Instr Sci (2014) 42:523–538
DOI 10.1007/s11251-013-9300-7

Making the failure more productive: scaffolding the invention process to improve inquiry behaviors and outcomes in invention activities

N. G. Holmes · James Day · Anthony H. K. Park · D. A. Bonn · Ido Roll

Received: 21 May 2013 / Accepted: 11 November 2013 / Published online: 20 November 2013
© Springer Science+Business Media Dordrecht 2013

Abstract Invention activities are Productive Failure activities in which students attempt (and often fail) to invent methods that capture deep properties of a construct before being taught expert solutions. The current study evaluates the effect of scaffolding on the invention processes and outcomes, given that students are not expected to succeed in their inquiry and that all students receive subsequent instruction. While socio-cognitive theories of learning advocate for scaffolding in inquiry activities, reducing students' agency, and possibly their failure rate, may be counter-productive in this context. Two Invention activities related to data analysis concepts were given to 87 undergraduate students in a first-year physics lab course using an interactive learning environment. Guided Invention students outperformed Unguided Invention students on measures of conceptual understanding of the structures of the constructs in an assessment two months after the learning period. There was no effect, however, on measures of procedural knowledge or conceptual understanding of the overall goals of the constructs. In addition, Guided Invention students were more likely to invent multiple methods during the Invention process. These results suggest that the domain-general scaffolding in Invention activities, when followed by instruction, can help students encode deep features of the domain and build on their failures during Productive Failure. These results further suggest not all failures are equally productive, and that some forms of support help students learn form their failed attempts.

Keywords Invention activities · Productive Failure · Scaffolding · Interactive learning environments

Electronic supplementary material The online version of this article (doi:10.1007/s11251-013-9300-7) contains supplementary material, which is available to authorized users.

N. G. Holmes (✉) · J. Day · A. H. K. Park · D. A. Bonn · I. Roll
Department of Physics and Astronomy, University of British Columbia, 6224 Agricultural Road, Vancouver, BC V6T 1Z1, Canada
e-mail: nholmes@phas.ubc.ca

I. Roll
Department of Educational and Counseling Psychology and Special Education, University of British Columbia, Vancouver, BC V6T 1Z1, Canada

Introduction

The term Productive Failure describes activities in which students generate solutions to novel problems prior to receiving instruction on the same topics. Students may be asked to generate methods that capture, for example, the variability of given data sets prior to being taught about standard deviation (Day et al. 2010; Kapur 2011; Kapur and Bielaczyc 2012). The, so-called, failure stems from the fact that students commonly fail to generate correct methods in these activities. In the variability example, students may take the range of the values or count the number of different values, ignoring the full distribution and number of data points. The failure is often productive, though, as students learn from the subsequent instruction and practice activities better than students who receive only instruction and practice, controlling for overall time on task (Kapur 2011, 2012; Roll et al. 2009; Westermann and Rummel 2012). That is, students who engage in a Productive Failure activity will learn the concept of standard deviation better than students who have longer instruction and practice on standard deviation without an opportunity to first generate a solution. In this way, it is said that the generation of solutions prepares students for the subsequent instruction—or that students are "Inventing to Prepare for Future Learning" (Schwartz and Martin 2004). The failure to solve problems may be productive as the struggle through ill-structured activities with little guidance allows students to explore the problem space and discover the construct better than being given the instruction at the outset (Kapur 2012; Schwartz and Bransford 1998).

Some forms of support or guidance, however, were found to be useful in Productive Failure activities. Invention activities, for example, are a class of Productive Failure activities that are relatively structured (Roll et al. 2009; Schwartz and Martin 2004). In Invention activities, students use carefully designed sets of data, called *contrasting cases*, to invent mathematical methods that capture deep properties of data (Schwartz et al. 2011). Contrasting cases are defined here as a deliberately chosen set of examples where contrasting the individual cases exposes important features about the task at hand. For example, the contrasting cases in Fig. 1 are given to students when asked to create a method for calculating a weighted average. The contrast between Carpenters A and B highlights the role of the spread of the data points. Contrasting Carpenters A and D introduces the role of sample size. The contrasting cases in Invention activities are domain specific and are designed to constrain the inquiry process by focusing the design on key features of the relevant construct. In contrast to many of the contrasting cases analyzed by Alfieri et al. (2013), it is the differences between the cases, rather than the similarities, that identify the critical features of the domain. The use of the contrasting cases in Invention activities also capitalizes on the benefits found in their meta-analysis for case comparisons being used prior to instruction. Schwartz et al. (2011) found that students working with contrasting cases in Invention activities make far more use of the contrasts than students who work with the contrasting cases in tell and practice situations. Simply having the contrasting cases may not be enough, however. It has been found that simply having contrasting cases available to students (without guidance or prompts to engage with them) during Invention activities does not improve their learning over Invention activities without the contrasting cases (Loibl and Rummel 2013). In fact, we have previously found that many students working with Invention activities do not engage with the available contrasts when developing their methods (Roll et al. 2012). This lack of engagement with the available contrasting cases motivates the need for guidance or prompts to help students encode the desired key features of the construct, as recommended by Alfieri et al. (2013).

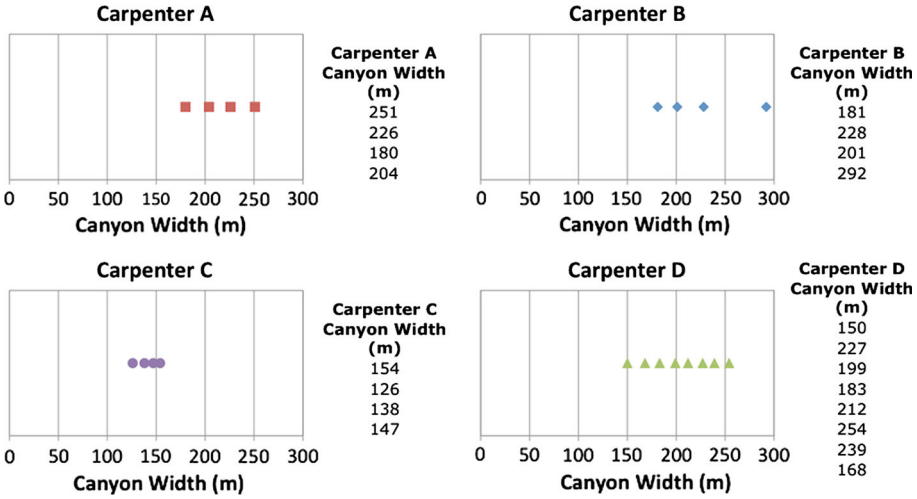

Fig. 1 Contrasting cases emphasize the roles of average, distribution, and sample-size in determining the weighted average of the four cases

The invention process itself, prior to instruction, is set up to resemble an inquiry process, as students attempt to discover the underlying structure of the problem (de Jong 2006; Roll et al. 2012). Students can orient themselves to the problem and generate hypotheses by contrasting the cases and extracting the relevant features of the topic. They can experiment as they generate solutions and apply the solutions to the contrasting cases. A final evaluation and reflection is a key phase in inquiry learning (de Jong 2006). In the invention process, students evaluate their solution by comparing their original hypotheses to the results of applying their solution to the contrasting cases. A cognitive task analysis revealed that experts indeed engage in these processes during Invention activities (Roll et al. 2012). In the absence of explicit support, however, it is of no surprise that students rarely invent valid methods (Kirschner et al. 2006; Matlen and Klahr 2012).

Identifying the correct level, form, and timing of support for inquiry learning remains a challenge (Koedinger and Aleven 2007; Wise and O'Neill 2009). For instance, support in the form of scaffolding in discovery learning activities (Hmelo-Silver et al. 2007; Manlove et al. 2007), prompts or guidance to compare and contrast cases (Alfieri et al. 2013), and prompts to self-explain correct or incorrect reasoning (Chi et al. 1994; Siegler 2002), have all been shown to be productive for learning. Supporting the collaboration process in Invention activities, through the composition and roles of group members, was shown to contribute to learning (Wiedmann et al. 2012; Westermann and Rummel 2012), however, little work has been done to evaluate the effect of guided support for inquiry on learning in Invention activities.

In one such evaluation, it was shown that support for inquiry behaviors improves the invention process and its outcomes (Roll et al. 2012). In that study, scaffolding students' orientation and reflection processes was found to improve the quality of students' invented methods (in terms of the number of included features and the accuracy of the method at ranking the contrasting cases) compared with students who did not receive the scaffolding. Furthermore, students who received the scaffolding were found to apply better invention behaviors as they gave better self-explanations even on unprompted components of the task. Students who received scaffolding also invented more alternative methods (that is,

potential solutions) for each problem. That paper did not, however, evaluate the effect of the scaffolding on students' learning. Thus, it is still unknown whether supporting inquiry during invention leads, in turn, to better learning.

Given that learning takes place mainly during the following instruction and practice, it is unclear whether supporting the invention process itself improves learning from, or preparation for, the subsequent instruction. Unlike other forms of guided discovery, in Invention activities students are not required to succeed in order to learn. Thus, improving their inquiry behaviors during invention may not lead to better learning outcomes. In fact, the given support may short-circuit key cognitive processes that are necessary to achieve a productive failure, and thus may reduce learning. It is, as yet, unclear what these key cognitive processes are, however, and so understanding the outcomes of scaffolding Invention activities on learning may provide a clearer understanding of the mechanisms through which Invention activities improve learning from subsequent instruction and practice.

Explaining the benefits of invention activities

Explanations for the effects of Invention and Productive Failure activities focus on the failure itself and on preparation for learning. It can be compared to impasse-driven learning (VanLehn 1988), which attributes learning to moments when students get stuck during a learning activity (that is, reach an impasse). In this way, failures or impasses could also be considered "desirable difficulties" (Bjork 1994). Reaching and identifying an impasse can reveal the gaps in the students' knowledge and reasoning up to that point, similar to assessing knowledge through testing effects (Roediger and Karpicke 2006). With the knowledge of what they do not know, rather than what they do know, they are better prepared to learn from the instruction and encode the missing features of the construct. Under this explanation, supporting the invention process may lead to more successful attempts, since it has previously been shown that such scaffolding increases the quality of invented methods (Roll et al. 2012). Thus, the scaffolding may have no effect on learning, or may even be counter effective. It may be that an initial failure is sufficient for learning, and support to help students overcome the failure will have *no effect* on learning. The scaffolding may, instead, reduce the failure or rates of impasse and subsequently *reduce* learning from the follow-up instruction.

Impasse-driven learning also attributes learning to the repair process once an impasse has been reached, either through using prior knowledge to get oneself past the impasse or through external assistance (VanLehn 1988). That is, beyond simply the act of failing, the key mechanism may, instead, be the process of attempting to overcome the failure. Attempting to repair the impasse uncovers rules or concepts about the learning activity as it requires students to engage more fully with the design space to determine the reasons for which they have reached the impasse (Roll et al. 2010). Schwartz et al. (2007) suggest that by doing so the invention process helps students gain essential experiences that prepare them to encode the subsequent instruction. In addition, Kapur and Bielaczyc (2011) explain that the process of solving novel problems gives students an opportunity to compare and contrast various solutions, which reveals critical features of the construct that may not be salient in the expert solution alone. Put together, attempting to repair failures during the invention process requires students to evaluate their failed attempts and reexamine the contrasting cases. In doing so, students spend more time deliberately engaged with the cases, which gives them more of an opportunity to extract the features of the construct and identify better mathematical tools to capture these features (Roll et al.

2011a). Thus, it may be the features, and not the overall solution, that are learned during the invention process and transfer to improve learning from subsequent instruction.

If attempts to overcome failure during Invention activities help students identify the key features of the domain, then supporting students' inquiry process, especially through exploration of the design space, could *improve* their subsequent learning from the instruction and practice. Designing support that encourages this exploration is, however, non-trivial. It has previously been found that using domain-specific guidance (that is, prompts and support that are specific to the subject or concept of the activity) improves content learning over support that is independent of the subject or concept, domain-general support (Bulu and Pederson 2010). The opposite was true, however, for learning meta-cognitive skills such as to monitor or evaluate the problem solving process (that is, domain-general prompts were better than domain-specific prompts at engaging students with these skills). If learning is a result of trying to overcome failure through improved inquiry processes, then domain-general scaffolding could maintain the initial rate of failure, but then support students' reasoning through the repair process. Under this explanation, the scaffolding would *increase* learning from the follow-up instruction, but also be mediated by a greater use of the design space.

The act of generating a solution at all is a key component to engaging with the design space. Several studies found that students who invent their own methods learn better than students who evaluate predesigned methods (Kapur and Bielaczyc 2011; Roll et al. 2011b). Inventing a variety of different solutions, rather than merely inventing a single solution, seems to be indicative of larger exploration of the design space. In particular, it would require the students to extensively move through the inquiry process as they evaluate and reflect on their different solutions. Indeed, several studies have found a relationship between the number of students' inventions and their learning outcomes (Kapur 2012; Kapur and Bielaczyc 2012; Loibl and Rummel 2013; Wiedmann et al. 2012). Previous work has also found that domain-general scaffolding in Invention activities increases the number of invented solutions and improves the quality of students' solutions, especially with regards to the number of features that are included (Roll et al. 2012). Thus, measuring students' inquiry process could be done by examining whether students invent multiple solutions, which one would expect to be generated as they move past initially incorrect solutions to develop more correct ones.

A third class of explanations focuses on motivation, generation, and agency. Belenky and Nokes-Malach (2012) found that generating solutions during Invention activities leads students to adopt more mastery-oriented goals for their learning. Adoption of these goals was also found to predict better performance on a transfer task. It is possible that it is the adoption of the mastery-oriented goals, and not the Productive Failure process during the activities, that promotes learning. If, under this explanation, high agency is essential for the invention effect, then reducing students' autonomy by scaffolding the invention process may *reduce* learning outcomes.

Evaluating the effect of scaffolding on learning can help us understand the mechanisms through which Invention and Productive Failure activities improve learning. The current study does so by evaluating the effect of scaffolding during Invention activities in two ways. First, we evaluate the effect of scaffolding on learning outcomes from the overall invention-instruction-practice process. We do so by comparing performance on an in-class assessment administered 2 months after their initial learning period. Second, we evaluate whether scaffolding supports students' invention process itself by measuring the likelihood that groups invent more than a single solution.

The current study extends our previous work in several ways (Roll et al. 2012). First and foremost, the former study evaluated the quality of students' inventions but did not evaluate learning. Second, neither of the Invention activities used in the current study was included in the former and, thus, the current study improves the external validity and generalizability of our results. Thirdly, students completed the Invention activities on a computer-based learning environment in the current study, which provides the log data of students' actions while they work on the activities, giving much more information about whether they invented multiple methods (described below in the follow sections).

Method

Design and procedure

We compared two forms of Invention activities using an in vivo pre-to-post design. The study was spread across a three-month term with the pre-test and two Invention activities given in three subsequent weeks at the beginning of the term. The two Invention activities covered topics of least-squares fitting and weighted averages. The final post-test was delivered at the end of the term, roughly 2 months after students had finished the second invention activity. Students worked on the Invention activities in class in pairs or groups of three, since it has been found that a level of peer interaction during these activities is beneficial for developing ideas (Roll et al. 2012; Westermann and Rummel 2012; Wiedmann et al. 2012). Students were randomly assigned to these groups and different groups were assembled for the two activities within the conditions. The activities were carried out on the computer through a dedicated computer-based learning environment, the Invention Support Environment (ISE; Holmes 2011), with one computer per group. The ISE is described in the "Materials" section, below.

Students were given approximately 30 min to complete each Invention activity, at which point the course instructor delivered a short lecture on the target domain, which included a discussion to direct students' attention to the important features of the data. Following the direct instruction, students worked on physics lab experiments for roughly two more hours. The analysis of data from these experiments provided opportunities for students to practice applying the expert solution from the Invention activity, adding an extra 10–15 min of practice with the construct. Topics from the Invention activities were revisited or built on in subsequent weeks through additional Invention activities (for example, a weighted least-squares activity followed the two presented here to combine the concepts of weighting and fitting). Students also used the two concepts in the analysis of other lab experiments later in the course.

Conditions

Students were randomly assigned, within each lab section, to one of two conditions: an Unguided Invention (UI) and a Guided Invention (GI) condition. Students in the UI condition acted as a control condition and received "traditional" Invention activities following the original Inventing to Prepare for Future Learning paradigm (Schwartz and Martin 2004). Students were first given contrasting cases (shown in the left side of Fig. 2) and a cover story (described in Online Resource 1). Students then were asked to invent a method. While the students had space to produce self-explanations, they were not explicitly prompted to do so. Following their invention, students were asked to apply their

methods to the contrasting cases using calculators or spreadsheet software, and to copy the final values back to the ISE. They were also invited to revise their method if they saw the need. Students in the GI condition received additional scaffolding in the form of explicit prompts modeled after the scaffolding that was used by Roll et al. (2012). The scaffolding was designed to promote expert scientific inquiry behaviors that were identified in a prescriptive cognitive task analysis using similar invention activities on a related topic. The scaffolding was domain-general in that the prompts focused on general inquiry strategies, were independent of any content language, and were applied across activity topics. Within the scope of this study, we chose to focus on scaffolding two key phases that bracket the invention process: orientation and reflection.

Orientation The goal of this level of scaffolding was to get students familiar with the contrasting cases prior to beginning to invent. As discussed in the introduction, deliberate guidance in comparing and contrasting the cases is recommended for students to encode the critical features of the domain to include in their invented solutions. In order to ensure students made use of the contrasting cases, we developed and validated prompts, following a prescriptive cognitive task analysis, that help students *orient* themselves to the given data (Roll et al. 2012). This is done by prompting students to compare pairs of contrasting cases and rank them according to the target concept (e.g., the fit line in the top case in Fig. 2 goes through more data points than the line in the second case). Students were then prompted to briefly explain each of the pairwise comparisons. For example, students would be asked to compare Advisor A and C in Fig. 2 to determine which line is a better fit to the data. In this example, students are confronted with the misconception that a best fit line should go through as many points as possible, since the line for Advisor C looks like a more balanced fit to the data. Once they obtain a conceptual grasp on the problem, they need to determine whether and how to incorporate this aspect into their mathematical solution.

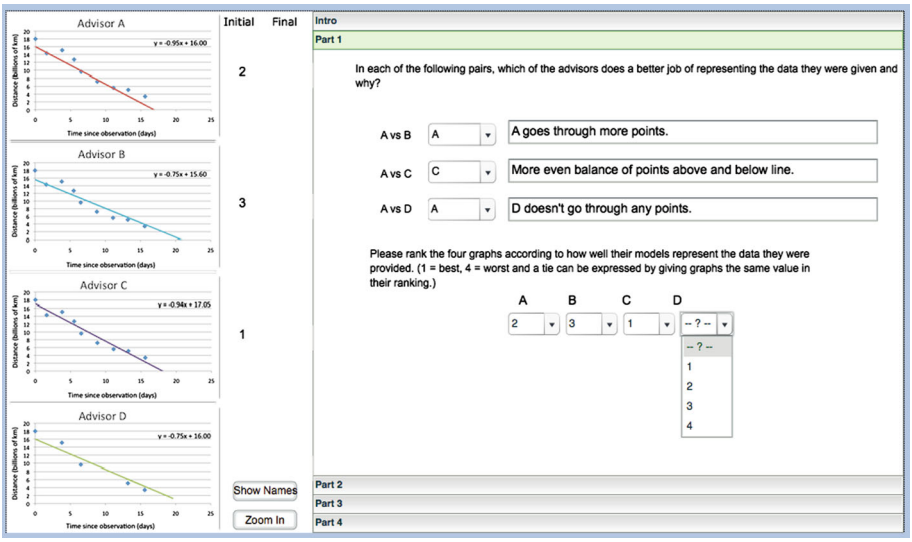

Fig. 2 A snapshot of the ISE for the 'Planet Phaedra' activity shows the contrasting cases on the left-hand side and the activity prompts in the accordion structure on the right-hand side. In this image, the orientation phase is featured with the drop-down menus to select cases for the pairwise contrasts and rankings and the automatic display of the rankings next to the cases. Only the GI condition received the orientation phase

Table 1 A comparison of prompts given to both conditions across different phases of the inquiry process

Invention phase	Guided Invention	Unguided Invention
Orientation	- Introduction story - Pairwise comparisons with self-explanation - Qualitative ranking of all data sets	- Introduction story
	- Invent a formula	- Invent a formula
Reflection	- Explain the invented formula - Implement the solution for all data sets - Explain discrepancies between intuitive and formula-based rankings - Instruction to revise as needed	- Implement the solution for all data sets - Instruction to revise as needed

Reflection A second process that we chose to focus on was *evaluation and reflection*. Once students developed their methods the scaffolding explicitly prompted them to explain how their invented methods related to the pairwise comparisons during the orientation. Students then applied their invented method to the contrasting cases and then were prompted to evaluate their method by comparing these results to the qualitative rankings they identified intuitively in the orientation phase.

It should be noted that while the UI group did not have explicit prompts to perform these particular steps, they still had the opportunity to engage in them spontaneously. For example, the implementation process often leads naturally to reflection as students recognize the shortcomings of their formulas, especially if the students spontaneously analyzed the contrasting cases first. Thus, the main difference between the conditions is the explicit prompting to carry out and reflect on each of the key stages. Table 1 summarizes the differences between conditions. Snapshots of the entire process can be found in Online Resource 2.

Materials

The Invention Support Environment Students in both conditions completed the Invention activities using the ISE (Holmes 2011), a computer-based learning environment that facilitates the invention process. The ISE interface has two main components. On the left-hand side of the interface (shown in Fig. 2), students see a graphical representation of the contrasting cases, while the right-hand side contains the stages of the invention activity that students work through. The stages were set in an accordion structure so that students could easily move back and forth without losing what they had input thus far. In this way, it is analogous to having the stages of the activity on different pages of an activity booklet. The contrasting cases, on the other hand, remained visible on the left-hand side as the student progressed through the activity so that they could always be used to influence their invention process. Online Resource 2 shows additional screenshots of the different stages of the ISE. In addition, students can click the "Zoom In" button to get a detailed view of the contrasting cases and their data.

The first "page" of the according included a cover story that described the contrasting cases and set up the task at hand. GI students then received the orientation page in which they were asked to rank selected pairs of contrasting cases (with prompted self-explanations) and then all four cases. The full ranking was automatically displayed next to the

respective cases (under the heading "Initial" as in Fig. 2) so that they could be seen throughout the invention process.

Students in both conditions then moved on to the design space where they invent their methods (featured in Fig. 3). This featured an equation editor for students to input their inventions and a text area for students to input any comments or explanations (this was visible to the UI students, but they were not prompted to input any text in that space). Figure 3 shows a sample method in the equation editor, which supports a combination of computer keyboard input and mathematical symbols that could be selected from a toolbar along the top.

The application stage of the activities, where students apply their invented methods to the contrasting cases provided, was not supported in the ISE, so students turned to spreadsheets or calculators to apply their methods and returned to the ISE to input their final values. The final rankings, also for the GI condition only, were input in the same way and were then displayed alongside the orientation rankings (under the heading "Final" as in Fig. 2) so that comparisons could be easily made during the evaluation phase. The final stage in the GI condition was the reflection stage, described above. Once all students had worked through the stages of the activity, they were invited to continue adjusting their invented methods if desired, which was again supported by the accordion structure.

Invention activities Two Invention activities were included in this study. 'Planet Phaedra' asked students to find a way to quantify the goodness of fit of a line to given data. 'Not so Grand Canyon' asked students to find a way to weigh measurements by the spread of the data when combining them into an average. The full activities are described in Online Resource 1. Figure 3 shows a snapshot of the system from the second activity with an example of a student solution and explanation.

Assessments We evaluated learning outcomes using a statistics test that was created by the researchers and course instructor. For both pretest and posttest, students were not informed about the test in advance. Notably, the post-test was administered 2 months after

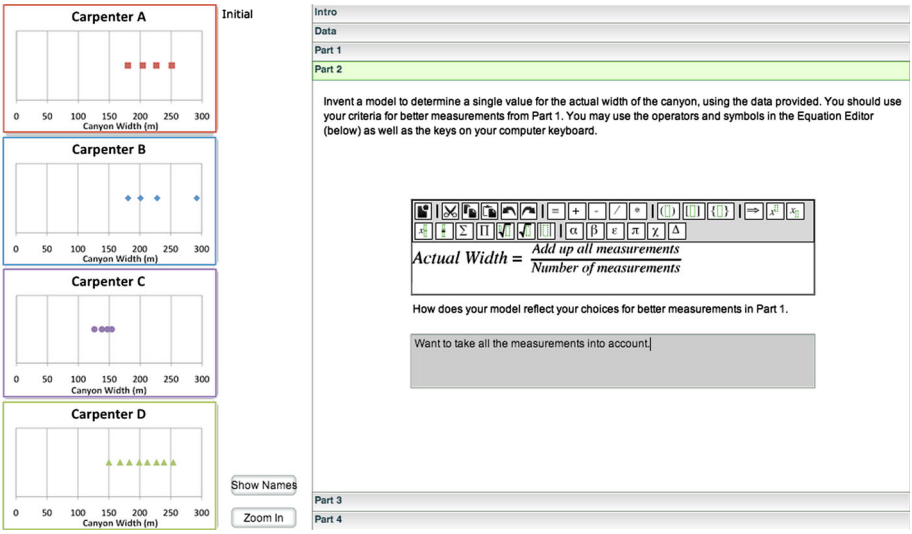

Fig. 3 A snapshot of the ISE for the 'Not so Grand Canyon' activity shows the equation editor, which students used to invent their methods, and a self-explanation text area. Both conditions received the text area but only GI students received prompts to self-explain their invented solutions

the second activity. On the days of the assessments students were told that their scores on the test could contribute positively to their course grade if they did well but could not negatively affect their grade. Students completed the pre- and post-test individually.

The test included three types of questions with each question type included for both invention topics. *Procedural* items asked students to calculate numeric answers by applying the domain formulas. *Conceptual* items asked students to apply the concepts without calculation to demonstrate understanding of the basic principles of the domains. *Knowledge-decomposition* items provided students with equations that were deliberately varied from the expert formulas and asked students to evaluate whether the formulas were reasonable ways to accomplish the same task. Successfully answering the knowledge-decomposition items requires a deep understanding of the structure of the formulas and the mapping of their components on the key features of the domain. While the conceptual questions ask students what the equation means or how it can be generally used or applied, the knowledge-decomposition questions ask for fine-grained analysis of how the formulas achieve their goals. This type of item was previously found to be useful in detecting the effect of Invention activities (Roll et al. 2011b), as students must decompose the target concept into its individual functional components. A sample question and solution were provided to the students. Procedural and conceptual items were multiple-choice and were common to both the pre- and post-test. The multiple-choice options were developed through think-aloud interviews conducted by the primary author with students not included in the study. The students completed open-ended versions of the items and common incorrect responses were used as distractor options. The knowledge-decomposition items were open-ended questions that were unique to the post-test. Scoring of these items is described in the "Data analysis" section below. There were, therefore, four multiple-choice items common to the pre- and post-test, and an additional two open-ended items on the post-test only. The test items and example solutions to the knowledge-decomposition items can be found in Online Resource 3.

Log files of students' actions in the ISE were used to measure the inquiry process with regards to whether students invented more than a single solution. Each submission into the equation editor was analyzed to observe whether its contents showed a change in the invented solution from the previous version during the course of the activity.

Participants

130 first-year students from four sections of a first-year physics laboratory course at the University of British Columbia participated in the study. The prerequisite course included Invention activities that required students to develop graphical representations of data. Thus, all students were familiar with Invention activities, but had not yet worked with activities that required the development of mathematical models and had not yet worked with the ISE. The lab course supplements concurrent honors physics courses that historically attract very high-achieving students. The students do not yet demonstrate strong data handling skills, however, as assessed through the Concise Data Processing Assessment (CDPA; Day and Bonn 2011) given to students at the beginning of the term.

Only students who completed both activities were included in the analysis of the learning outcomes. Even though assignment to conditions was random, many more students were missing from the GI group. Specifically, 38 students from the GI group missed at least one activity (out of 65 students) and 5 students from the UI group missed at least one activity (out of 65 students). To evaluate the risk of self-selection, we compared the subset of students who missed an activity to the other students using two measures: pre-test

scores and scores on the CDPA at the end of the term. On both measures, two-sample, equal-variance t-tests found no differences between students who missed an activity and students who completed both activities across the entire student group and for the GI-assigned students only.[1] Overall, 27 students in the GI group and 60 students in the UI group who completed both activities were included in the analysis of the learning outcomes. Log data from all groups of students for each Invention activity was included in the analysis of the invention process.

Data analysis

The procedural and conceptual test items were multiple-choice and scored binary for each item (correct or incorrect choice). Students could, therefore, obtain a maximum of two points for each question type (one coming from each of the activity topics). Knowledge-decomposition items were open-ended and analyzed through binary scoring such that students either correctly or incorrectly identified the conceptual error in the formulas. Analysis was done by two of the authors with an agreement of 92 %. An agreement regarding the remaining answers was reached in a discussion. A two-sample equal-variance t test was used to check equivalence between conditions on pre-test. Learning from pre- to post-test across conditions was evaluated using paired t-test on items that were shared by both tests. The effect of scaffolding on learning outcomes was evaluated using a single MANCOVA with post-scores for procedural, conceptual, and knowledge-decomposition items as dependent variables, condition as a factor, and controlling for pre-test scores.

Since the Invention activities themselves were carried out in groups, analysis of the learning process was done on the groups of students in each activity. The effect of scaffolding on the learning process was evaluated by measuring whether the groups created more than one method in the process of inventing a model. We fit a logistic regression model to evaluate the log likelihood of inventing more than a single method as a function of condition. We report the slope (B), error (SE B), odds ratio (e^B), Z value, and p value. $B > 0$ suggests that groups in the GI condition were more likely to invent multiple methods. A somewhat more intuitive interpretation of the results is suggested by the odds ratio, e^B. An odds ratio of 1 corresponds to no effect, an odds ratio bigger than 1 corresponds to a positive effect for the scaffolding, and values between 0 and 1 correspond to a negative effect.

Results

The two groups did not differ on pre-test: $t(85) = 1.14$, $p = .258$ (see Table 2), suggesting that students in each condition were initially equivalent in their content knowledge. A paired t test found significant learning, across conditions, from pre-test to post-test on both the procedural and conceptual items (which were shared by both tests, see Table 2). The effect sizes in both cases were small to medium: procedural: $t(86) = 3.23$; $p = .002$, $d = 0.35$; conceptual: $t(86) = 2.82$; $p = .006$, Cohen's $d = 0.30$.

[1] Pre-test differences between students who missed at least one activity and students who completed both activities for both conditions: $t(128) = 0.59$, $p = .557$; pre-test differences for the GI condition: $t(63) = 1.672$, $p = .099$; CDPA10973-differences for both conditions: $t(128) = 0.84$, $p = .402$; CDPA differences for GI students: $t(63) = 0.17$, $p = .864$.

Table 2 Overall pre- and post-test scores on procedural and conceptual items (Standard deviations are presented in parentheses)

Item type	Pre-test	Post-test
Procedural	32 % (33 %)	47 % (29 %)**
Conceptual	63 % (33 %)	74 % (27 %)**

* $p < 0.05$, ** $p < 0.01$

Table 3 Mean pre- and post-test scores on procedural, conceptual, and knowledge-decomposition items by condition (Standard deviations are presented in parentheses)

Item type	Unguided Invention ($N = 60$)	Guided Invention ($N = 27$)
Pretest		
Procedural	28 % (31 %)	33 % (32 %)
Conceptual	66 % (34 %)	59 % (35 %)
Posttest		
Procedural	47 % (30 %)	48 % (26 %)
Conceptual	73 % (28 %)	74 % (25 %)
Knowledge-decomposition	20 % (28 %)	35 % (33 %)*

* $p < 0.05$

Learning outcomes

The overall MANCOVA evaluating the effect of scaffolding on learning, controlling for pre-test scores, was significant, $F(3,82) = 41.76$, $p < .001$. Separate ANCOVAs for the three dependent variables found no significant effect for condition on procedural or conceptual knowledge (see Table 3): procedural: $F(1,84) = 0.05$, $p = .817$, $\eta^2 = .001$; conceptual: $F(1,84) = 0.08$, $p = .784$, $\eta^2 = .001$. However, condition had a significant effect on knowledge-decomposition items, with the GI condition outperforming the UI condition, $F(1,84) = 4.46$, $p = .038$, $\eta^2 = .050$, which is a medium effect size. To address the issue of the differing sample sizes, Levene's test for homogeneity of variance was evaluated during the MANCOVA and found that the variances were not statistically different for the three dependent variables: procedural: $F(1,85) = 1.38$, $p = .243$; conceptual: $F(1,85) = 0.05$, $p = .945$; knowledge-decomposition: $F(1,85) = 1.24$, $p = .268$.

Learning process

Overall, 56 groups of students worked on the first activity and 55 groups worked on the second. A logistic regression model found that groups in the GI condition were significantly more likely to create multiple methods, controlling for task, GI = 51 %, UI = 38 %; $B = 1.13$, $SE(B) = 0.56$, $e^B = 3.09$, $Z = 4.02$, $p = .045$.

Discussion

Analysis of learning by condition demonstrated that scaffolding the Invention activities led to a higher rate of multiple methods during the invention process and to improved learning outcomes of knowledge-decomposition items 2 months after the initial learning period.

The scaffolding had no effect on procedural and conceptual items. This is not surprising since the benefits of Invention activities are often on knowledge-decomposition items and less so on conceptual or procedural skills (Roll et al. 2009; 2011a, b; Schwartz and Martin 2004). The primary challenge during the Invention activities is in decomposing the construct into the key features in order to build a comprehensive solution, which is more strongly tied to the decomposition knowledge, where improved learning was observed. Thus, modifying the invention process amplified the familiar benefits of these activities.

One key question to be answered is how the scaffolding resulted in the observed improvements. In the introduction, it was suggested that if the effects of Invention activities relative to direct instruction were attributable to the preparation caused by the failure to invent the correct method or to students' motivation and adopted mastery goals, then scaffolding the invention process to promote more successful Invention attempts would result in no effect or in *reduced* learning. Since learning was in fact increased, a more likely explanation relates to the importance of connecting students to the key features of the domain through improved inquiry processes as students attempt to repair their failures (VanLehn 1988). In the introduction, it was described that under this explanation the scaffolding would support students as they reason through a repair process once they have reached an impasse (or a failure) and thus increase learning. Results showed that students in the GI condition not only demonstrated increased learning, but also that they were more likely to invent multiple methods. Inventing more than one solution could occur as students move past initially incorrect solutions to develop more correct ones. The improved learning for the GI group was only found on items that required a deep understanding of the key features and functional components of the domain. The support provided here focused on the use of the contrasting cases, and, thus, identification of the key features, through orientation and evaluation. Indeed, it has been previously shown that guided comparisons of contrasting cases can improve learning (Alfieri et al. 2013; Loibl and Rummel 2013; Rittle-Johnson and Star 2009) and the given scaffolding during Invention activities was shown to increase the number of features that students include in their methods (Roll et al. 2012). During the activities, failure was most likely recognized if the outcomes of students' application (and the quantitative, final ranking of the cases) did not match the original orientation goals (and qualitative ranking). The self-explanation and evaluation prompts in the reflection phase highlighted what was known versus what was yet to be learned. Thus, the combination of a high-quality orientation, through guided comparisons of the cases, followed by a prompted evaluation to compare the outcomes of the initial and final rankings may have prompted students to return to their solutions and the cases and attempt to invent a better solution. In this way, the scaffolding may have supported students to recognize their failures, identify knowledge gaps and missing features of the domain, and then continue working to correct the failure by looking for better mathematical manipulations that incorporate these features (Roll et al. 2012).

While the explanations described above seek to explain what transfers from invention to instruction, perhaps we should look at what transfers from invention to assessment. An alternative explanation suggests that inquiry skills, and not domain knowledge, are learned during the activities and contribute to the increased number of methods and the subsequent improved *performance* on the knowledge-decomposition items. Students in the GI condition were asked to explain and evaluate their solutions. These scaffolding prompts could provide the students with opportunities to practice the knowledge-decomposition questions—opportunities that the UI students did not receive. Several reasons make this explanation less likely. First, it has been previously shown that students who practice evaluation of faulty methods do not do better on isomorphic decomposition items than

students who generate their own methods (Kapur and Bielaczyc 2011; Roll et al. 2011b). Second, the two-month delay between learning and testing makes it very unlikely that a general strategy would transfer after being practiced only twice. Transfer of self-regulated learning strategies usually requires much more practice, compared with transfer of domain knowledge (Roll et al. 2011a, b).

This additional variable of long-term retention has not been studied in the context of Invention activities in college-level classrooms and so the interaction between instruction and delay of assessment is especially of interest. A study by Strand-Cary and Klahr (2008) found a striking effect for delay of test on students' learning. Specifically, comparing high-scaffold with low-scaffold instruction on an immediate test found benefits for high-scaffold, while the effect was washed out in a delayed test. The fact that the effect in the current study was still significant after a long delay raises an important question about the direction and strength of the effect on immediate assessments. We aim to address this question in future studies by analyzing students' responses to the practice questions that follow the Invention activity and instruction. Future work should also evaluate the relationship between quality of invention and quality of learning.

The results also demonstrated that students developed a broad understanding of weighted averages and least-squares fitting across the term through a combination of Invention activities, instruction, and ongoing practice opportunities. Students entered the term with a moderate understanding of the concepts and how they relate to certain data situations, but they demonstrated little fluency with how to apply the concepts to data. By the end of the term, there were significant, although small, learning gains in both areas.

The study, more importantly, demonstrates that Invention activities work not simply because support should be delayed. Instead, it is the transmission of domain knowledge that should be withheld, while supporting the inquiry process is beneficial for learning. Of course, there may be other forms of support, or timings to provide the support, that are additionally beneficial for learning using the Productive Failure paradigm (Wise and O'Neill 2009). These could include supporting students' collaborations (Westermann and Rummel 2012) or only providing support in response to the models that students invent (Roll et al. 2010). Further research could be carried out to assess whether additional forms of support are beneficial for learning, if there is an optimal level of support, or if effects differ for activity topics other than data analysis. One practical benefit for the support that was discussed in this paper is its domain general nature. This support could be reused, as we have done, across topics. Our previous study (Roll et al. 2012), suggests that the scaffolding also works across mediums, be it on paper or through a computer-based learning environment, such as the ISE. The scaffolding also complements the use of Invention activities in science labs, as was done here, as it reinforces the invention process as an inquiry process in which students engage in iterations of design and evaluation.

The current study found that scaffolding students' orientation and reflection in Invention activities improves the learning process and outcomes. These results are important for several reasons. First, they demonstrate a direct link between students' invention process and learning outcomes. Second, they help identify forms of support that can improve learning in inquiry activities without short-circuiting critical elements of students' reasoning and the Productive Failure process. The study further suggests that learning improves because students who invent become more likely to notice the deep features of the domain and look for mathematical manipulations that incorporate these features, especially once they have recognized a failure and try to overcome it. Notably, adding guidance during Invention activities helps learning even though students commonly fail to invent the expert solutions. Thus, not only is the failure to invent, indeed, productive, but

also attempts to overcome these failures before instruction may be more so. This study demonstrates how engaging students with good scientific practices helps them achieve a more productive failure.

Acknowledgments This work was supported by the Pittsburgh Science of Learning Center, which is funded by the National Science Foundation, award number (#SBE-0836012), and by the University of British Columbia through the Carl Wieman Science Education Initiative.

References

Alfieri, L., Nokes-Malach, T. J., & Schunn, C. D. (2013). Learning through case comparisons: A meta-analytic review. *Educational Psychologist, 48*(2), 87–113. doi:10.1080/00461520.2013.775712.

Belenky, D. M., & Nokes-Malach, T. J. (2012). Motivation and transfer: The role of mastery-approach goals in preparation for future learning. *Journal of the Learning Sciences, 21*(3), 399–432. doi:10.1080/10508406.2011.651232.

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). Cambridge: The MIT Press.

Bulu, S., & Pedersen, S. (2010). Scaffolding middle school students' content knowledge and ill-structured problem solving in a problem-based hypermedia learning environment. *Educational Technology Research and Development, 58*(5), 507–529. doi:10.1007/s11423-010-9150-9

Chi, M. T. H., De Leeuw, N., Chiu, M.-H., & Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18*(3), 439–477. doi:10.1016/0364-0213(94)90016-7.

Day, J., & Bonn, D. (2011). Development of the concise data processing assessment. *Physical Review Special Topics, 7*(1), 010114. doi:10.1103/PhysRevSTPER.7.010114.

Day, J., Nakahara, H., & Bonn, D. (2010). Teaching standard deviation by building from student invention. *The Physics Teacher, 48*(8), 546. doi:10.1119/1.3502511.

de Jong, T. (2006). Scaffolds for scientific discovery learning. In J. Elen, R. E. Clark, & J. Lowyck (Eds.), *Handling complexity in learning environments: Theory and research* (pp. 107–128). Howard House: Emerald Group Publishing.

Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist, 42*(2), 99–107. doi:10.1080/00461520701263368.

Holmes, N. G. (2011). *The invention support environment: using metacognitive scaffolding and interactive learning environments to improve learning from invention.* Circle: UBC's Digital Repository: Electronic Theses and Dissertations (ETDs) 2008 + . http://hdl.handle.net/2429/37904.

Kapur, M. (2011). A further study of productive failure in mathematical problem solving: unpacking the design components. *Instructional Science, 39*(4), 561–579. doi:10.1007/s11251-010-9144-3.

Kapur, M. (2012). Productive failure in learning the concept of variance. *Instructional Science, 40*(4), 651–672. doi:10.1007/s11251-012-9209-6.

Kapur, M., & Bielaczyc, K. (2011). Classroom-based experiments in productive failure. In *Proceedings of the 33rd annual conference of the cognitive science society* (pp. 2812–2817).

Kapur, M., & Bielaczyc, K. (2012). Designing for productive failure. *Journal of the Learning Sciences, 21*(1), 45–83. doi:10.1080/10508406.2011.591717.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: an analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*(2), 75–86. doi:10.1207/s15326985ep4102_1.

Koedinger, K. R., & Aleven, V. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review, 19*(3), 239–264. doi:10.1007/s10648-007-9049-0.

Loibl, K., & Rummel, N. (2013). The impact of guidance during problem-solving prior to instruction on students' inventions and learning outcomes. *Instructional Science, 1*, 22. doi:10.1007/s11251-013-9282-5.

Manlove, S., Lazonder, A. W., & de Jong, T. (2007). Software scaffolds to promote regulation during scientific inquiry learning. *Metacognition and Learning, 2*(2–3), 141–155. doi:10.1007/s11409-007-9012-y.

Matlen, B. J., & Klahr, D. (2012). Sequential effects of high and low instructional guidance on children's acquisition of experimentation skills: Is it all in the timing? *Instructional Science, 41*(3), 621–634. doi:10.1007/s11251-012-9248-z.

Rittle-Johnson, B., & Star, J. R. (2009). Compared with what? The effects of different comparisons on conceptual knowledge and procedural flexibility for equation solving. *Journal of Educational Psychology, 101*(3), 529–544. doi:10.1037/a0014224.

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning taking memory tests improves long-term retention. *Psychological Science, 17*(3), 249–255. doi:10.1111/j.1467-9280.2006.01693.x.

Roll, I., Aleven, V., & Koedinger, K. R. (2009). Helping students know 'further'-increasing the flexibility of students' knowledge using symbolic invention tasks. In *Proceedings of the 31st annual conference of the cognitive science society* (pp. 1169–74).

Roll, I., Aleven, V., & Koedinger, K. R. (2010). The invention lab: Using a hybrid of model tracing and constraint- based modeling to offer intelligent support in inquiry environments. In V. Aleven, J. Kay, & J. Mostow (Eds.), *Proceedings of the 10th international conference on intelligent tutoring systems* (pp. 115–24). Berlin: Springer.

Roll, I., Aleven, V., & Koedinger, K. R. (2011a). Outcomes and mechanisms of transfer in Invention activities. In *Proceedings of the 33rd annual conference of the cognitive science society* (p. 2824–2829).

Roll, I., Aleven, V., McLaren, B. M., & Koedinger, K. R. (2011b). Improving students' help-seeking skills using metacognitive feedback in an intelligent tutoring system. *Learning and Instruction, 21*(2), 267–280. doi:10.1016/j.learninstruc.2010.07.004.

Roll, I., Holmes, N., Day, J., & Bonn, D. (2012). Evaluating metacognitive scaffolding in guided invention activities. *Instructional Science, 40*(4), 691–710. doi:10.1007/s11251-012-9208-7.

Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. *Cognition and Instruction, 16*(4), 475–5223. doi:10.1207/s1532690xci1604_4.

Schwartz, D. L., Chase, C. C., Oppezzo, M. A., & Chin, D. B. (2011). Practicing versus inventing with contrasting cases: The effects of telling first on learning and transfer. *Journal of Educational Psychology, 103*(4), 759–775. doi:10.1037/a0025140.

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. doi:10.1207/s1532690xci2202_1.

Schwartz, D. L., Sears, D., & Chang, J. (2007). Reconsidering prior knowledge. In M. C. Lovett & P. Shah (Eds.), *Thinking with data* (pp. 319–344). New York: Routledge.

Siegler, R. S. (2002). *Microgenetic studies of self-explanation* (pp. 31–58). Microdevelopment: Transition processes in development and learning.

Strand-Cary, M., & Klahr, D. (2008). Developing elementary science skills: Instructional effectiveness and path independence. *Cognitive Development, 23*(4), 488–511. doi:10.1016/j.cogdev.2008.09.005.

VanLehn, K. (1988). Toward a theory of impasse-driven learning. In D. H. Mandl & D. A. Lesgold (Eds.), *Learning issues for intelligent tutoring systems* (pp. 19–41). New York: Springer. http://link.springer.com/chapter/10.1007/978-1-4684-6350-7_2.

Westermann, K., & Rummel, N. (2012). Delaying instruction: Evidence from a study in a university relearning setting. *Instructional Science, 40*(4), 673–689. doi:10.1007/s11251-012-9207-8.

Wiedmann, M., Leach, R. C., Rummel, N., & Wiley, J. (2012). Does group composition affect learning by invention? *Instructional Science, 40*(4), 711–730. doi:10.1007/s11251-012-9204-y.

Wise, A. F., & O'Neill, K. (2009). Beyond more versus less: A reframing of the debate on instructional guidance. In T. Duffy & Tobias (Eds.), *Constructivist instruction: success or failure*.