (taken together) and test it in a classroom-based setting would be that instructional designs need not take learners through a narrow path to success or the "correct" answer in the most efficient manner especially when learners are engaged in solving ill-structured problems. Therein lies the purpose of this study.

Purpose

The purpose of this study was to design a productive failure instructional cycle for mathematics classrooms in a Singapore school and compare it with a conventional lecture and practice instructional cycle. We wanted test whether or not there is a hidden efficacy in delaying structure in the learning and performance space of students by having them engage in unscaffolded problem-solving of complex, ill-structured problems first before any direct instruction. To achieve this, two classroom-based, quasi-experimental studies with 7th-grade mathematics students were carried out; each study targeting a two-week curricular unit. This paper focuses on the second unit (targeting the curricular unit on rate and speed) because the second unit provided confirmatory evidence for productive failure, and thus bolstered the pedagogical tractability of productive failure in real classroom contexts.

Method

Participants

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

Research Design

A 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 same number of lessons for the targeted unit totaling seven, 55-minute periods over two weeks. Thus, the amount of instructional time was held constant for the two conditions. Before the unit, all students wrote a 30-minute, 9-item pre-test (*Cronbach 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 two post-tests (described below).

Productive Failure (PF) Class 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 were designed as what-if scenarios that 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 rate and speed. The design of the ill-structured problem scenarios was closely aligned to the design typology for problems espoused by several scholars (e.g., Goel & Pirolli, 1992; 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. To achieve this, validation of the problem scenarios was carried out through multiple iterations of design with two mathematics teachers at the school as well as pilot-testing with a small group of students (4 triads, 1 dyad) from the previous cohort of students from the school. The validation exercise informed the time allocation for group and individual tasks as well as the design element of persistence.

Lecture & Practice (LP) Class 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., 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.

Hypothesis We hypothesized that compared to the LP condition, the delay of structure in the PF condition may result in students attempting to assemble key ideas and concepts underlying rate and speed, as well as explore various ways of representing and solving the ill-structured problems (Spiro et al., 1992). Although we expected students not to be successful in their problem-solving efforts, as long as they persisted in the problem-solving process, we hypothesized that it may engender sufficient knowledge differentiation which may help them better discern and understand those very ideas, concepts, and representations when presented in a well-assembled, structured form during the consolidation lecture (Marton, 2007; Schwartz & Bransford, 1998).

It is important to note that the research design allows for a comparison between two 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 from the start, and therefore, we put greater emphasis on an ecological comparison of designs vis-à-vis causal attribution of effects to design elements (Brown, 1992; Schwartz & Martin, 2004).

Data Sources & Analysis

Due to space constraints, data analysis procedures are described together with the results in the following section. Both process and outcome measures were analyzed. For process measures, we analyzed the *problem representations* produced by groups, *group solutions*, *individual solutions* to the what-if extension problems, and the corresponding *confidence ratings*. Group discussions were also captured in audio but are currently under transcription. *Post-test scores* formed the outcome measures.

Process Analyses & Results

Problem representation Each group handed in all their work, including rough work and calculations. These 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) from one of the ill-structured problems that the groups solved.

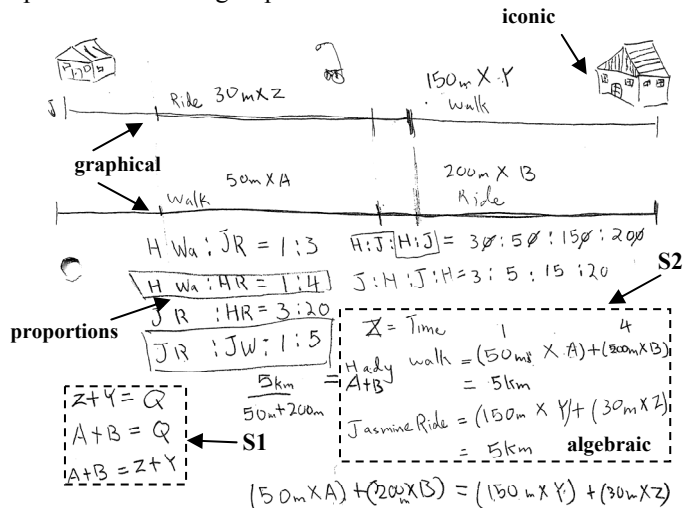


Figure 1: Example of Linked, Representational Diversity. S1 and S2 refer to systems of algebraic equations.

Before discussing the representation in Figure 1, we describe the essence of the ill-structured problem scenario that Figure 1 refers to (the actual problem scenario was two pages long). Essentially, the problem presented a scenario where two friends, Jasmine and Hady, had to get to an exhibition by a certain time. They could walk and/or ride a bicycle. The constraint was that they had to reach the exhibition at the same time despite having different walking and biking speeds. Furthermore, a little while into their journey, one of the bicycles breaks down, forcing them to re-strategize. Groups had to determine ways in which Jasmine and Hady could ride and/or walk for different periods of times and distances to reach the exhibition.

Figure 1 reveals that the group used a diverse but linked set of *iconic* (e.g., house, bicycle), *graphical* (e.g., straight lines for Jasmine and Hady), *proportional* (e.g., ratios between Jasmine's and Hady'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, S1 and S2. S1 and S2 are conceptually correct representations of the problem but because they contain more unknowns than equations, they are obviously not solvable. The use of letter-symbolic algebraic representations is significant because the introduction of algebra in the formal curriculum does not occur until after the unit on rate and speed. It seemed that the students were able to produce algebraic representations and domain-specific strategies without having had any prior, formal instruction in algebra.

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. 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.

Group solutions Content analysis of the groups' solutions suggested that 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, to the extent that groups were actually able solve the problems successfully, we noticed a clear, bimodal distribution, i.e., groups were either able to solve the problem successfully or not despite their extensive exploration of the problem and solution spaces. By solving a problem successfully we mean that groups were able to build on their representations to devise either domain-general and/or domain-specific strategies, develop a solution, and support it with quantitative and qualitative arguments. For the first ill-structured problem, only 11% of the groups managed to solve the problem. For the second ill-structured problem, 21% of the groups managed to solve the problem. Thus, the average success rate was evidently low at only 16%. This was not totally surprising because the problem scenarios were carefully designed and validated for students persisting in their attempt to solve the problem without necessarily solving it successfully.

Individual solutions to what-if extension problems Content analysis of the solutions produced by students for the what-if extension problems displayed a pattern similar to that of group performance. For the first and second extension problems, 3% and 20% of the students respectively managed to solve the problem. Thus, the average success rate for solving the extension problem scenarios was also low at only 11.5%.

Confidence ratings After solving the individual what-if extension problems, students in the PF design condition rated the confidence they had in their solutions using a 5-point Likert scale from 0 (0% confidence) to 4 (100% confidence). The average confidence reported by students was low, $M = 1.22$, $SD = .82$.

Summary In sum, the process findings suggested that students used a variety of ways to represent the problems but were ultimately unable to successfully solve the problems, be it in groups or individually. Their confidence in the solutions was also reportedly low. In a traditional classroom setting, these findings would be considered a failure on the part of the PF students. Thus, on conventional measures of efficiency, accuracy, and performance success, students in the PF condition seemed to have failed¹.

Outcome Analyses & Results

Individual outcomes were measured using two post-tests. The first post-test (post-test 1) targeted content covered during the unit on rate and speed. The second post-test (post-test 2) targeted an extension concept—relative speed—that was *not* covered during the unit.

Post-test 1 Students from both the PF and LP classes were given 40 minutes to complete a 7-item post-test (*Cronbach alpha* = .78) comprising six well-structured items similar (not same) to those on the pre-test as well as one ill-structured problem item (see below for an example of each).

A post-test well-structured item: *David travels at an average speed of 4km/hr for 1 hour. He then cycles 6km at an average speed of 12 km/hr. Calculate his average speed for the entire journey.*

A post-test ill-structured item: *Hummingbirds are small birds that are known for their ability to hover in mid-air by rapidly flapping their wings. Each year they migrate approximately 8583 km from Canada to Chile and then back again. The Giant Hummingbird is the largest member of the hummingbird family, weighing 18-20 gm. It measures 23cm long and it flaps its wings between 8-10 times per second. For every 18 hours of flying it requires 6 hours of rest. The Broad Tailed Hummingbird beats its wings 18 times per second. It is approximately 10-11 cm and weighs approximately 3.4 gm. For every 12 hours of flying it requires 12 hours of rest. If both birds can travel 1 km for every 550 wing flaps and they leave Canada at approximately the same time, which hummingbird will get to Chile first?*

Controlling for the effect of prior knowledge (as measured by the pre-test), an ANCOVA revealed a statistically significant effect of condition (PF vs. LP) on post-test 1 scores, $F(1, 72) = 10.69$, $p = .002$, ES (effect size) = .75. The adjusted (i.e., after controlling for prior knowledge) mean performance of students in the PF class, $M = 37.8$, $SD = 4.57$, was better than those in the LP class, $M = 33.4$, $SD = 6.52$; an average difference of 10% points given that the maximum score possible on post-test 1 was 43. We also conducted further analysis by considering the well- and ill-structured items on post-test 1 separately. Findings suggested that:

¹ The process findings also double up as a manipulation check demonstrating that students in the PF condition experienced failure at least on the abovementioned conventional measures. In contrast, 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. This is also the reason why the process analyses focuses only on the PF condition.

- i. On the well-structured items, students from the PF class scored higher, $M = 30.8$, $SD = 4.09$, than those from the LP class, $M = 28.9$, $SD = 5.13$. This effect was statistically significant, $F(1,72) = 4.87$, $p = .019$, $ES = .42$. However, this difference amounted to only 6% points (maximum score on these items was 32) with a low effect size. Notwithstanding, it was remarkable that PF students who were not given any homework or practice assignments during instruction still managed to outperform LP students who did receive such practice and feedback on well-structured type of items.
- ii. On the ill-structured item, students from the PF class scored higher, $M = 7.0$, $SD = 3.60$, than those from the LP class, $M = 4.5$, $SD = 3.55$; an average difference of 23% points (maximum score possible on this item was 11). This effect was statistically significant, $F(1,72) = 8.95$, $p = .004$, $ES = .98$.

Thus, students from the PF class outperformed those from the LP class on both the well-structured items and the ill-structured item on post-test 1 thereby suggesting that the productive failure hypothesis held up to empirical evidence.

Post-test 2 Post-test 2 immediately followed post-test 1 and lasted 15 minutes. The objective of post-test 2 was to determine if there were any differences between students from the PF and LP classes in their ability to learn and apply the extension concept of relative speed on their own. Note that the concept of relative speed is more difficult because it involves the motion of two bodies moving at the same time. Not surprisingly therefore, the formal curriculum actually does not cover the concept of relative speed until grade 10. Inspired by the assessment experiment designed by Schwartz and Martin (2004), two versions (A and B) of post-test 2 were created, each comprising 2 items on relative speed. Version A comprised the following two items:

Item 1: *You and your friend start running at the same time from the same position but in opposite directions on a 400m running track. You run at 5m/s whereas your friend runs at 3m/s. In 1s, how many meters:*

- do you travel towards your friend?
- does your friend travel towards you?
- do the two of you travel towards each other in total?
- Therefore, how many seconds will it take for the two of you to first cross each other?

Item 2: *Two MRT trains on separate but parallel tracks are traveling towards each other. Train A is 100m long and is traveling at a speed of 100km/hr. Train B is 200m long and is traveling at a speed of 50km/hr. How many seconds will it take from the time that the two trains first meet to the time they have completely gone past each other?*

For version B, item 2 was exactly the same as in version A, whereas for item 1, parts a, b, and c were removed leaving only part d. Thus, in version A of post-test 2, students first received a structured-response scaffold for item 1 following which the scaffold was removed for item 2. In contrast, both items were unscaffolded in version B of the post-test 2. Further note that item 2 was designed to be conceptually more challenging than item 1; unlike item 1,

item 2 required students to first ascertain the relative distance using the lengths of the two trains before using relative speed to calculate the time.

Recall that all students from both the experimental conditions took the same post-test 1. However, for post-test 2, roughly half of the students in each condition (PF and LP) were randomly assigned to take either version A or B of post-test 2. Thus, version A and B of post-test 2 created a nested factor within the experimental condition (PF or LP), i.e., a 2(PF or LP) by 2(version A or B of post-test 2) design. This created four groups of students, namely, PF-A (18 students), PF-B (19 students), LP-A (19 students), and LP-B (19 students), where PF-A refers to students in the PF condition who received version A of post-test 2, and so on.

Analysis of post-test 2 revealed a bimodal distribution in student scores (very high or very low) on both the items. Students' solutions were therefore categorized as either 'successful' or 'unsuccessful' based on whether they scored high or low respectively for each item; the inter-rater reliability was .98 and .94 (*Krippendorff's alphas*) for items 1 and 2 respectively. Figure 2 shows the percentage of students who were successful or unsuccessful for each item in each of the four groups PF-A, PF-B, LP-A, and LP-B.

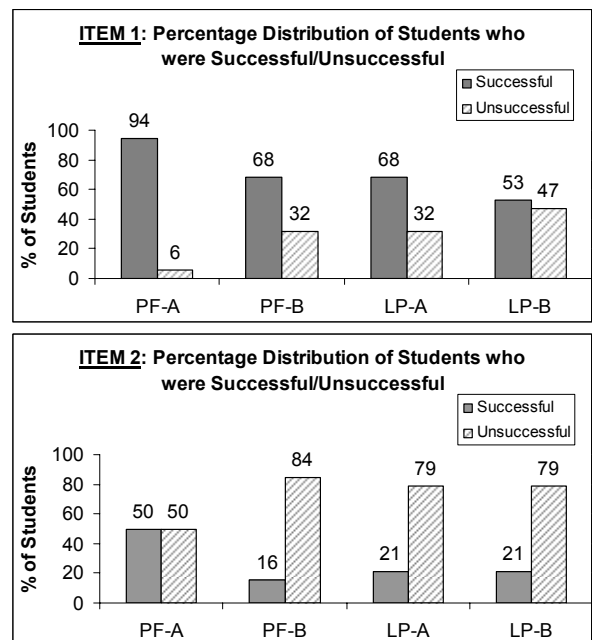


Figure 2: Post-test 2 Success Rate Distributions

For item 1, two findings are noteworthy. First, students from PF-A condition (PF students who received scaffolded version of item 1) had a significantly greater success rate than all the other students, including students from the PF-B condition who did not receive the scaffolded version of item 1, $\chi^2(1) = 7.98$, $p = .005$. Second, students from the PF-B condition who solved item 1 *without* a scaffold were just as successful as students from the LP-A condition who solved item 1 *with* the scaffold. For item 2, the success rates were generally low across the four groups. This was expected because the item was designed to be conceptually more

challenging, as described earlier. Still, the fact that 50% of students managed to solve item 2 successfully was statistically significant, $\chi^2(1) = 6.77, p = .009$. One may argue that a success rate of 50% is not high but considering that the concept of relative speed was not covered during the unit and that this concept is not formally covered until grade 10, a 50% success rate is arguably significant. Finally, although it may seem from Figure 2 that students in the PF-B condition were more unsuccessful in solving item 2 than their counterparts in the LP condition, post-hoc analysis showed that this difference was statistically not significant.

Discussion

Findings from our study suggest: a) despite seemingly failing in their collective and individual problem-solving efforts, students from the productive failure condition significantly outperformed their counterparts from the lecture and practice condition on the targeted content on post-test 1, and b) the productive failure instructional design better prepared students to use the structured-response scaffolds as an opportunity to learn an advanced concept of relative speed that was not even targeted during instruction. Taken together, findings from our study are consistent with and add to a growing body of research underscoring the role of productive failure in problem solving and learning. Note, however, that these findings may only be attributed to the productive failure instructional design as a whole and not to its constituent elements of collaboration, unscaffolded solving ill-structured problem scenarios designed for persistence, and delay of structure (Brown, 1992).

As hypothesized, explanation for the above findings comes from the notion of *knowledge assembly* (Schwartz & Bransford, 1998; Spiro et al., 1992); the idea 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. They were evidently not successful in solving the problems but it may have possibly engendered sufficient knowledge differentiation (Schwartz & Bransford, 1998) that prepared them to better *discern* and understand those very ideas and concepts when presented in a well-assembled, structured form during the consolidation lecture (Marton, 2007). Solving the ill-structured problems influenced how they learned from the consolidation lecture, which, in turn, helped them discern critical and relevant aspects of the concept of rate and speed. It is plausible that having struggled with various representations and methods, they perhaps better understood the affordances of the representations and methods when delivered by the teacher during the consolidation lecture. Furthermore, these students were also better prepared to make use of the structure (in post-test 2) to learn a new concept on their own.

It is of course much too early to attempt any generalization of the claims; the scope of inference technically holds only under the conditions and settings of the respective study and is thus circumscribed by the content domain, communication modality, age-group, and socio-cultural factors, among other factors. Plus, this study essentially presents an existence proof for a productive

failure instructional design and much future research needs to be carried out to provide a fuller explanatory support. Yet, we contend that this study opens up an exciting line of inquiry into the hidden efficacies in ill-structured, problem-solving activities. Perhaps one should resist the near-default rush to structure learning and problem-solving activities for it may well be more fruitful to first investigate conditions under which instructional designs lead to productive failure.

Acknowledgements

The research reported in this paper was funded by a grant (LSL13/06MK) 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.
- Goel, V., & Pirolli, P. (1992). The structure of design problem spaces. *Cognitive Science, 16*, 395-429.
- Kapur, M. (2006). Productive failure. In S. Barab, K. Hay, & D. Hickey (Eds.), *Proceedings of the International Conference of the Learning Sciences*. NJ: Erlbaum.
- Kapur, M. (in press). Productive failure. *Cognition and Instruction*.
- 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.
- Reiser, B. J. (2004). Scaffolding complex learning: The mechanisms of structuring and problematizing student work. *The Journal of the Learning Sciences, 13*(3), 423-451.
- 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.