# Productive Failure in Learning the Concept of Variance

**Manu Kapur**
National Institute of Education, Singapore

**Abstract**
In a study with 140, ninth-grade mathematics students on learning the concept of variance, students experienced either direct instruction (DI) or productive failure (PF), wherein they were first asked to generate a quantitative index for variance without any guidance before receiving direct instruction on the concept. Whereas DI students relied only on the canonical formulation of variance taught to them, PF students generated a diversity of representations and formulations for variance but were ultimately unsuccessful in developing the canonical formulation. On the posttest however, PF students performed on par with DI students on procedural fluency, and significantly outperformed them on data analysis, conceptual insight, and transfer items. These results challenge the claim that there is little efficacy in having learners solve problems targeting concepts that are novel to them, and that direct instruction alone is the most effective approach for teaching novel concepts to learners.

## Introduction

Proponents of direct instruction bring to bear substantive empirical evidence against un-guided or minimally-guided instruction to claim that there is little efficacy in having learners solve problems that target novel concepts, and that learners should receive direct instruction on the concepts before any problem solving (Kirschner, Sweller, & Clark, 2006). Kirschner et al. (2006) argued that "Controlled experiments almost uniformly indicate that when dealing with novel information, learners should be explicitly shown what to do and how to do it" (p. 79). Commonly-cited problems with un-guided or minimally-guided instruction include increased working memory load that interferes with schema formation (Tuovinen & Sweller, 1999; Sweller, 1988), encoding of errors and misconceptions (Brown & Campione, 1994), lack of adequate practice and elaboration (Klahr & Nigam, 2004), as well as affective problems of frustration and de-motivation (Hardiman et al., 1986).

Klahr & Nigam's (2004) often-cited study compared the relative effectiveness of discovery learning and direct instruction approaches on learning the control of variable strategy (CVS) in scientific experimentation. On the acquisition of basic CVS skill as well as ability to transfer the skill to evaluate the design of science experiments, their findings suggested that students in the direct instruction condition who were explicitly taught how to design un-confounded experiments outperformed their counterparts in the discovery learning condition who were simply left alone to design experiments without any instructional structure or feedback from the instructor (I will return to this study in more detail in the discussion section). Further experiments by Klahr and colleagues (Chen & Klahr, 2008; Strand-Cary & Klahr, 2008), and others as well have largely bolstered the ineffectiveness of discovery learning compared with direct instruction (for reviews, see Kirschner et al., 2006).

Be that as it may, the above findings do not necessarily imply that there is little efficacy in having learners solve novel problems, that is, problems that target concepts they have not learnt yet (Schmidt & Bjork, 1992). To determine if there such an efficacy, a stricter comparison for direct instruction would be to compare it with an approach where students first generate representations and methods on their own followed by direct instruction. Expectedly, the generation process will invariably lead to failure, that is, students are rarely able to solve the problems and discover the canonical solutions by themselves. However, this very process can be productive for learning *provided* direct instruction on the targeted concepts is subsequently provided (Kapur, 2008; Koedinger & Aleven, 2007; Schwartz & Bransford, 1998; Schwartz & Martin, 2004).

As a case in point, I present evidence from an on-going research program on *productive failure* (Kapur, 2008; Kapur & Kinzer, 2009; Kapur et al., 2007).

## Designing for Productive Failure

There are at least two problems with direct instruction in the initial phase of learning something new or solving a novel problem. First, students often do not have the necessary prior knowledge differentiation to be able to discern and understand the affordances of the domain-specific representations and methods underpinning the targeted concepts given during direct instruction (e.g., Schwartz & Martin, 2004). Second, when concepts are presented in a well-assembled, structured manner during direct instruction, students may not understand why those concepts, together with their representations, and methods, are assembled or structured in the way that they are (Chi et al., 1988; diSessa et al., 1991; Schwartz & Bransford, 1998).

To overcome these two problems, a learning design should focus squarely on first engaging students in processes that serve two critical cognitive functions, which in turn, prepare students for subsequent direct instruction: a) activating and differentiating prior knowledge in relation to the targeted concepts, and b) affording attention to critical features of the targeted concepts.

Productive failure is one such learning design. It comprises two phases—a generation and exploration phase followed by a direct instruction phase. In the generation and exploration phase, the focus is on affording students the opportunity to leverage their formal as well as intuitive prior knowledge and resources to generate a diversity of *structures*—concepts, representations and solution methods—for solving a complex problem; a problem that targets concepts that they have not been formally taught or learnt yet[1]. Research suggests that students do have rich *constructive resources* (diSessa & Sherin, 2000) to generate a variety of structures for solving novel problems (diSessa et al., 1991; Schwartz & Bransford, 1999). At the same time,

[1] The complexity of the problem is in relation to the learner. The problem is complex to the learner because the learner does not know the canonical representations and methods for solving it. To someone who knows these, the problem is no longer complex.

research also suggests that one cannot expect students, who are novices to the target content, to somehow generate or discover the canonical representations and domain-specific methods for solving the problem (Kirschner et al., 2006).

However, the expectation for the generation and exploration phase is not for students to be able to solve the problem successfully. Instead, it is to generate and explore the affordances and constraints of a diversity of structures for solving the problem. To the extent that students can persist in this process, the process not only activates but also differentiates their prior knowledge (as evidenced in the diversity of student-generated concepts, representations and methods). Furthermore, a comparison and contrast between the various structures also affords opportunities to attend to critical features of the targeted concepts (more on this in results section). Consequently, the generation and exploration phase provides the necessary foundation for developing deeper understanding of the canonical concepts, representations, and methods during direct instruction.

Empirical evidence for PF comes from a series of design experiments in grades seven through nine in Singapore mathematics classrooms (Kapur, 2009a, 2009b; Kapur et al., 2008; Kapur & Lee, 2009). Working with approximately 300 students from four public schools, the studies compared PF and DI designs for a two-week, curricular unit on average speed. Findings suggested that PF students produced a diversity of linked problem representations and methods for solving the problems but were ultimately unsuccessful in their efforts. Despite seemingly failing in their problem-solving efforts, PF students significantly outperformed DI students on both procedural fluency and complex analysis problems on the posttests. Furthermore, PF students also demonstrated significantly better transfer performance in adapting and building upon the targeted concepts to learn new concepts on their own.

These findings are consistent with other research programs that suggest that conditions that maximize performance in the shorter term are not necessarily the ones that maximize learning in the longer term (Clifford, 1984; Schmidt & Bjork, 1992). Examples of such research programs include VanLehn's (2003) work on *impasse-driven learning*, Schwartz and Bransford's (1998) work on *preparation for future learning*, Schwartz and Martin's (2004) work on *inventing to prepare for learning*, diSessa's (1991) work on *meta-representational competence*, Koedinger and Aleven's (2007) work on the *assistance dilemma*, among others (Kapur & Rummel, 2009).

Collectively, these research programs support the argument for designing conditions for learners to persist in the process of solving novel, complex problems without instructional support structures initially. Even though such a process invariably leads to failure in the shorter term, the extent to which this process affords learners opportunities to explore and generate a variety of representations and methods, the process can be germane for learning.

The purpose of this paper is to report findings from an on-going, classroom-based research program on productive failure in a public school in Singapore.

# Method

## Participants

Participants were 140, ninth-grade mathematics students (14-15 year olds) from an all-boys pubic school in Singapore. Students were almost all of Chinese ethnicity. Students were from four mathematics classes; three classes taught by one teacher (teacher A), and the fourth class by another teacher (teacher B). Students had no instructional experience with the targeted concept—variance—prior to the study, although they had learnt the concepts of mean, median, and mode in grades 7 and 8.

## Research Design

A quasi-experimental, pre-post design was used with two classes ($n$ = 31, 35) taught by teacher A assigned to the 'Direct Instruction' (DI) condition, and the other two classes ($n$ = 35, 39), under teachers A and B, assigned to the 'Productive Failure' (PF) condition.

First, all students took a five-item paper and pencil pretest ($\alpha$ = .75) on the concept of variance. Not surprisingly, not a single student demonstrated canonical knowledge of the concept, and there was no significant difference between the four classes either, $F(3,136) = 1.665$, $p$ = .177. Next, all classes participated in four, 55-minute periods of instruction on the concept as appropriate to their assigned condition. After the second and fourth periods, students from all classes took a five-item, five-point (1(low) - 5(High)) Likert scale engagement survey ($\alpha$ = .79). Finally, all students took a six-item, paper and pencil posttest ($\alpha$ = .74) comprising items on procedural fluency, data analysis, conceptual insight, and transfer.

In the DI condition, the teacher first explained the concept of variance and its canonical formulation as the square of the standard deviation ($SD^2 = \frac{\sum_{1}^{n}\left(x_i - \bar{x}\right)^2}{n}$) using a data analysis problem. Next, the teacher modeled the application of the concept by working through several data analysis problems, highlighting common errors and misconceptions, and drawing attention to critical features of the concept in the process. The data analysis problems required students to compare the variability in 2-3 given data sets, for example, comparing the variability in rainfall in two different months of a year, or comparing the consistency of performance of three soccer players, and so on. Thereafter, students worked face-to-face in triads on more data analysis problems. The teacher then discussed the solutions with the class. After each period, students were given similar data analysis problems for homework, which the teacher marked and returned to the students, usually by the following period.

The PF condition differed from the DI condition in only one important aspect. Instead of receiving direct instruction upfront, students spent two periods working face-to-face in triads to solve one of the data analysis problems on their own. The data analysis problem presented a distribution of goals scored each year by three soccer players for a twenty-year period. Students were asked to generate a quantitative index to determine the most consistent player. During this

generation phase, no instructional support or scaffolds were provided. Following this, two periods were spent on direct instruction just like in the DI condition. Note that because students in the PF condition spent the first two periods generating an index for variance, they solved fewer data analysis problems overall than their counterparts in the DI condition. To make this contrast even sharper, PF students did not receive any data analysis problems for homework.

**Hypothesis** The hypothesis tested was that *productive failure will be more effective than direct instruction in learning the concept of variance.* That is, expecting to replicate earlier work on productive failure (Kapur, 2008, 2009; Kapur & Lee, 2009), I hypothesized that students from the PF condition will be able to generate and explore various representations and methods for generating an index for variance (diSessa et al., 1991), but will not be successful in developing or discovering the canonical formulation on their own (Kirschner et al., 2006). However, this seeming failure would be integral for: a) engendering the necessary prior knowledge differentiation (evidenced in the diversity of student-generated structures), and b) drawing attention to critical features of the concept of variance (evidenced in the comparisons between the student-generated structures), which may help students better understand the concept when presented by the teacher during direct instruction subsequently (Schwartz & Bransford, 1998). This better understanding would result in better procedural fluency, data analysis, conceptual insight, and transfer.

## Process Results

Process data included group-work artifacts produced on A4 sheets of paper. These provided a rich source of data about the nature of problem representations and methods generated by the students in the PF and DI conditions.

In the PF condition, groups produced four major and *progressively sophisticated* categories of methods and representations. The four categories were: a) central tendencies, b) qualitative methods, c) frequency methods, and d) deviation methods.

*Category 1: Central Tendencies*. Groups started by using mean, median, and in some cases, mode for data analysis. This was not surprising because students had been taught these concepts in the earlier grades. However, relying on central tendencies alone, it was not possible to generate a quantitative index for variance because the problem was designed in a way to keep the central tendencies invariant.

*Category 2: Qualitative methods*. Groups generated graphical and tabular representations that organized the data visually and were able to discern which player was more consistent. The visual representations (see Figure 1) afforded a qualitative comparative analysis between the players, but did not provide a quantitative index for measuring consistency even though the ideas of spread and clustering are quite evidently important qualitative conceptual underpinnings for the concept of variance.

*Category 3: Frequency methods*. Groups built on the qualitative methods to develop frequency-based measures of consistency. For example in Figure 2, groups used the frequency of goals scored within certain intervals to argue that the player with the highest number of goals in the interval containing the mean was the most consistent. Other groups counted the frequency with which a player scored above, below, and at the mean. Frequency methods demonstrated that students could quantify the clustering and bunching up trends in the qualitative representations.

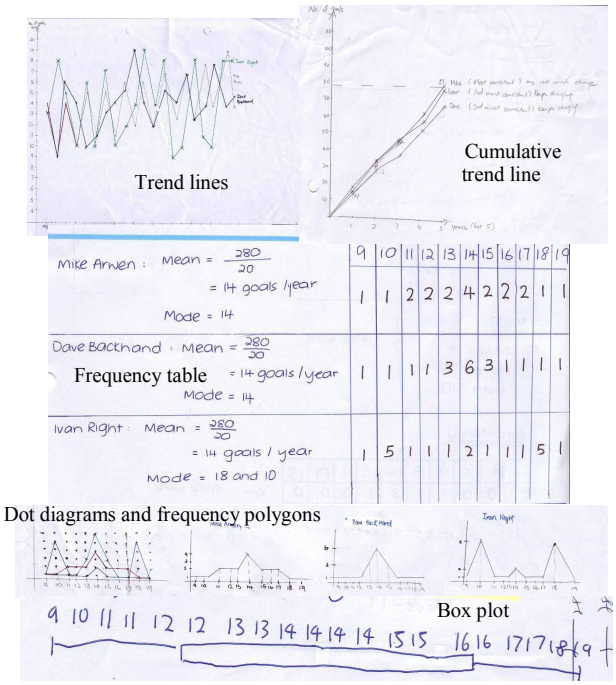


*Figure 1* Examples of qualitative representations/methods

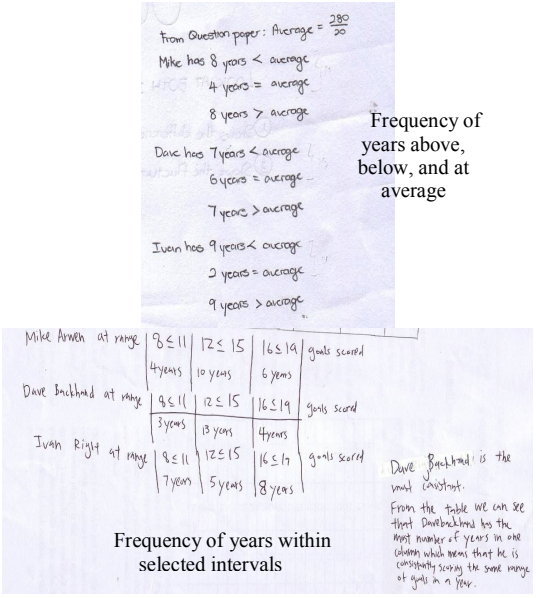


*Figure 2* Examples of frequency representations/methods

*Category 4: Deviation methods*. Figure 3 presents some examples of the deviation methods. The simplest deviation method generated was the range (Deviation method 1, or simply D1). Some groups calculated the sum of year-on-year deviations (D2) to argue that the greater the sum, the

lower the consistency. Among these, there were those who considered absolute deviations (D3) to avoid deviations of opposite signs cancelling each other—an important conceptual leap towards understanding variance. Finally, there were some groups who calculated deviations about the mean (D4) only to find that they sum to zero. For both the D3 and D4 categories, some groups further refined their method to consider not the sum of the deviations, but the average (D5).

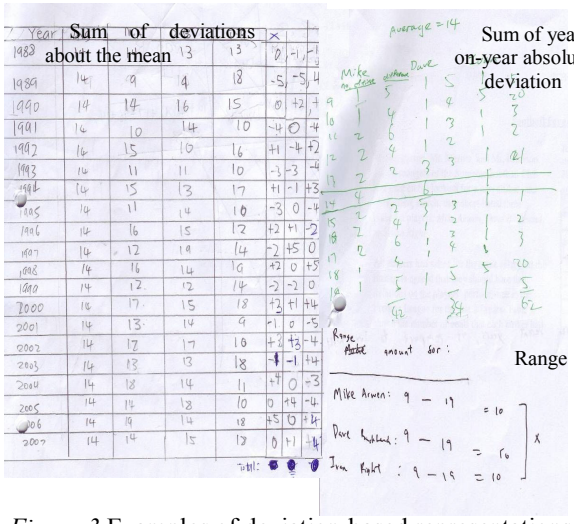


*Figure 3* Examples of deviation-based representations and methods

In both the PF classes, all groups demonstrated representational competence at the Category 3 level or greater. Only 2 groups from PF-A and 1 group from PF-B did not reach Category 4. Consistent with the hypothesis, none of the groups were able to develop let alone use the canonical formulation on their own.

More importantly, note that these structures evidence the hypothesis that students will in fact be able to generate a rich diversity of structures to solve the problem without having first learnt the targeted concept of variance, and that comparisons between these structures will afford students the opportunities to attend to deep conceptual features of the concept. The latter needs more elaboration:

i. Comparing central tendencies with qualitative representations afforded an opportunity to attend to the feature that central tendencies alone cannot convey information about variance, and that different distributions with the same mean can have different variance.
ii. A comparison between the frequency methods and the qualitative methods afforded the opportunity to attend to the quantification of qualitative data into a mathematical index that returns a value for consistency.
iii. Because the deviation methods consider the relative position of a data point, a comparison with the frequency methods afforded students the opportunity to attend to the feature that, for consistency, it is not only important to count a point but also consider its position in relation to other points.
iv. Range (D1) afforded students the opportunity to attend to the feature that considering just the extreme points may not be a good measure of consistency, because it tells us nothing about the distribution in the middle. Comparing D1 with any of the qualitative representations easily afforded attention to this feature.
v. A comparison between D2 and D3 afforded students the opportunity to attend to the feature of why deviations must be positive. The comparison clearly shows that when deviations are left with their signs intact, positive and negative deviations cancel out resulting in a case where the variance could be highly underestimated.
vi. A comparison of D3 and D4 methods afforded students the opportunity to attend to the feature of why the reference point must be a fixed point (e.g., the mean), or else the index is sensitive to ordering of data. If the reference point for the deviation is not a fixed point, then a re-ordering of the data will result in a different value of consistency for the same formulation.
vii. A comparison between the sum and the average afforded the opportunity to attend to the feature of how dividing by the number of data points helps compare samples of different sizes.

In the DI condition, analysis of students' classroom work revealed that all students relied *only* on the canonical formulation to solve data analysis problems. This was not surprising given that the canonical formulation is relatively easy to compute and apply, and was corroborated with data from homework assignments. The average performance (i.e., percentage of problems solved correctly) on the homework assignments was high, $M$ = 93.2%, $SD$ = 5.3%. Finally, on the mean of the two self-reported engagement ratings, there was no significant difference between the PF condition, $M$ = 3.84, $SD$ = .51, and the DI condition, $M$ = 3.82, $SD$ = .43, $F(1, 138) = .035$, $p = .852$.

These process findings serve as a manipulation check demonstrating that students in the PF condition experienced "failure" at least in the conventional sense. In contrast, DI students were not only just as engaged as PF students but also demonstrated successful application of the canonical formulation to solve several data analysis problems. The high engagement ratings and performance results also suggest that the DI condition was not simply a case of poor instruction.

## Outcome Results

**Post-test** The six-items on the posttest comprised:

i. one item on procedural fluency (calculating SD for a given data set),
ii. two items on data analysis (comparing means and SDs of two samples; these items were similar to the data analysis problems covered during instruction),
iii. two items on conceptual insight (one item dealing with sensitivity to ordering of data points, and another with outliers), and
iv. one item on transfer (item requiring the development of a normalized score for comparing incommensurable distributions. Note that normalization was not taught

during instruction, and therefore, students needed to flexibly adapt and build upon what they had learnt.).

Maximum score for each item was 10; two raters independently scored the items using a rubric with an inter-rater reliability of .96. Performance on the four types of items formed the four dependent variables. Controlling for the effect of prior knowledge as measured by the pretest, $F(4, 134) = 1.890$, $p = .112$, a MANCOVA revealed a statistically significant multivariate effect of condition (PF vs. DI) on posttest scores, $F(4, 134) = 16.802$, $p < .001$, partial $\eta^2 = .33$. There was no significant difference between the classes *within* the PF or DI conditions, nor was there any significant interaction between prior knowledge and experimental condition.

i. On the procedural fluency item, there was no significant difference between the PF condition, $M = 7.66$, $SD = 3.97$, and the DI condition, $M = 7.98$, $SD = 3.89$, $F(1, 137) = .819$, $p = .367$.
ii. On the data analysis items, students from the PF condition, $M = 14.11$, $SD = 4.20$, significantly outperformed those from the DI condition, $M = 11.38$, $SD = 4.86$, $F(1, 137) = 10.290$, $p = .002$, partial $\eta^2 = .07$.

It is important to note that PF students who were not given any homework and exposed to fewer data analysis problems still managed to perform on par with DI students on procedural fluency, and better than DI on data analysis in spite of DI students receiving homework and more practice and feedback on data analysis problems during instruction.

iii. On the conceptual insight items, students from the PF condition, $M = 16.40$, $SD = 6.41$, significantly outperformed those from the DI condition, $M = 8.20$, $SD = 6.15$, $F(1, 137) = 51.359$, $p < .001$, partial $\eta^2 = .27$.
iv. On the transfer item, students from the PF condition, $M = 4.93$, $SD = 2.99$, significantly outperformed those from the DI condition, $M = 3.07$, $SD = 2.35$, $F(1, 137) = 14.505$, $p < .001$, partial $\eta^2 = .10$.

## Discussion

These findings are consistent with previous studies on productive failure with other mathematical topics and profile of students (Kapur, 2009a, 2009b; Kapur et al., 2008; Kapur & Lee, 2009), and also with other studies (e.g., Schwartz & Bransford, 1998; Schwartz & Martin, 2004). Notwithstanding the limitations of what can be achieved in a single study carried out within a particular domain, context and classroom-based setting, implications arising from the findings are simple and significant: There is indeed an efficacy in having learners generate and explore representations and methods for solving problems on their own even if they do not formally know the underlying concepts needed to solve the problems, and even if such un-supported problem solving leads to failure initially. The process analysis showed that this seeming failure was integral for: a) engendering the necessary prior knowledge differentiation (evidenced in the diversity of student-generated structures), and b) drawing attention to critical features of the concept of variance (evidenced in the comparisons between the student-generated structures), which may help students better understand the concept when presented by the teacher during direct instruction subsequently (Schwartz & Bransford, 1998).

This study contributes to the ongoing debate comparing the effectiveness of direct instruction with discovery learning approaches (e.g., Kirschner et al., 2006; Klahr & Nigam, 2004; Dean & Kuhn, 2007); discovery learning being often epitomized as the constructivist ideal. It is perhaps worth clarifying that a commitment to a constructivist epistemology does not necessarily imply a commitment to discovery learning. Simply leaving learners to generate and explore without consolidating is unlikely to lead to learning, or at least learners cannot be expected to "discover" the canonical representations by themselves as indeed our findings suggest. Instead, a commitment to a constructivist epistemology requires that we build upon learners' prior knowledge. However, one cannot build upon prior knowledge if one does not know what this prior knowledge is in the first place. It follows that at the very least the burden on the designer (e.g., teacher, researcher) is to first understand the nature of learners' prior knowledge structures; the very structures upon which the claimed "building" will be done. Designing for productive failure presents one way of doing so, wherein students first generate and explore representations and methods, and in the process externalize their prior knowledge structures, before direct instruction.

Interestingly, one could argue that Klahr & Nigam's (2004) study supports the above contention although it is often cited as a stellar example of the superior effectiveness of direct instruction over discovery learning. A careful reading of the study suggests that before assigning students to either a direct instruction or a discovery learning condition, Klahr and Nigam conducted a baseline assessment where they asked students to design four experiments on their own. As expected, only 8 out of the 112 students were able to design four un-confounded experiments, that is, the success rates before any instruction on the control of variables strategy (CVS) were very low. Students who were subsequently assigned to the discovery learning condition simply continued to design these experiments but without any instruction on CVS or any feedback. However, for students in the direct instruction condition, the instructor modeled and contrasted the design of both confounded and un-confounded experiments with appropriate instructional facilitation and explanation to make them attend to critical features of why CVS, unlike confounded experiments, helps isolate the effects of a factor. It was not surprising therefore that Klahr and Nigam found direct instruction to be more effective than discovery learning as described earlier in this paper.

From the perspective of productive failure however, the baseline assessment in Klahr and Nigam's (2004) study seems to function very much like the generation and exploration[2] phase where students generate their own structures (in this case, experiments) to solve a problem that targets a concept (in this case, CVS) that they had not learnt yet. If so, the very effects that Klahr and Nigam attribute to

[2] Indeed, Klahr & Nigam (2004) themselves termed it the "exploration phase."

direct instruction *alone* seem more appropriately attributed to a generation and exploration phase (their baseline assessment) followed by direct instruction. Therefore, much as Klahr and Nigam set out to show, in part, that there is little efficacy in students exploring and solving problems requiring concepts they have not learnt yet, their findings can be reinterpreted to support precisely the opposing contention that such exploration can in fact be efficacious provided some form of direct instruction follows, for without it, students may not learn much (as indeed the performance of the students in the discovery learning condition revealed). Thus argued, designing for a certain level of failure (as opposed to minimizing it) in the initial learning phase may well be productive for learning in the longer run. Future research would do well not to (over)simplistically compare discovery learning with direct instruction, but instead understand conditions under which these approaches can complement each other productively.

## Acknowledgements

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

## References

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

Chen, Z., & Klahr, D. (2008). Remote transfer of scientific reasoning and problem-solving strategies in children. In R. V. Kail (Ed.), *Advances in child development and behavior* (pp. 419–470). Amsterdam: Elsevier.

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

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

Dean, D., & Kuhn, D. (2007). Direct instruction vs. discovery: The long view. *Science Education*, *91*, 384–397.

diSessa, A. A., Hammer, D., Sherin, B. L., & Kolpakowski, T. (1991). Inventing graphing: meta-representational expertise in children. *Journal of Mathematical Behavior*, *10*(2), 117–160.

diSessa, A. A., & Sherin, B. L. (2000). Meta-representation: An introduction. *Journal of Mathematical Behavior*, *19*(4), 385–398.

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

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

Kapur, M. (2009b). The role of productive failure in mathematics teaching and learning. In B. Kaur, B. H. Yeap, & M. Kapur (Eds.), *Mathematical Problem Solving* (pp. 43-68), Singapore: World Scientific.

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

Kapur, M., Dickson, L., & Toh, P. Y. (2008). Productive failure in mathematical problem solving. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 1717-1722). Austin, TX: Cognitive Science Society.

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

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

Kapur, M., & Lee, J. (2009). Designing for productive failure in mathematical problem solving. In N. Taatgen & V. R. Hedderick (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 2632-2637). Austin, TX: Cognitive Science Society.

Kapur, M., & Rummel, N. (2009). The assistance dilemma in CSCL. *Proceedings of the Computer-Supported Collaborative Learning Conference*, Rhodes, Greece.

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

Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction: Effects of direct instruction and discovery learning. *Psychological Science*, *15*(10), 661–667.

Koedinger, K. R., & Aleven V. (2007). Exploring the assistance dilemma in experiments with Cognitive Tutors. *Educational Psychology Review, 19*(3), 239-264.

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.

Strand-Cary, M., & Klahr, D. (2008). Developing elementary science skills: Instructional effectiveness and path independence. *Cognitive Development*, *23*(4), 488-511.

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

Tuovinen, J. E., & Sweller, J. (1999). A comparison of cognitive load associated with discovery learning and worked examples. *Journal of Educational Psychology*, *91*, 334–341.

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