The Design and Evaluation of Optimal Computerized Guidance for Invention Activities: The Invention Coach

Catherine C. Chase, Teachers College, Columbia University, chase@tc.columbia.edu
Helena Connolly, Teachers College, Columbia University, connolly2@tc.columbia.edu
Marianna Lamnina, Teachers College, Columbia University, lamnina@tc.columbia.edu
Vincent Aleven, Human-Computer Interaction Institute, Carnegie Mellon University, aleven@cs.cmu.edu

Abstract: A common dilemma in educational technology is designing the optimal level and type of guidance to support open-ended learning activities. We explored this question by designing a computer-based coach for Invention activities – a form of ill-structured problem-solving followed by expository instruction. The Coach was designed to elicit the core learning mechanisms of Invention by problematizing students' solutions and mimicking naturalistic teacher guidance. This research tests both the efficacy of our designed guidance and the appropriate amount of guidance for Invention activities. In an experimental study, 205 middle schoolers worked with full, minimal, or no guidance versions of the Coach before receiving a lecture on the target knowledge (ratio structures in science). Students who received the full guidance Coach were better able to transfer their knowledge to novel domains. The work has implications for the design of guidance in open-ended learning environments.

Introduction

A key question in the design of technology-supported open-ended learning environments, is how to effectively guide students as they engage in complex, exploratory, ill-structured, and inquiry-focused activities. Many approaches have been tried (Quintana et al., 2004; de Jong & van Joolingen, 1998). A critical challenge is to provide guidance that encourages the learner to generate and construct her own ideas, without quelling the exploratory nature of the task or robbing students of essential discovery moments (Mavrikis, Gutierrez-Santos, Geraniou, & Noss, 2013; Koedinger & Aleven, 2007). We explored this question by designing and evaluating a computer-based coach to guide students through Invention activities.

Invention is an instructional method that combines ill-structured problem-solving with subsequent instruction (Schwartz & Martin, 2004). During Invention, learners attempt to invent the deep principles of a domain. In the current research, students were inventing ratio-based equations for physical science concepts (e.g. density=mass/volume, speed = distance/time). The goal of Invention activities is to prepare students to learn from later instruction. Prior work suggests that Invention creates "a time for telling," preparing students to appreciate the "mathematical work" of equations (Schwartz & Martin, 2004). Many studies have shown that Invention and related pedagogies boost conceptual learning and transfer to novel situations (Kapur, 2008; Schwartz & Bransford, 1998; Schwartz & Martin, 2004; Schwartz, Chase, Oppezzo, & Chin, 2011).

Of course, scaling Invention activities is difficult because students require individual guidance as they invent. Classroom studies of Invention often involve a ratio of 1 instructor to every 5 students (Schwartz et al., 2011), which is not practical for widespread adoption. Thus, we began designing and implementing a computerized Invention Coach that would provide optimal, individualized guidance (Marks, Bernett, & Chase, 2016). The Invention Coach was built using Cognitive Tutor Authoring Tools (Aleven et al., 2016), which are used to build Intelligent Tutoring Systems (ITSs). However, the Invention Coach differs from typical ITSs, which tend to provide structuring scaffolds, such as correctness feedback on problem steps and next-step hints. Instead, the Invention Coach scaffolds provide less structuring and more *problematizing* guidance. Problematizing scaffolds highlight a facet of the student's work that is problematic, encourage students to grapple with deep ideas, and contradict students' erroneous solutions (Reiser, 2004). Our Coach also differs from the Invention Lab (Roll, Aleven, & Koedinger, 2010), which favors a fairly structured approach to Invention activities. The project is part of an emerging line of research in ITSs that focuses on creating adaptive support for learners in open-ended learning environments for inquiry learning (Gobert, Sao Pedro, Raziuddin, & Baker, 2013; Poitras & Lajoie, 2014), exploratory learning (Mavrikis et al., 2013), and learning with simulations (Borek, McLaren, Karabinos, & Yaron, 2009).

In the literature on Invention and productive failure (a related pedagogy), the level of optimal guidance is disputed. For instance, Loibl & Rummel (2014a) found that guidance in the form of contrasting cases had no effect on learning outcomes. Kapur (2011) found that intermittent teacher support and brief benchmark lessons during the Invention process hindered learning compared to a no-guidance condition. In contrast, Holmes et al. (2014) found that computer-based guidance in the form of orienting and reflection prompts led to greater

learning than unguided Invention. However, results across these studies may differ either because the type of guidance varies with each experiment or because the guidance focuses on a single learning mechanism, when many learning processes are at play in effective Invention activities. Because our main goal was to build an effective system that would preserve the generative and exploratory style of Invention, we chose to provide an indirect, *problematizing* style of guidance that would support multiple learning mechanisms while providing naturalistic, human-like guidance. In this paper, we describe the process and rationale behind the Invention Coach design. We then report on a study that tests the optimal level of guidance for Invention and evaluates the effectiveness of the Coach's particular style of guidance in promoting learning and transfer.

An example invention activity

Figure 1A shows an example Invention activity. The goal of the task is to invent an index of "clown crowdedness" that describes how crowded the clowns are in each bus. Students are given a few constraints that are necessary for solving the problem: buses from the same company are equally crowded, a bigger index number means a bus is more crowded, the method for finding the index should be the same for all buses, etc. Though they don't know it, students are inventing the formula for density (density = mass/volume), where density is conceived as a measure of how crowded clowns are in different-sized buses (e.g., #clowns/#box cars).

Invention often occurs with the aid of *contrasting cases* that highlight important features of a problem solution while simultaneously illustrating the invariant structure across all cases (Bransford, Franks, Vye, & Sherwood, 1989). Many students begin the clown crowdedness task with a simple "counting" solution, where the number of clowns in each bus represents crowdedness, but they overlook the feature of space (or bus size). However, by contrasting the 3-compartment Crazy Clowns bus to the 6-compartment Clowns 'r' Us bus in Figure 1A, students often come to realize that bus size is a critical feature of crowdedness. Both buses have the same number of clowns but the Crazy Clowns bus is clearly more crowded. This contrast highlights the significance of the number of bus cars, which should give students the idea that their index must account for space. By looking across the cases, students may induce the invariant ratio structure (clowns to bus cars) that is common to all of them. Most students begin the task with an intuitive (but vague) notion of crowdedness which gets further differentiated and developed as they attempt multiple Inventions.

After students finish an Invention activity, they receive some other form of instruction, often a lecture or reading on the canonical solutions and related deep structures. Many students do not generate the correct equation during the initial Invention phase, but attempting to do so helps them notice deep domain features and explore how they may relate in a mathematical structure. This exploration prepares students to gain a deep and flexible understanding of the target knowledge presented in future instruction.

Our design process

To design the optimal type and amount of guidance for Invention tasks, we took a three-pronged approach. First, we studied teachers' naturalistic guidance of Invention to explore how teachers walk the line between giving and withholding assistance. Second, we implemented *problematizing* scaffolds, which make student solutions "problematic" without providing direct or corrective feedback. Third, we focused on the core cognitive processes invoked by successful Invention tasks. Finally, we created two prototype versions of the software which were improved based on extensive pilot testing.

Study of human teacher guidance and problematizing guidance

Since computer-based guidance can sometimes feel unnatural or lack the sophistication of human teaching tactics (du Boulay & Luckin, 2001), we modeled our system on human teacher guidance. To do this, we ran a study of experienced science teachers guiding students one-on-one through paper-based Invention activities (Chase, Marks, Bernett, Bradley, & Aleven, 2015). We asked teachers to guided students naturally, as they saw fit. Gains from pretest to posttest showed that the teachers were quite successful in increasing students' conceptual knowledge (effect size $d = 0.6$) and ability to transfer to novel domains ($d = 0.7$). Overall, we found that teachers used an "ask more, tell less style" style of dialogue. Teachers asked questions twice as often as they gave explanations, and they rarely gave direct right/wrong feedback. Moreover, the more teachers posed deep questions and withheld explanations, the more students transferred their learnings to novel problems.

Given these findings, we designed the Invention Coach with an "ask more, tell less" style of guidance, with a focus on deep questions. As such, the Coach avoids giving didactic explanations, instead prompting students to reflect on and explain their answers. We also drew on Reiser's (2004) construct of *problematizing* scaffolds. In many educational technologies, scaffolds serve to structure, simplify, or ease the task in some way. Problematizing scaffolds, on the other hand, add complexity to the task in the short term, by making students confront and grapple with deep disciplinary ideas. The Invention Coach problematizes student understanding by

contradicting and poking holes in wrong solutions, and encouraging students to diagnose their own errors. Thus, the Coach avoids giving direct right/wrong feedback or explicitly stating students' errors or how to fix them.

Supporting learning mechanisms

Our study of human teacher guidance of Invention gave us a feel for the types of prompts and dialogue to provide in the Coach, but we felt the Coach would be most effective if it also supported the key learning mechanisms of Invention (Loibl et al., 2016). There are three core learning processes invoked by Invention activities. The first is activation of prior knowledge. By asking learners to Invent their own solutions before telling them the expert solutions, we invite learners to draw out prior knowledge and skills that can then be augmented, built upon, or modified (Kapur, 2008). The second core learning process in Invention is uncovering knowledge gaps. While attempting to invent solutions and often failing, students come to realize their solution is insufficient and may identify holes in their knowledge, which they can then seek to fill during later instruction (Loibl & Rummel, 2014b). The third core learning mechanism is noticing deep features. Contrasting cases can highlight deep features of the target knowledge, preparing students to learn (from later instruction) how these features relate in a mathematical structure (Bransford et al., 1989; Schwartz et al., 2011). To maximize the effectiveness of our Invention Coach, we designed instructional modules that would support each of these three core learning mechanisms.

The Invention Coach

Drawing on our study of human teacher guidance, the problematizing framework, and the core learning mechanisms identified in the literature, we designed and implemented the Invention Coach system. Figure 1 depicts our third iteration of the Coach with an example of the crowded clowns Invention activity. In this activity students are asked to invent a numerical index to describe how crowded the clowns are in each bus. Students input their invented index numbers (B) next to each contrasting case (A). The Coach (C) provides hints and guidance along the way in the dialogue box (D). If students get stuck, they can access several resources such as a calculator, rules sheet (which describes task goals and constraints), a notepad, and a "help" button to solicit guidance (E). Students tend to invent iteratively by generating solutions, receiving guidance, then generating new solutions in Invention-guidance cycles. While the Coach never explicitly tells students whether their Inventions are right or wrong, it gives indirect feedback by posing comments, questions, or activities that keep students reflecting on and evaluating their inventions.
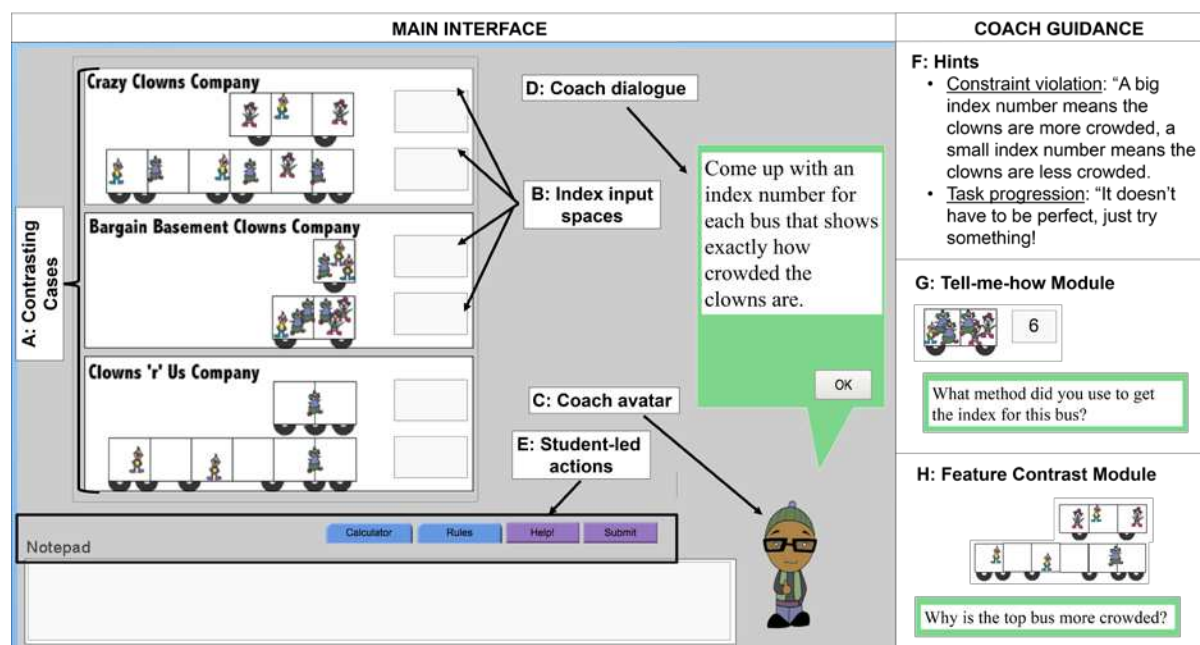

Figure 1. Invention Coach main interface and forms of guidance.

The Invention Coach guidance comes in several forms. Any time students "submit" their solutions to the Coach, they receive a short motivational message. We included motivational messages because we found in our pilot work that students often felt frustrated when they could not generate the right solution immediately

(since typical school math and science problems are not solved iteratively). The motivational message is followed by either a hint or a module. Most hints remind students of problem constraints their solutions violate, while others encourage students to progress through the task (Figure 1F). Modules are longer, interactive sequences where students respond to prompts and complete activities designed to engage the learning mechanisms discussed above.

There were three main modules: ranking, tell-me-how, and feature-contrast. The *ranking module* was designed to help students activate their intuitive, prior knowledge of crowdedness (Kapur, 2008). In this module, learners are asked to order the companies from most to least crowded. Most students can visually distinguish between the most and least crowded bus companies (though they often do not know how to quantify crowdedness yet). This ranking activity can be a good form of early guidance for students who are initially lost. Later on in the task, students can evaluate their index numbers by comparing their intuitive ranking of the cases to the ranking provided by their indices. Some form of this ranking activity occurred fairly frequently in our study of human teacher guidance, as well. The *tell-me-how module* asks learners to reflect on and explain how they generated their index numbers (Figure 1G). It also mimics a question that was frequently posed by the teachers in our study of naturalistic guidance: "Tell me how you got that number." In the tell-me-how module, students first describe their solution method in an open text box before selecting from a set of explicit strategy types derived from our piloting ("I counted", "I estimated", or "I calculated"). They are then pushed to describe their answers in disciplinary terms, such as mathematical expressions and units. In essence, the tell-me-how module elicits students' mathematical self-explanations. Self-explaining is one way to surface gaps in knowledge, a key learning process in Invention (Chi, De Leeuw, Chiu, & LaVancher, 1994; Loibl & Rummel, 2014b; Roll et al., 2010). Finally, the *feature-contrast module* (Fig. 1G), encourages students to compare specific sets of contrasting cases to help them notice key features of the domain, another key learning mechanism of Invention (Roll et al., 2010; Schwartz et al., 2011). In the example in Figure 1H, the Coach compares two cases and asks the student why the top bus is more crowded. The comparison is designed to highlight the often-overlooked feature of bus size, as only bus size differs across the cases.

Both modules and hints are designed to enact the "ask more, tell less" style, while *problematizing* students' solutions. The Coach avoids giving didactic explanations and never gives direct feedback on students' Inventions. Rather, it points out how solutions are problematic by noting a constraint the solution violates (in hints), encouraging articulation of ideas (tell-me-how module), helping students notice a feature missing in their solution (feature contrast module), or encouraging comparison of intuitive notions of crowdedness to invented solutions (ranking module). Moreover, the majority of hints and within-module prompts are structured as deep questions, asking students to generate, explain, and connect their ideas, modeled on the teacher guidance observed in our study.

How does the Coach decide when to give which forms of guidance? Based on data from a study with a Wizard-of-Oz version of our system (where guidance was selected by a human "Wizard"), we constructed a novel model of an adaptive coaching strategy to determine when to give each kind of guidance (Aleven et al., 2017). Every time a student submits their answer, the system classifies their solution into one of five broad categories: unclassifiable, single-feature, two-feature, mathematical two-feature, and ratio. The system responds with guidance related to the most significant knowledge gap expressed in the solution. For instance, if the student's solution only considers a single feature (e.g., number of clowns), then guidance will focus on getting the student to notice the importance of space. If the student's solution considers both features but does not relate them mathematically, then the guidance focuses on the importance of precision (e.g., index numbers should be exact). Each solution category has a sequence of alternating hints and modules. When students submit a solution of the same type, the system provides the next piece of guidance in the sequence. Students often generate a number of different solution types in a session, jumping from one branch of guidance to another throughout.

The current study

The current experiment was designed to address two questions: (1) Can a computerized Coach that provokes key learning mechanisms via problematizing guidance effectively support learning and transfer from Invention?, and (2) How does the level of guidance during Invention (full, minimal, or none) impact learning and transfer? To answer these questions, we compared our full version of the Invention Coach system to "minimal guidance" and "no guidance" versions of the same system. This allowed us to test the efficacy of our designed guidance, while controlling for use of the computer, task structure, presence of the Coach avatar, and so on. The "full guidance" condition had access to our complete Invention Coach with hints and modules as guidance. The "minimal guidance" condition received only hint-based guidance, but no modules. This amounted to giving students guidance about the goals and constraints of the task. Thus, the "minimal guidance" condition served as a useful comparison condition when testing the Coach's efficacy, because it does not provide additional

information that go beyond the given task goals and constraints (it merely reiterates them in different words and gives them adaptively, in response to a student's specific solution). Finally, the "no guidance" condition received only repeated suggestions to keep working (e.g. "Keep going, you're not quite there yet."). Since the "full guidance" condition would have the benefit of our modules, which were designed to provoke the key learning mechanisms of Invention, we hypothesized that they would perform best on learning and transfer measures. Further, we hypothesized that the "minimal guidance" condition's adaptive hints that problematize students' solutions by pointing to constraints their solutions violate would produce a middle level of learning and transfer. In contrast, we predicted that the "no guidance" condition, which would have no hints or modules to help them understand why their solutions failed, would have the lowest learning and transfer outcomes. Recall that Invention activities are followed by expository instruction, so we really tested how well these various levels of guidance prepared students to learn and transfer from a later lecture.

Methods

Participants
Students from 9 seventh- and eighth-grade classes (N = 205) in a public middle school in New Jersey participated in this study over a total of five days during their regular science classes. The school population was 96% Hispanic and 56% male, with 87% receiving free and reduced-price lunch. Condition was randomized at the student level, so students within the same class were assigned to different conditions.

Design and procedure
Two weeks prior to the start of instruction, students took a pretest to assess their prior knowledge of ratio structures as they relate to density and speed concepts. Instruction lasted for three days, 35 minutes per day. On the first day, students worked with the Invention Coach on the clown crowdedness Invention activity (see Figure 1A). On the second day, students worked with the Invention Coach on a car fastness Invention activity, where their goal was to develop an index of car fastness (developing the equation for speed). Invention activities were introduced by a short video explaining task goals and constraints. Students worked individually on each Invention task. On the third day of instruction, all students received a PowerPoint lecture from an experimenter that (1) gave the scientifically accurate solution to the Invention activities (2) related these activities to the science concepts of density and speed and (3) highlighted the importance of ratio structures in physical science equations. The lecture was integrated with a set of word problems which students completed on a paper worksheet, where they practiced applying these equations in simple, well-defined problem scenarios, similar to those that would be found at the end of a textbook. The goal of this instruction was to enhance students' understanding and ability to notice ratio structures in science. The day after this, all students completed a learning and transfer posttest.

Conditions were implemented using the full, minimal, and no guidance versions of the Invention Coach system described earlier. Each time students "submit" their indices, they received some form of guidance, which varied by condition. Additionally, all three conditions received motivational messages (e.g. "I can see the gears turning in your brain, you're working hard!"). Also, all students had access to student-initiated tools such as a calculator, a notepad, and rules list. Both full and minimally guided versions of the system used a similar adaptive coaching strategy to select the appropriate guidance in response to a student's solution type.

Measures
Log data from all interactions in the Coach were collected. To get a feel for the Invention problem-solving process, for each student, we calculated the number of different sets of indices submitted to the Coach, the number of unique solution types submitted (according to the five categories to which the system responds), and whether or not the student was ever able to invent a ratio-based solution.

Paper test items targeted students' understanding of ratio structures in physical science concepts. The posttest contained 8 items: 3 conceptual, 2 application, and 3 transfer items. Conceptual items required students to reason about the ratio structure of density and speed (e.g. determining whether a large or small pillow is more "tightly packed" if the number of feathers is constant). Application items asked students to reason about density and speed ratios in novel ways (e.g. design a flower pot that has a specific density of flowers). Transfer questions assessed whether students could notice and implement ratio structures in novel domains (e.g. describe the "spray strength" [i.e. pressure=force/area] of several fountains). The pretest contained a total of 6 items consisting of a limited set of 4 isomorphic versions of the posttest items (counterbalanced), plus 2 items to test students' prior knowledge of the density and speed equations.

Results

While working in the Invention Coach software, students submitted an average of 17 sets of indices per problem to the Coach, which did not vary significantly by condition, $F(2, 199) = 1.07$, $p = .35$. During this iterative process, students in the full guidance condition invented significantly fewer solution types ($M = 1.7$, $SE = .11$) than in the minimal ($M = 2.1$, $SE = .11$), $p = .007$ and unguided conditions ($M = 2.2$, $SE = .11$), $p = .001$. However, the number of students who invented a ratio solution on either of the Invention activities did not differ by condition (full = 49%; minimal = 59%, none = 65%), $\chi^2(2) = 3.64$, $p = .16$.

Conditions did not differ significantly on pretest scores, $F(2, 202) = .05$, $p = .96$. To analyze how condition affected posttest measures, we conducted a MANCOVA, with the three posttest measures (conceptual understanding, application, and transfer) as dependent variables, and condition, class, and whether students invented a ratio solution as independent variables, and pretest score as covariate. All variables had significant main effects on the dependent variables (p's < .02). Interactions were not significant, so they were excluded from our MANCOVA model. The omnibus MANCOVA revealed a significant main effect of condition, $F(6, 382) = 2.50$, $p = .02$, $\eta_p^2 = .04$. Follow-up ANOVAs revealed that all three groups performed similarly on conceptual and application items, p's > .14, but differed significantly on transfer items, $F(2, 192) = 5.04$, $p = .007$. As shown in Table 1, the descriptive pattern on posttest transfer items is that the full guidance condition performed the best, followed by the minimal guidance condition, and then the no guidance condition. However, posthoc tests (with Bonferonni correction) on posttest transfer scores revealed that only the full guidance condition performed significantly better than the no guidance condition, $p = .005$, however the minimal guidance condition did not differ significantly from either condition.

Table 1. Adjusted means and standard errors for each item type on posttest. Max scores = 1

Item Type	Full Guidance		Minimal Guidance		No Guidance	
	M	SE	M	SE	M	SE
conceptual	0.71	0.02	0.68	0.02	0.68	0.03
application	0.49	0.04	0.56	0.04	0.44	0.04
transfer**	0.55	0.03	0.50	0.03	0.43	0.03

To compare the efficacy of our full Invention Coach to the human teachers in our previous study (Chase et al., 2015), we explored gains on items that were common on pre and posttests. While there were no significant gains on conceptual items, the full guidance group made significant gains on the common transfer item, effect size $d = 0.6$. Thus, the full guidance provided by the Invention Coach was almost as effective as the human teachers used in our previous study in enhancing transfer ($d = 0.7$), but not conceptual learning ($d = 0.6$). While this is not the most rigorous comparison given differences in the participant populations and time spans across these two studies, it adds evidence to the conclusion that our system is effective.

Discussion

We have detailed the design of the Invention Coach, which was based on a study of human teacher guidance of Invention along with prior research and theory on the cognitive processes that make Invention effective at enhancing transfer. Our full guidance version of the Coach has an adaptive, "ask more, tell less" style of guidance. It attempts to problematize student solutions by contradicting and poking holes in them. It also contains activities (modules) that encourage students to activate prior knowledge, identify gaps in their understanding, and notice deep features of the problem. It implements a style of guidance that is distinctly different from that found in traditional ITSs, which tend to focus on providing step-level guidance for practice in solving complex problems (e.g., Aleven et al., 2016; Kulik & Fletcher, 2015; Vanlehn, 2011).

This work provides initial evidence of the full Invention Coach's efficacy in enhancing transfer. Our study showed that only the full guidance version was more effective than the no guidance condition at inducing transfer on the posttest. This research also provides evidence for the optimal level and type of guidance for scaffolding students through Invention activities, when the goal is transfer. Our results indicate that in computer-based learning environments, scaffolding Invention tasks can be effective, when the guidance substantially problematizes students' solution attempts and encourages core learning mechanisms.

This study also demonstrated that extensive guidance provided by the full guidance version of the Coach did not hinder student exploration. While students in the full guidance condition created fewer solution

types, they submitted the same number of solutions as the other conditions, and students in the full guidance were just as likely to generate the correct solution as students who received minimal or no guidance.

However, we also found that the full guidance Coach was not effective at enhancing conceptual learning or the ability to apply and manipulate learned ratios. There are two explanations for this outcome. One is that students may have learned these concepts and applicative abilities largely from the lecture and practice that followed – which were identical for all three conditions. Unfortunately, we cannot present hard evidence of this, since the worksheets students completed during the lecture were not done independently. A second possibility is that Invention pedagogies are uniquely designed to enhance transfer, but not learning, as has been found in other work (Schwartz & Martin, 2004). Thus, guidance designed to maximize the benefits of Invention may improve transfer only.

A limitation of this work is that the current Coach is only equipped to support Invention of ratio-based equations. In future work, we aim to adapt the Coach to support Invention of a broader variety of equation types (additive, multiplicative, exponential, etc.). A second limitation is that we have not isolated exactly what makes the Invention Coach effective in comparison to the "no guidance" condition. Future work could test the efficacy of problematizing guidance and the addition of guidance modules separately.

Despite limitations, this research makes both practical and theoretical contributions. The work adds to the small body of evidence regarding the question of how much guidance might be optimal during Invention activities. The results seem consistent with that of Holmes et al. (2014), who found that a modest amount of scaffolding led to better conceptual learning outcomes. Our results diverge from Loibl and Rummel (2014a) and Kapur (2011), who found that guidance during early instruction was either detrimental or simply not helpful. In the case of Loibl & Rummel (2014a), the lack of effect may be due to their focus on a singular learning mechanism, while guidance may need to address the full suite of learning processes to be effective. In contrast, our guidance (as well as that of Holmes et al., 2014) taps into additional, related learning processes such as identifying knowledge gaps and activating prior knowledge. In the case of Kapur (2011), it is possible that the guidance, some of which contained mini lectures, provided too much structure, dampening the exploratory nature of the task (Chase et al., 2015). Thus, future research could investigate not just *how much* guidance is optimal in Invention, but also *what kind* of guidance produces the greatest learning gains.

More broadly, we have created the first technology designed to support Invention with adaptive guidance, modeled on human teacher guidance. This adds to the body of work on the efficacy of various forms of technology-based scaffolds for open-ended learning tasks (Quintana et al., 2004; de Jong & van Joolingen, 1998; Reiser, 2004). Moreover, this paper provides a blueprint for designing software to guide open-ended problem-solving with deep questions, few explanations, problematizing guidance, and modules that invoke specific learning mechanisms.

References

Aleven, V., Connolly, H., Popescu, O., Marks, J., Lamnina, M., & Chase, C. (2017, June). An Adaptive Coach for Invention Activities. In *International Conference on Artificial Intelligence in Education* (pp. 3-14). Springer, Cham.

Aleven, V., McLaren, B. M., Sewall, J., van Velsen, M., Popescu, O., Demi, S., . . . Koedinger, K. R. (2016). Example-Tracing tutors: Intelligent tutor development for non-programmers. *International Journal of Artificial Intelligence in Education*, *26*(1), 224-269. doi:10.1007/s40593-015-0088-2

Borek, A., McLaren, B. M., Karabinos, M., & Yaron, D. (2009). How much assistance is helpful to students in discovery learning? In U. Cress, V. Dimitrova, & M. Specht (Eds.), *Proceedings 4th European Conference on Technology Enhanced Learning, EC-TEL 2009* (pp. 391-404). Springer Berlin Heidelberg. doi:10.1007/978-3-642-04636-0_38

Bransford, J. D., Franks, J. J., Vye, N. J., & Sherwood, R. D. (1989). New approaches to instruction: Because wisdom can't be told. *Similarity and analogical reasoning*, *470*, 497.

Chase, C. C., Marks, J., Bernett, D., Bradley, M., & Aleven, V. (2015, June). Towards the development of the invention coach: A naturalistic study of teacher guidance for an exploratory learning task. In *International Conference on Artificial Intelligence in Education* (pp. 558-561). Springer, Cham.

Chi, M. T., De Leeuw, N., Chiu, M. H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive science*, *18*(3), 439-477.

De Jong, T., & Van Joolingen, W. R. (1998). Scientific discovery learning with computer simulations of conceptual domains. *Review of educational research*, *68*(2), 179-201.

Du Boulay, B., & Luckin, R. (2001). Modelling human teaching tactics and strategies for tutoring systems. *International Journal of Artificial Intelligence in Education*, *12*(3), 235-256.

Gobert, J. D., Sao Pedro M., Raziuddin J., & Baker, R.S. (2013) From log files to assessment metrics: Measuring students' science inquiry skills using educational data mining. *Journal of the Learning Sciences 22*(4), 521-563.

Holmes, N. G., Day, J., Park, A. H., Bonn, D. A., & Roll, I. (2014). Making the failure more productive: scaffolding the Invention process to improve inquiry behaviors and outcomes in productive failure activities. *Instructional Science*, *42*(4).

Kapur, M. (2008). Productive failure. *Cognition and Instruction*, *26*(3), 379-425.

Kapur, M. (2011). A further study of productive failure in mathematical problem solving: Unpacking the design components. *Instructional Science*, *39*(4), 561-579.

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

Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of intelligent tutoring systems: a meta-analytic review. *Review of Educational Research*, *86*(1), 42-78.

Loibl, K., Roll, I., Rummel, N. (2016). Towards a Theory of When and How Problem Solving Followed by Instruction Supports Learning. *Educational Psychology Review.*

Loibl, K., Rummel, N. (2014a). The impact of guidance during problem-solving prior to instruction on students' Inventions and learning outcomes. *Instructional Science, 42*, 305-326.

Loibl, K., & Rummel, N. (2014b). Knowing what you don't know makes failure productive. *Learning and Instruction*, *34*, 74-85.

Marks, J., Bernett, D., & Chase, C. C. (2016). The Invention Coach: Integrating Data and Theory in the Design of an Exploratory Learning Environment. *International Journal of Designs for Learning*, *7*(2).

Mavrikis, M., Gutierrez-Santos, S., Geraniou, E., & Noss, R. (2013). Design requirements, student perception indicators and validation metrics for intelligent exploratory learning environments. *Personal and Ubiquitous Computing, 17*(8), 1605-1620.

Poitras, E. G., & Lajoie, S. P. (2014) Developing an agent-based adaptive system for scaffolding self-regulated inquiry learning in history education. *Educational Technology Research and Development 62*(3), 335-366.

Quintana, C., Reiser, B. J., Davis, E. A., Krajcik, J., Fretz, E., Duncan, R. G., ... & Soloway, E. (2004). A scaffolding design framework for software to support science inquiry. *The journal of the learning sciences*, *13*(3), 337-386.

Reiser, B. J. (2004). Scaffolding Complex Learning: The Mechanisms of Structuring and Problematizing Student Work. *Journal of the Learning Sciences*, *13*(3), 273–304.

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.), *Lecture Notes in Computer Science: Proceedings of the 10th International Conference on Intelligent Tutoring Systems, ITS 2010* (Vol.1, pp. 115-124). Berlin: Springer.

Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. *Cognition and instruction*, *16*(4), 475-5223.

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.

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.

Vanlehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, *46*(4), 197-221.

Acknowledgments