Instr Sci
DOI 10.1007/s11251-010-9151-4

Testing the instructional fit hypothesis: the case of self-explanation prompts

Timothy J. Nokes · Robert G. M. Hausmann · Kurt VanLehn · Sophia Gershman

Received: 11 May 2010 / Accepted: 29 September 2010

Abstract Cognitive science principles should have implications for the design of effective learning environments. The self-explanation principle was chosen for the current work because it has developed significantly over the last 20 years. Early formulations hypothesized that self-explanation facilitated inference generation to supply missing information about a concept or target skill, whereas later work hypothesized that self-explanation facilitated mental-model revision (Chi, Handbook of research on conceptual change, 2000). To better understand the complex relationship between prior knowledge, cognitive processing, and changes to a learner's representation, two classes of self-explanation prompts (gap-filling and mental-model revision) were tested in the domain of physics problem solving. Prompts designed to focus the learner on gap-filling led to greater learning and a reduction in the amount of tutoring assistance required to solve physics problems. The results are interpreted as support for the *instructional fit hypothesis*—the idea that the efficacy of instruction is contingent on the match between the cognitive processing that the instruction elicits, how those processes modify the underlying knowledge representations for the task, and the utility of those representations for the task or problem.

Keywords Self-explanation · Prompting · Worked examples · Intelligent tutoring systems

T. J. Nokes (✉)
Department of Psychology, Learning Research and Development Center,
University of Pittsburgh, 3939 O'Hara Street, Pittsburgh, PA 15260, USA
e-mail: nokes@pitt.edu

R. G. M. Hausmann
Carnegie Learning, Inc., Pittsburgh, PA, USA

K. VanLehn
School of Computing, Informatics and Decision Systems Engineering,
Arizona State University, Tempe, AZ, USA

S. Gershman
Watchung Hills Regional High School, Warren, NJ, USA

Published online: 11 November 2010

In many formal domains such as mathematics, physics, and logic, it is common for students to learn from both studying examples and solving problems. Instructors typically require students to solve problems, which often motivates them to use and study the examples presented in the textbook or lecture. Unfortunately, many students study the examples using shallow learning strategies such as paraphrasing (Hausmann and Chi 2002) or making analogies based on how the surface features match the target problem (Ross and Kilbane 1997; VanLehn 1998). If students rely on these strategies, then they may flounder when solving more complex problems or those that change in surface features (Aleven and Koedinger 2002; Ross 1984). In an effort to avoid such shallow processing, researchers have developed several kinds of prompts that facilitate *self-explanation*, an activity that promotes deep processing of examples and robust learning (Chi et al. 1994).

The purpose of the current study is twofold. The first goal is to conduct translational research that tests different theory-driven implementations of the self-explanation principle in an applied classroom setting. Much work on learning in the cognitive sciences has implications for educational practice; however, a large divide often separates the principles discovered in the laboratory from their potential implementation in the classroom. Implementing a learning principle requires more than simply translating a textual description of that principle for a particular problem or task. It relies upon understanding the principle and the factors that affect its implementation. For self-explanation, one needs to take into account the learner's prior knowledge, the target knowledge or skill to be acquired, and the structure of the domain, task, or activity. Our aim here is to begin to close this divide by taking a principled approach to testing different implementations of the self-explanation principle in a classroom setting.

Our second goal is to test the *instructional fit hypothesis*, which is the idea that the efficacy of prompting is contingent on the match between the cognitive processing that the prompting elicits, how those processes modify the underlying knowledge representations for the task, and the utility of those representations for understanding and solving the task or problem. We test the efficacy of two classes of self-explanation prompts: prompts designed to promote gap-filling inference processes and prompts designed to promote mental-model revision for students learning to solve electrodynamics problems in a high school physics class. We hypothesized that gap-filling prompts would be a better instructional fit than mental-model revision prompts because the students were not expected to have prior misconceptions for these problem types (Maloney et al. 2001). Through pursuing these goals, we hope to make progress in both applying cognitive research to real-world problems (i.e., classroom learning) as well as developing a better understanding of the basic learning mechanisms of the mind.

In the next section, we describe two classes of self-explanation prompts: those targeted for inference generation and those targeted for mental-model revision. In the sections that follow, we elaborate on the instructional fit hypothesis and describe our experiment to test it in the context of physics problem solving. In the last section, we describe the implications of the results for understanding the relation between basic learning processes, the learning environment, and the translation of cognitive principles into pedagogy.

Different types of self-explanation prompts

A worked example in the domain of physics consists of a series of problem-solving steps that terminate with the problem's solution. Worked examples demonstrate the application of domain principles and expert solution strategies. However, these examples are often

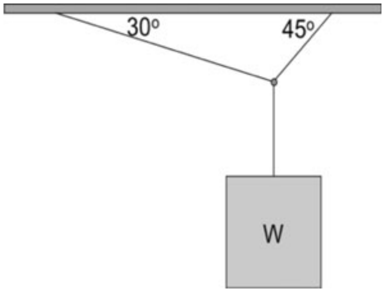

Fig. 1 A figure from a worked example

incomplete with respect to the conditions under which a step applies. For instance, consider the following steps from a statics example: "Fig. 1 shows an object of weight **W** hung by strings. Consider the knot at the junction of the three strings to be 'the body'." Unfortunately for the student, the example does not explicate a reason for choosing the knot as the body (i.e., it is a choice of problem-solving convenience because the sum of the forces at the knot is zero).

In an effort to facilitate deep processing of an example or expository text, a number of self-explanation interventions have been developed including: instructional prompts that ask the student to justify the correct steps (Berthold et al. 2009; Rittle-Johnson 2006), prompts that ask the student to engage in reflection (Chi et al. 1994), prompts that ask students to explain by referencing principles in a glossary (Aleven and Koedinger 2002), instructional explanations designed to elicit self-explanations (Renkl 2002), and training through a tutoring system that provides feedback and guidance for the development of self-explanations (McNamara 2004). In the current work, we focus on two classes of instructional prompts that correspond to two hypothesized mechanisms of self-explanation, specifically, gap-filling (Chi and Bassok 1989) and mental-model revision (Chi 2000).

Gap-filling

The first class of prompting we call "gap-filling" because it was designed around the hypothesis that self-explanation leads to learning gains because it prompts the necessary information missing in the examples. Chi and Bassok (1989) argued that good students generated self-explanation inferences to add coherence and completeness to examples that skipped steps or did not include the application conditions for the principle or concept being applied. By filling in the gaps, the student can focus on the links between those steps and the sub-goals of the problem (Catrambone 1998). If the student encodes the problem type and the sub-goals from the example, then this should facilitate schema acquisition. Problem-solving schemata typically consist of the representative features of the problem, including the problem type, associated concepts and principles and their application conditions, sub-goals, and procedures for solving the problem (Marshall 1995; Nokes et al. 2010). Gap-filling prompts are particularly well suited to facilitate the acquisition of a problem-solving schema. We focus on two types of gap-filling prompts used in the literature, justification and step-focused.

Justification prompts ask students to generate the principle justification for a step in an effort to focus her or his processing on the underlying concepts and application conditions for the step (see Table 1a for examples from physics). Critically, self-explaining each step facilitates inference generation connecting the underlying principles and concepts in the

Table 1 Examples of gap-filling and mental-model revision prompts

Gap-filling: prompts that focus the learner on generating inferences to fill in the information that is missing from the example
a. Gap-filling justification prompts (Conati and VanLehn 2000)
What principle is being applied on this step?
This choice is correct because…
What is the justification for this step? Why is it correct?
What law, definition, or rule allows one to draw that conclusion?
b. Gap-filling step-focused prompts (Hausmann and Chi 2002)
What does that step mean to you?
Do you have any more thoughts about that step?
Could you restate or summarize that step in your own words?
So, specifically, what else does this step tell us?
Mental-model revision: prompts that focus the learner on revising his or her prior knowledge
c. Mental-model revision prompts (Chi et al. 1994)
What new information does each step provide for you?
How does it relate to what you've already seen?
Does it give you a new insight into your understanding of how to solve the problems?
Does it raise a question in your mind?

domain to the equations used in the problem. Justification prompts have been used successfully to assist students in the early stages of learning. Conati and VanLehn (2000) designed an intelligent tutoring system to coach students while they studied examples. The system included prompts for self-explanation that focused on the justification for taking a problem-solving step. Students who were initially learning the material demonstrated learning gains with a large effect size ($d = 1.07$).

A second kind of gap-filling prompt is the *step-focused prompt* (Table 1b, Hausmann and Chi 2002). These prompts focus the learner on explaining the steps of an example without specific direction as to what aspects of the step to explain. Although these prompts are much less directive than the justification prompts—because they do not explicitly instruct the student to explain the conceptual justification or reason for a step—they still focus the student's attention on explaining the "meaning" of an individual step. Explaining the meaning of the step provides an opportunity for the learner to generate inferences to fill in gaps in her or his understanding for that step. These prompts have been used successfully in some problem-solving domains (Hausmann and Chi 2002). However, this type of prompt may lead some students to focus their attention on the equations and the superficial relations between the steps and not necessarily on the deeper conceptual aspects of the problems. If students engage in this more superficial type of processing, then we would expect it to lead to less robust learning and knowledge transfer than students given the justification prompts. We include them in the current study to examine the robustness of prompts that promote gap-filling. Do step-focused prompts lead to robust learning or do students need to be explicitly prompted to generate the justifications underlying those steps?

Mental-model revision

The second class of prompting we call "mental-model revision" because it is based on the hypothesis that learning occurs through revising a flawed mental model (see Table 1c for example prompts). Chi (2000) argued that when students read a line of text, they compare their interpretation of the text to their prior understanding. If they detect a discrepancy, then they generate a self-explanation that repairs their flawed representation. This view assumes that the student has a prior misconception or knowledge that is 'in conflict' with the to-be-learned concept (Chi 2008). Therefore, this class of prompts is aimed at highlighting discrepancies between students' prior knowledge and the to-be-acquired knowledge.

Mental-model prompts have been used successfully to facilitate learning from an expository text on the circulatory system (Chi et al. 1994). Students were shown to have misconceptions about this topic at pretest. For example, many students believed that the circulatory system functioned as a 'single-loop model' where blood is oxygenated in the heart, rather than in the lungs, then is distributed to the body where the oxygen is used, and later is returned back to the heart for oxygenation. Students who were given the mental-model revision prompts while reading the expository text acquired a more scientifically accurate mental model of the concepts than students who studied the text without prompting. Moreover, the students in the prompting condition demonstrated higher scores than the unprompted students on the difficult items of a posttest, which suggests that many of these students showed *conceptual change* by revising their prior misconception into a correct conception ($d = 1.14$). The same prompts have also been used to successfully facilitate learning in other, more procedural domains, such as physics problem solving ($d = .57$, Hausmann and VanLehn 2007).

Both gap-filling and mental-model prompts have been demonstrated to be effective methods for eliciting self-explanations while studying worked examples or reading expository texts. However, the two differ in the underlying theoretical assumptions as to whether the self-explanation primarily facilitates the generation of new knowledge (gap-filling) or revising one's prior knowledge (mental-model revision).

The instructional fit hypothesis

To evaluate these two theoretical positions, we posit the instructional fit hypothesis, which is the idea that the cognitive processes triggered by an instructional intervention must match, or fit, the learning scenario (i.e., learner's prior knowledge, the task structure, and the knowledge required to perform well on the task). In the case of self-explanation, we hypothesize that the gap-filling and mental-model revision prompts are best suited for particular types of learning scenarios that depend on the student's prior knowledge (e.g., whether or not they have a prior misconception), the relation between that knowledge and the task domain, and the structure of the task or learning activity. See Table 2 for an illustration of the instructional fit hypothesis for two 'high fit' self-explanation prompt scenarios.

For example, mental-model revision prompts are likely to facilitate deep learning when the student has a prior misconception about the target concept, whereas gap-filling prompts do not require such prior knowledge to be effective and may be less effective if the student has a robust misconception. Mental-model revision prompts are designed to focus the learner on generating inferences to integrate her or his prior knowledge with the newly acquired information from the target task. These prompts promote the detection of errors or discrepancies between one's prior knowledge and the target knowledge and provide an opportunity to engage in error-correction to revise that knowledge (Chi 2000). The prompts have the learner consider the implications of the current step for their prior knowledge.

Table 2 Illustration of the instructional fit hypothesis for two 'high fit' scenarios for self-explanation prompts, prior knowledge, the elicited cognitive processes, and target knowledge

	Prior knowledge	Prompt type	Cognitive processes elicited	Target knowledge
High fit scenario for gap-filling	Little/no prior knowledge of concepts or procedures	Gap-filling	Inference generation; addition of new knowledge	Problem-solving schemas with step-based procedures
High fit scenario for mental-model revision	Prior misconceptions; knowledge that is in conflict with target knowledge	Mental-model revision	Inference generation to repair and revise prior knowledge; error-correction processes	Accurate mental models consisting of interrelations between features and propositions

Prompts such as 'does the current step give a new insight into the task' or 'does it raise a question in your mind' assume that the student's prior knowledge conflicts in some way with the new information. Indeed, when there is a conflict between what one knows and the current situation, these prompts can act as a way to confront the misconception and provide an opportunity to resolve it (Chi et al. 1994). Students in this kind of learning scenario are hypothesized to engage in inference generation and error-correction in an attempt to reconcile their previous knowledge with their new knowledge. This should promote the revision of prior knowledge in light of the new information.

However, if the student did not have any prior misconceptions, then such prompts should be ineffective and result in a different kind of processing than was intended. The prompts should seem odd to the learner resulting in either simple responses, such as "no, it does not raise a question in my mind" or "no, it does not provide a new insight because I did not know how to solve it previously", or, alternatively "yes, it did provide a new insight because I did not know it before". These kinds of responses should highlight the limited utility of such prompts, and may encourage students to ignore such prompts in the future.

In contrast, gap-filling prompts do not assume any misconceived prior knowledge, but instead focus the learner on the to-be-understood step. In fact, students could have no knowledge about the step but still be able to effectively respond to the prompt by using the instructional resources to help them generate inferences and fill in the gaps in their understanding (e.g., using worked examples, textbooks, peers, etc.). Gap-filling prompts are designed to help the learner generate inferences that highlight why and when the problem-solving steps should be taken. By generating justifications for each step, a learner connects that step to the underlying concepts and principles (i.e., the reasons for that step). Furthermore, explaining each step and the links between them, should focus participants on the sub-goals of the problem and provide an opportunity to abstract that sub-goal structure and incorporate it into a problem-solving schema.

If a student had a prior misconception, and the goal was to revise that misconception, then gap-filling prompts would likely leave those misconceptions inert, or potentially strengthen them. In fact, participants could respond to such gap-filling prompts with the wrong or incorrect justification based on their misconception. Gap-filling prompts appear ideal for self-explaining step-based tasks (e.g., problem solving) whereas mental-model revision prompts appear better suited for contexts in which the learner has a known misconception. The two classes of prompts are best designed for certain types of prior knowledge and task structures (see Table 2).

We argue that aligning the cognitive processes, the resulting knowledge representations, and their hypothesized utility for the target task is critical for reaching the maximal effectiveness of particular instructional interventions. As an example of how these different types of interactions can play out in the laboratory, we take another look at the research reported above.

We can use the instructional fit hypothesis to reinterpret the findings reviewed in the previous section. Chi et al. (1994) reported large effect sizes for a study that used mental-model revision prompts to elicit self-explanation inferences when students learned about the circulatory system, whereas Hausmann and VanLehn (2007) found a much smaller effect size ($d = 1.14$ vs. .57) for the same prompts when used to encourage students to self-explain electrodynamics examples. The instructional fit hypothesis predicts a reduction in the effectiveness of mental-model revision prompts for the second study because, unlike the Chi et al. (1994) study, participants were hypothesized not to have conflicting knowledge at the start of the experiment. The students were new to the field of electrodynamics and were unlikely to hold any misconceptions about the specific topic (Maloney et al. 2001, p. S17).

In sum, this analysis suggests that mental-model revision prompts are best for tasks in which students enter with misconceptions and where the goal is to revise those misconceptions. In contrast, gap-filling prompts appear best suited for tasks in which the goal is for the students to acquire a problem-solving schema. Next, we describe a study to test this hypothesis in a classroom.

Current study

To test the instructional fit hypothesis we conducted an in vivo classroom experiment in which high school students were randomly assigned to one of three prompting conditions where they were given either gap-filling justification prompts, gap-filling step-focused prompts, or mental-model revision prompts when studying worked examples in electrodynamics. Students solved three electrodynamics problems in the Andes intelligent tutoring system, and the examples were interleaved between each problem in the form of a video in which an expert solved a problem in Andes.

The problems required that students draw free-body diagrams, define vector and scalar quantities, identify relevant equations, and recognize when the problem is solved. This set of problem-solving activities is best characterized as learning a cognitive skill (Singley and Anderson 1989). We hypothesize that the acquisition of a step-based, problem-solving schema should facilitate solving these types of problems. To maximize learning from worked examples in this domain, students need to engage in (at least) three types of cognitive processing. First, they need to attend to the type of problem that is being solved. Problem-solving schemata typically include information about the problem type that is associated with the particular procedures and equations used to solve the problem. Second, the students need to attend to the individual problem-solving steps illustrated in the example. That is, they need to learn the steps for solving a given problem and the order in which they are implemented. Third, they need to generate inferences regarding the justifications for each step. These justifications provide knowledge for how and when to apply the steps. They provide the application conditions that go beyond the surface features of the original examples and support transfer to new problems with different surface features. The extent to which students engage in these three types of processing should predict the robustness of their learning.

The instructional fit hypothesis predicts that different types of prompts will be differentially beneficial because of the match between the cognitive processes elicited and the learner's prior knowledge, the target knowledge, and the task constraints. First, the fit between the cognitive processes elicited by mental-model revision prompts and acquiring a cognitive skill is *low* because the skill is step-based and the students are not expected to have prior misconceptions that they can repair and revise. Second, the fit between gap-filling prompts and learning a cognitive skill is *high* because the problem-solving activities are commensurate with attending to steps and generating inferences (i.e., reasons for taking each problem solving step). In addition, we expect justification prompts to facilitate more robust learning than step-focused prompts, because justification prompts should facilitate inferences connecting the steps to the underlying concepts and application conditions. The step-focused prompts test how direct the prompts have to be in promoting inference generation. Will simply explaining the steps be enough to facilitate construction of a problem-solving schema?

In addition to examining impact of the prompts on example study and tutored problem solving, we also examine how those prompts impact learning particular knowledge components. A *knowledge component* is a hypothesis of the "pieces" of knowledge that are critical to correctly solving a problem. Knowledge components are specified by an expert analysis and can be verified with empirical data (see Anderson and Lebiere 1998 for in-depth illustrations). Some examples of knowledge components in this domain include knowing the *definition of the electric field*, which is measured by specifying the appropriate equation, $\mathbf{F} = q\mathbf{E}$, or correctly *drawing and defining the electric field vector*, which is measured by correctly specifying the magnitude, direction, region, source, and label of the vector.

If gap-filling prompts promote learning the justifications of the steps, then we should expect to see large learning gains for those knowledge components in which the justification is a critical part of the step, such as knowing the problem type or a definition of a core concept in the problem. In contrast, steps that have less conceptual content such as learning to draw a vector may not show similar amounts of improvement. Furthermore, a gap-filling process suggests that the knowledge gains will be selective to the missing aspects of the skill (i.e., the gaps), whereas a mental-model revision process suggests that multiple components of the model must be revised to facilitate conceptual change. Therefore, mental model prompting should result in learning gains across multiple knowledge components related to the mental model. In the current experiment, we select three knowledge components that are present across the three problem-solving tasks to examine these hypotheses. The knowledge components include *defining the electric field*, *drawing the electric field*, and *drawing the electric-force vectors*. We expect the gap-filling prompts to be particularly helpful when learning the principle *definition of the electric field*.

Method

Participants

Forty-seven students were recruited from three sections of a second-semester physics course taught at an eastern high school. Volunteers were given course credit for their participation. The experiment took place in one of the open class periods, which were approximately 90 min in duration.

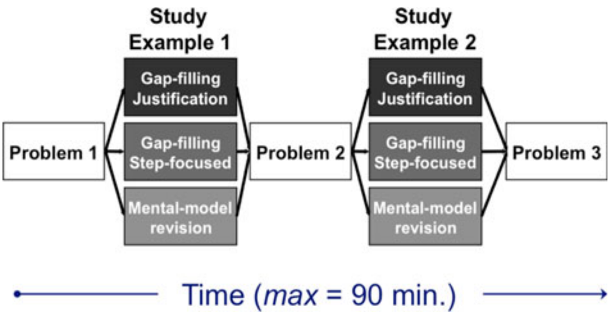

Fig. 2 Overview of experiment design and procedure

Design

The experiment was a mixed design with participants randomly assigned to one of three experimental conditions: justification ($n = 16$), step-focused ($n = 16$), and mental-model revision ($n = 15$). There were two within-subjects factors: examples and problems. Each student received three problem-solving tasks and two worked examples. See Fig. 2 for an overview of the design.

Materials

Three electrodynamics problems and two worked examples were designed in collaboration with the instructors of the course. Each example was isomorphic to the immediately preceding problem (see Appendix for test problems and examples). The students attempted to solve the problem first, with the assistance of an intelligent tutoring system, and then they studied the isomorphic example. The topics covered by the problems and examples included the definition of the electric field, the weight law, Newton's second law, and several kinematics equations. The problems grew in complexity such that they required the application of 16, 22, and 24 knowledge components, respectively. See Fig. 3 for a screenshot of problem 1.

The purpose of problem 1 was to provide a measure of initial problem-solving performance using the tutor. Because participants were randomly assigned to a condition, we did not expect participants to differ in their performance on this problem. We used students' scores on problem 1 as a covariate for prior knowledge and physics problem-solving ability when analyzing performance on problems 2 and 3. The expectation is that students will find the first problem difficult, thus motivating them to learn from the following worked example, which will impact subsequent problem solving and example study. However, the extent of student learning from examples is expected to vary as a function of the instructional fit of the self-explanation prompts.

Students solved the problems with the assistance of the Andes physics tutoring system, and the examples were presented as a video of the Andes screen, with an expert describing what actions were being taken. The reasons for each solution step were omitted from the examples because one of the goals of the experiment was to see if students were able to supply the missing information. Examples 1 and 2 had 10 steps each. The prompts were presented at the bottom of the screen, and the students were required to answer one prompt for each step of the example.

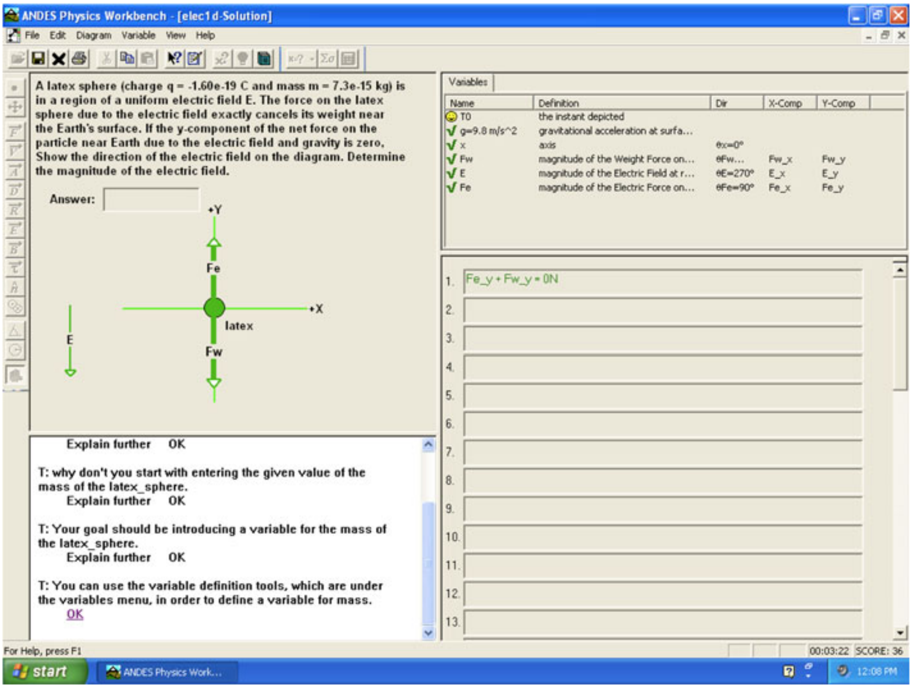

Fig. 3 Screenshot of problem 1 in Andes tutoring system

There were four prompts used for each condition (see Table 1). All of the prompts were used in previous research. The gap-filling justification prompts correspond to those used in Conati and VanLehn (2000) and were designed to elicit the generation of justifications for each step. The step-focused prompts were used by Hausmann and Chi (2002) and were constructed to focus the students' attention on each step, but they did not require students to generate justifications. The mental-model revision prompts were taken from a prior study that elicited self-explanations while reading an expository text (Chi et al. 1994) and were designed to facilitate revision of one's prior knowledge. Prompts were administered only during the example study.

All of the problem solving took place within the Andes physics tutor (VanLehn et al. 2005). Andes provides several problem-solving scaffolds. First, Andes gives color-coded, instant feedback on each step. If the student enters an incorrect entry, Andes will flag the attempted step red. Second, Andes provides the student with on-demand hints, which are graded in terms of their depth. The top-level hint directs the student's attention to a critical feature of the problem. For instance, in the statics problem from Fig. 1, the hint reminds the student that the body is near the earth, which should help the student to remember that gravity is acting on the body. The next level of hint teaches the student a concept or principle. Continuing the example, if pointing out the fact that the body is near the Earth is not enough, the next hint says, "When an object is near a planet, the planet exerts a weight force on the object." This is a verbal description of the applicability conditions for the weight law. If this is still not enough information to take the next step, the "bottom-out" hint directly tells the student what to do (e.g., draw a force on the body due to the earth of type weight). It is important to note that the information missing in the examples can be found in the on-demand hints.

Finally, Andes also provides a list of equations that the students can reference at any time. Students are required to work with symbolic expressions while solving a problem. Once the symbolic expressions are clearly written, Andes provides the students with the algebraic solution. That is, Andes handles all of the mathematical operations of substituting values into the variables and solves the system of equations.

Procedure

Students were recruited from three high school physics courses, and the instructors explained to them that they were going to be solving problems with Andes as part of their classroom exercises. On the day of the experiment, the students logged into Andes and read a short set of instructions explaining that they would be studying video-based examples. The prompts were shown to the students to orient them to what would be expected of them later. Then they solved the first problem with Andes. Afterwards, they studied the first example, which was broken down into 10 steps. At the conclusion of each step, the students were prompted to self-explain the step. Below the playback screen were the same four prompts that were displayed during the introduction. Talk-aloud instructions were given to the students, and the prompts were always available for the students to reference. The teacher instructed them to select and answer at least one prompt per step. Students explained each step aloud by talking into a headset microphone and their explanations were recorded.

After completing all of the example steps, the students then solved the next problem. This cycle of solving problems and studying examples continued until the students completed all of the materials, or until the class period ended. The order of the problems and examples were fixed for all of the students. Each student solved a problem first and then studied an isomorphic example (see Fig. 2 for an overview of the procedures).

Predictions

The gap-filling prompts were designed to help construct a problem-solving schema. By generating inferences explaining the meaning of each step, students in both the justification and step-focused conditions are expected to encode the problem type, the steps required to solve the problems, and the order of the steps. Furthermore, if they generate justifications for each step, they are expected to acquire the application conditions for those steps, as well as how the step relates to the underlying physics concepts and principles. Because students in both conditions are focused on explaining each step, we expect them to encode the sub-goal structures of the problems, providing them an opportunity to integrate that knowledge into an abstract problem schema.

Mental-model revision prompts were designed to facilitate the cognitive processes that repair a flawed mental model. If a student does not yet have a well-formed mental model, then such prompting should not be effective. Since physics novices are expected to have a nascent or incomplete problem-solving schema for these problem types, the justification and step-based prompts are expected to be helpful to these students. The justification prompts are expected to lead to the most robust learning (i.e., gap-filling and schema acquisition) followed by the step-focused prompts, because step-focused prompts are less explicit about generating inferences linking the steps to the underlying concepts. The mental-model revision prompts should show the least learning because they should be relatively ineffective in facilitating gap-filling processes. Therefore, we predicted that students would attempt to use the prompts to study the examples for equal amounts of time on the first exposure, not yet knowing the utility of such prompts. However, for the second example, we predicted that the students given justification and step-focused prompts would maintain a consistent level of engagement with the examples finding the utility of the prompts high, whereas the students given mental-model revision prompts would decrease their study of the example, finding the utility of the prompts low or unhelpful.

What happens while studying examples should have an impact on problem-solving performance. If students in the justification and step-focused conditions learn more from the examples, then they should require less pedagogical assistance while solving problems. Specifically, students in the mental-model revision condition are predicted to ask for more assistance and bottom-out hints from the tutor than the other two conditions when problem solving. We also explore the type of knowledge acquired by examining the learning curves of particular knowledge components required to solve the problems. If students were acquiring a mental model, then we would expect them to show consistent gains across different knowledge components because those knowledge components are interconnected vis-à-vis the interrelations of the model's features and propositions. Revision of a mental model would have a substantial impact on multiple components of that model for problem solving. Alternatively, if students are acquiring a problem-solving schema, then students should show differential learning gains on specific knowledge components. They should show large gains on those knowledge components that were missing from their original problem-solving schema or that could not be induced from problem solving. To provide a test of this hypothesis, we examined three knowledge components to track the learning gains and losses across problems for each condition. Because we expected learning gains to be associated with gap-filling processes, we expected improvements on only some of the knowledge components. Specifically, we predicted learning gains for students given the gap-filling prompts for a knowledge component that depends on understanding the underlying concept (i.e., the principle definition of an electric field).

Results

The results are divided into three sections: worked examples, problem solving, and a learning curve analysis. In the worked examples section, we examined the effect of the different prompts on students' study times and the relationship between the study times and coached problem solving. In the problem-solving section, we examined students' performance on solving the problems in the tutoring system. In the learning curves analysis, we examined how much assistance students required from the tutor for particular knowledge components. The alpha level was set to $\alpha = .05$ for all main effects, interactions, and planned comparisons (Keppel 1991). Effect sizes (partial eta squared, η_p^2) were calculated for main effects, interactions, and planned comparisons. Cohen (1988; see also Olejnik and Algina 2000) has suggested that effects be regarded as small when $\eta_p^2 < .06$, as medium when $.06 < \eta_p^2 < .14$ and as large when $\eta_p^2 > .14$.

Studying worked examples

In this section, we examined how the prompt type affects the amount of time students study each example. Figure 4 shows the average time spent studying examples 1 and 2 as a function of prompt. A 3 (prompt: justification, step-focused, or mental-model revision) $\times$ 2 (example: 1 vs. 2) mixed-model analysis of variance (ANOVA) revealed a large effect for example (F (1, 44) $= 13.57$, $p < .05$, $\eta_p^2 = .24$) with students spending more

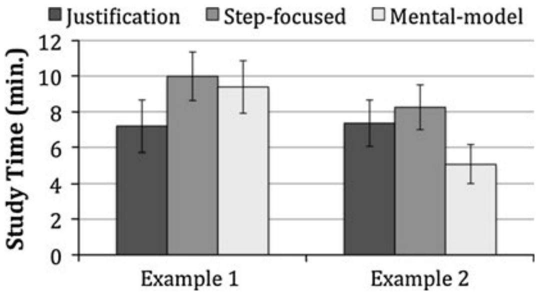

Fig. 4 Mean (±SE) study time for each example as a function of condition

time studying the first example than the second. There was no effect for prompt, $F < 1$. However, there was a large interaction of prompt by example, F (2, 44) = 5.84, $p < .05$, $\eta_p^2 = .21$. To better understand the interaction, we conducted a simple-effects analysis for each group. Both the step-focused and the mental-model revision groups spent significantly less time studying the second example, F (1, 14) = 12.44, $p < .05$, $\eta_p^2 = .47$ and F (1, 15) = 7.46, $p < .05$, $\eta_p^2 = .33$ respectively. In contrast, the justification group showed no difference in the amount of time spent studying both examples, $F < 1$. One plausible explanation for these differences is that the participants in the step-focused and mental-model revision groups found the prompts less useful than the justification group; therefore, they decided to spend less time with the second example.

Next, we examined the relationship between study time and problem-solving performance. Collapsing across conditions, study times were negatively correlated with the amount of assistance (i.e., number of hint requests + errors) that the students required from the tutor to solve the three problem-solving tasks, $r(47) = -.42$, $p < .05$. This was also true when the analyses were restricted to the assistance scores for only the second and third problems (i.e., those problems that directly followed example study, $r(47) = -.40$, $p < .05$). Moreover, this pattern of correlations was significant for only the justification and step-focused groups, $r(16) = -.53$, $p < .05$ and $r(16) = -.52$, $p < .05$ respectively. In contrast, the correlation for the mental-model revision group was not significant, $r(15) = -.17$, ns.

These results suggest that the more time students spent on the examples, the fewer hint requests and errors they made while solving problems; therefore, study time was an indicator of learning from the examples. Furthermore, this relationship was the strongest for the justification and step-focused groups, suggesting that they learned more from example study than the mental-model revision group. Next we examine how example study impacts problem-solving performance and how much help students required from the tutor.

Problem solving

We predicted that the students given the mental-model revision prompts would learn less from the examples than the students in the other two conditions and thus rely more on the scaffolding supplied by the tutoring system for successful problem solving. The first measure we used to test this prediction was the students' *problem-solving scores* for each problem. The problem-solving score can vary from 0 to 100 and is displayed to students on the lower right corner of the application screen when working on a problem (see Fig. 3). The score is determined by the number of points earned for correct entries and the number of penalties for requesting help. For each help request, five points are subtracted from their score. Points are awarded based on the following rubric: 35% based on proportion of correct *equations* entered, 25% based on the proportion of *answer boxes* correctly entered, 20% based on the proportion of correct *given quantities* entered, and 20% based on the proportion of correctly *drawn components* in the diagram. The score reflects the number of help requests and errors that students make when solving a problem. Problem 1 served as a measure of prior knowledge and problem-solving ability for each student before the intervention. We conducted a one-way analysis of variance and there were no differences across the groups for problem 1, $F < 1$, suggesting that the groups were similar in initial knowledge and ability. Next we examined how the manipulation of prompts at example study impacts problem-solving performance on problems 2 and 3. See Fig. 5 for the mean problem-solving scores and standard errors for each condition.

We conducted a 3 (prompt: justification, step-focused, mental-model revision) × 2 (problem: 2 vs. 3) mixed analysis of covariance (ANCOVA) with students' problem-solving scores on problem 1 serving as the covariate. The analysis revealed a large effect of prompt, $F\ (2, 43) = 3.93$, $p < .05$, $\eta_p^2 = .16$. There was no effect of problem, covariate, or any interactions, F's < 1.61. Follow-up analyses for the effect of prompt revealed that both the justification and step-focused groups performed better than the mental-model revision group on problem-solving performance, $F\ (1, 28) = 4.03$, $p = .054$, $\eta_p^2 = .12$ and $F\ (1, 28) = 5.73$, $p < .05$, $\eta_p^2 = .17$ respectively. This result is consistent with our expectations that the students given mental-model revision prompts would require more assistance from the tutor than the students given the justification and step-focused prompts. However, in contrast to our expectations, these results suggest that there were no benefits of the justification prompts over the step-focused prompts. Next, we consider the number of bottom-out hints students requested from the tutor.

The bottom-out hint provides a measure of assistance from the tutor that indicates students have requested multiple hints and are requesting the tutor provide them with the solution for the current step. The bottom-out hint is a measure of the most extensive assistance the student can request. Figure 6 presents the average number of bottom-out hints each group requested from the tutor on problems 2 and 3.

We conducted a 3 (prompt: justification, step-focused, and mental-model revision) × 2 (problem: 1 vs. 2) mixed ANCOVA on the number of bottom-out hints requested for problems 2 and 3 with the number of bottom-out hints from problem 1 used as the covariate. We use the number of bottom-out hints on problem 1 as the covariate because it provides a measure of initial assistance each group required before the intervention, again providing a proxy of prior knowledge and problem-solving ability in Andes. The analysis revealed an effect of the covariate, $F\ (1, 43) = 9.08$, $p < .05$, $\eta_p^2 = .17$, showing that

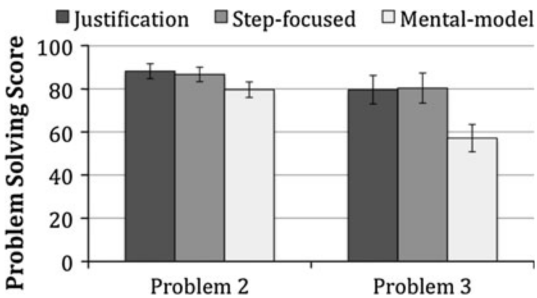

Fig. 5 Mean problem-solving scores for each group for problems 2 and 3

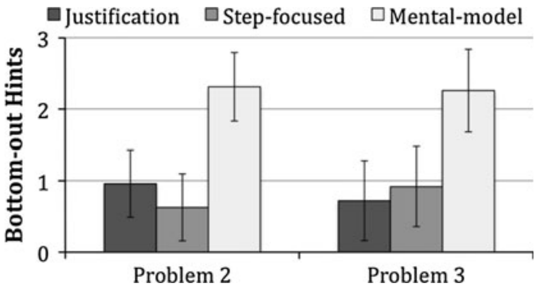

Fig. 6 Mean (±SE) frequency of bottom-out hint requests

bottom-out hints on problem 1 predicted students bottom-out hints on problems 2 and 3. There was also a large effect of prompt, F (2, 43) = 3.53, $p < .05$, $\eta_p^2 = .14$, showing that the groups differed in the number of bottom-out hints requested on problems 2 and 3 when controlling for bottom-out hints on problem 1. There was no effect for problem or interaction of prompt by problem, F's < 1.81. Follow-up analyses for the effect of prompt revealed that both the justification and step-focused groups made fewer bottom-out hint requests than the mental-model revision group, F (1, 28) = 3.70, $p = .065$, $\eta_p^2 = .12$ and F (1, 28) = 4.04, $p = .054$, $\eta_p^2 = .13$ respectively.

These results are consistent with the problem-solving score showing that participants in the mental-model revision group requested more extensive assistance from the tutor during problem solving than either of the other two groups. This set of results suggests that students given the justification and step-focused prompts learned more from example study than the students given mental-model prompts and thus required less assistance on subsequent problem solving.

Learning curves analysis

The analyses reported above provide evidence that learning did in fact occur from the worked examples. To support this hypothesis, we used an "embedded assessment" technique, which follows from theories of cognitive skill acquisition (Anderson 1993; Anderson and Lebiere 1998). Specifically, as an individual becomes more proficient in deploying a skill, her error rate and amount of assistance decreases with each opportunity to apply that skill. In the present case, assistance came in the form of on-demand hints and immediate error flagging. To evaluate learning, an assistance score, which is the sum of the number of hints and errors for each knowledge component, was calculated for each opportunity (i.e., problem). The score varies from 0 to 1 depending on the amount of assistance the student required in applying that knowledge component for each opportunity. For example, if the knowledge component was perfectly executed with no assistance on first attempt, then the student would receive a score of 0 for that knowledge component on that opportunity. If we examine performance across the problems, the result is a learning curve that represents the assistance score as a function of opportunity. Note that a lower assistance score indicates more domain-relevant knowledge and less reliance on the tutor for help.

For this analysis we focus on three knowledge components that were present across all three problems: drawing and defining the electric-field vector, drawing and defining the electric-force vector, and writing the principle definition of an electric field. Our prediction is that the gap-filling prompts would be particularly beneficial on the principle definition knowledge component to the extent that participants would generate explanations of the principle. In contrast, we expected relatively less prompt-based learning for knowledge components that required drawing vectors. We examined each knowledge component for evidence of learning and whether there were differential learning gains across opportunities as a function of condition. We begin with the knowledge component of an electric-field vector. This knowledge component consists of correctly specifying the magnitude, direction, region, source, and label of the vector. The learning curve for the electric field vector for each condition as a function of opportunity can be found in Fig. 7.

To analyze the learning gains, we conducted 3 (prompt: justification, step-focused, mental-model revision) $\times$ 3 (opportunity: 1, 2 and 3) mixed ANOVA for the amount of assistance required for the electric field vector knowledge component. The analysis revealed a large effect of opportunity, $F\,(2, 88) = 48.63, p < .05, \eta_p^2 = .52$, but no effect of prompt or interactions between prompt and opportunity, F's $< .1$. Follow-up analyses for opportunity revealed a significant linear decrease across opportunities, F (1, 44) $= 106.29$, $p < .05$, $\eta_p^2 = .71$. These results show that students across all three groups required less assistance from the tutor across opportunities. Students showed much improvement across trials 1–2 and from 2 to 3. These improvements indicate large learning gains for this knowledge component. However, since all three prompting groups showed equal improvements across the opportunities it is unlikely that this knowledge component was the source of the problem-solving improvements for the gap-filling conditions. Each condition was equally effective in supporting learning of this knowledge component across practice opportunities.

The second component we focus on is correctly drawing and defining the electric force vector. Similar to the electric field vector, this knowledge component consists of correctly specifying the magnitude, direction, body and time, source and type, and label of the vector. The learning curve for the electric force vector for each group as a function of opportunity can be found in Fig. 8.

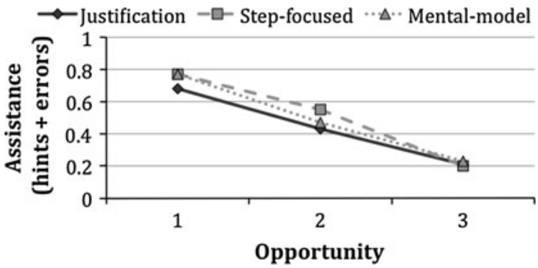

Fig. 7 Learning curves for drawing and defining the electric-field vector

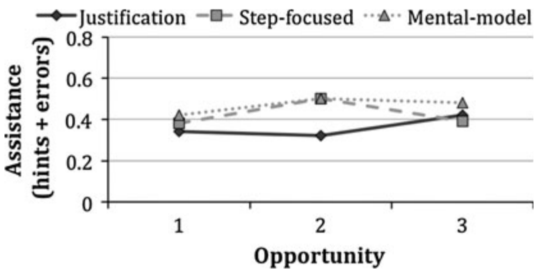

Fig. 8 Learning curves for correctly drawing and defining an electric force vector

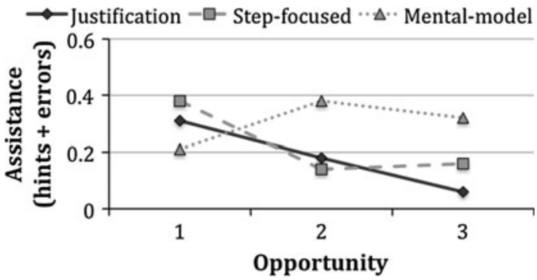

Fig. 9 Learning curves for the definition of an electric field

Inspection of the means suggests that a moderate amount of assistance was required initially to successfully use this component and that there were no learning gains across trials for any of the groups. This observation was confirmed by a 3 (prompt: justification, step-focused, mental-model revision) × 3 (opportunity: 1, 2 and 3) mixed ANOVA that revealed no effect of prompt, opportunity, or interactions, all F's < 1.03. These results indicate that there were no learning gains for any of the groups. In contrast to the previous analysis, this result highlights that, although participants showed learning for some components of problem solving, there were others that were not impacted by either problem solving or example study.

Finally, we examined the definition of the electric field. This knowledge component consists of correctly specifying the electric field principle by writing the following equation: $\mathbf{F} = q\mathbf{E}$. Figure 9 shows the average assistance for each group as a function of opportunity.

To analyze the learning gains, we conducted 3 (prompt: justification, step-focused, mental-model revision) × 3 (opportunity: 1, 2 and 3) mixed ANOVA for the amount of assistance required on the definition of an electric field. The analysis revealed no effect of opportunity or prompt, F's < 2.21. However, there was a medium-sized interaction of prompt by opportunity, $F\ (4, 88) = 3.03$, $p < .05$, $\eta_p^2 = .12$. To follow up the interaction we conducted a simple-effects analysis with separate linear ANOVAs for each group. Analyses revealed that the justification and step-focused groups show significant linear decreases across opportunities, $F\ (1, 15) = 13.53$, $p < .05$, $\eta_p^2 = .47$ and $F\ (1, 15) = 5.21$, $p < .05$, $\eta_p^2 = .25$, respectively. In contrast, the mental-model revision group did not show an effect of opportunity, $F < 1$. In fact, they showed an increase in the amount of assistance required from opportunities 1 to 2.

Several interesting results are illustrated in this knowledge decomposition analysis. First, it is clear that some learning takes place at the knowledge component level (Figs. 7, 9). Furthermore, such an analysis helps pinpoint which components of problem solving are affected by the intervention. From a simple analysis of three knowledge components, we observed three very different patterns of performance across the groups. For the electric field vector, we saw large learning gains across all three prompting conditions, suggesting much learning but similar improvements across conditions. For the electric-force vector, we saw less initial difficulty with the knowledge component but no learning gains across opportunities for any of the groups. Finally, for the definition of the electric field, the learning curve evidence suggests that, as predicted, the participants given the justification and step-focused prompts learned more than the participants given mental-model revision prompts because their assistance scores decreased at a faster rate. This result is consistent with the hypothesis that justification and step-focused prompts facilitate inference generation connecting the solution step (e.g., electric field equation) to the application conditions for when and how to use that concept.

Discussion

One of the goals of cognitive science is to develop theories that have a broad impact. The self-explanation principle has the potential to have a large effect on instructional design and improving student learning in school settings. We tested two classes of self-explanation prompts in the domain of electrodynamics problem solving. We hypothesized that gap-filling prompts would be the most effective because they would facilitate inference generation and promote the acquisition of a problem-solving schema. The acquisition of such knowledge should facilitate solving new problems that use those same concepts and problem-solving steps. In contrast, we expected that the mental-model revision prompts would be the least effective because they were designed to focus the student on revising her prior misconceptions, and previous work has shown that students do not typically have robust misconceptions for these problem types. The results supported these hypotheses, showing that the largest learning gains were for students who were given gap-filling prompts when studying the worked examples. The goal here was to examine how a cognitive science principle translates into an educational application, which critically depends on the interpretation of that principle (i.e., the theoretical assumptions upon which it is based) and the fit with the learning scenario (i.e., prior knowledge, target knowledge, and task constraints).

The second goal was to test the instructional fit hypothesis, which is the idea that there is an interaction between the student's prior knowledge, the cognitive processes that are evoked by different types of instruction (e.g., prompting), the modifications those processes make to the knowledge representation, and the utility of that representation for the given task. The results provide support for this hypothesis showing that students in the justification and step-focused conditions benefited more from studying examples than the mental-model revision condition. This result is consistent with the idea that the prompts in the justification and step-focused conditions facilitated the acquisition of a problem-solving schema and the generation of justifications for each step of the problem. Further evidence for this conclusion comes from the knowledge component analysis that shows the students in the gap-filling conditions showed improvements on knowledge components that were missing conceptual justifications (i.e., electric field definition) after example study, whereas the students given mental-model revision prompts did not. The fact that both justification and step-focused groups performed equally well suggests that students in the step-focused condition spontaneously generated justifications for each step.

The fact that the mental-model revision group relied more heavily on the tutoring support is consistent with the hypothesis that students learn less from prompts triggering cognitive processes that are not a good match for that particular learning scenario (student prior knowledge and task structure). These results are also consistent with the notion that students might ignore instructional activities that they feel do not help them achieve their learning goals. For instance, the study times for the students given the mental-model revision prompts dropped dramatically between the first and second examples, suggesting either the students did not perceive the prompts as useful, or they did not have any existing prior knowledge to revise. Instead, they relied on the on-demand hints from the tutor to solve the problems. Another possible explanation is that the students adopted a 'gaming the system' problem-solving strategy because they did not find the mental-model prompts useful. Prior work has shown that gaming the system behaviors, such as exploiting the tutor's feedback and help features to solve the problems, have a negative correlation with learning (Baker et al. 2004). Future work is needed to differentiate between these two possibilities.

There are two limitations to the current study. First, we only tested one aspect of the instructional-fit hypothesis, by comparing the instructional fit of two classes of prompts in a learning scenario where the participants were expected to have no prior misconceptions. Future work should examine other components of the hypothesis by conducting a full factorial study crossing knowledge (known misconception vs. no misconception) with prompting type (gap-filling vs. mental-model revision). In this type of design, gap-filling would be expected to be most helpful on the task where students have no misconceptions (as shown in the current study) and the mental-model revision prompts would be expected to be most helpful on the task where students were predicted to have a misconception. A second limitation of the current study was that our posttest assessment was relatively brief, in that we examined tutored problem-solving performance on only two challenging problems. Although the results showed large learning and problem-solving differences after a very short instructional period, future work should examine additional measures of learning, including qualitative reasoning measures and measures of long-term retention.

This work has several implications for classroom instruction and pedagogy. First, this study highlights the difficulty of translating learning science principles discovered in the laboratory for effective classroom instruction. One cannot simply take a learning principle off the shelf and apply it to any academic topic or task and expect to get robust learning results. Every learning principle comes with a theory of *cognitive change* (i.e., the underlying cognitive processes and the mental representations that those processes act on), and is best suited for a particular type of learning scenario (i.e., prior knowledge, target knowledge, task constraints, domain, etc.). To maximize the effectiveness of a given learning principle a teacher must understand the mechanics for how that principle works and the application conditions for which that principle will be most effective.

A challenge for cognitive and learning scientists is to articulate the application conditions upon which the principles will operate most effectively and to test different instantiations of the principle. Our work takes some first steps towards this goal for the self-explanation principle but similar studies should be conducted for other principles including analogical comparison, worked examples, structured inquiry tasks, and spaced practice, among others. Similar approaches can be found at the Pittsburgh Science of Learning Center (PSLC; in their section on developing instructional principles, www.learnlab.org) and the research that has contributed to the development of the *practice guides* produced by the Department of Education, Institute for Education Sciences (e.g., Pashler et al. 2007).

Second, this work has implications for the design of intelligent tutoring systems. Intelligent tutoring systems are well known to be powerful learning environments because they can be adaptive to student learning and provide scaffolding and support to students who struggle (Anderson et al. 1995; Koedinger et al. 1997; VanLehn et al. 2005). Much of the support students receive comes in the form of immediate feedback, instructional prompts, worked examples, direct instruction, and hints. The current work suggests that great care needs to be taken when determining what types of prompts to use and that some prompts will be more or less effective depending on the student's prior knowledge and the nature of the task and domain. Our work suggests that tutors can become even more effective if they adaptively implement multiple instructional prompts (e.g., gap-filling or mental-model revision for self-explanation) depending on student knowledge (knowledge tracing tutors) and the task constraints. Such adaptations will make an already powerful learning technology even more effective.

The instructional fit hypothesis applies beyond tutoring environments and has similar implications for in-class instruction. Most closely related to the prompts explored in the current study are the types of questions and prompts that teachers ask during class (e.g., Chapin et al. 2003). These can be directed as gap-filling or mental-model revision. The current work suggests that the teacher adapts the use of such prompts based on the prior knowledge of the class, the target knowledge, learning goals, etc. This view of instructional fit suggests that the instructor focuses on the alignment of the curricula and instruction to the students' prior knowledge, target knowledge, and task constraints. In conclusion, there is a great deal of importance and difficulty when interpreting learning principles and implementing them as instructional interventions. Research at the nexus of cognitive theory, technological tools, and classroom work is uniquely situated to address such multidisciplinary problems.

Acknowledgments This work was supported by the National Science Foundation, Grant Number SBE-0354420 to the Pittsburgh Science of Learning Center (http://www.learnlab.org). No endorsement should be inferred. Portions of the results were presented at the Cognitive Science Society's 31st Annual Conference. We would like to thank the members of the Cognitive Science Learning Laboratory and several anonymous reviewers for their many helpful comments and suggestions.

Appendix

Test problems and examples

Problem	Problem statement	Equations
Problem 1	A latex sphere (charge $q = -1.6\text{e}{-19}$ C and mass $m = 7.3\text{e}{-15}$ kg) is in a region of a uniform electric field E. The force on the latex sphere due to the electric field exactly cancels its weight near the Earth's surface. Show the direction of the electric field on the diagram. Determine the magnitude of the electric field.	$\mathbf{F_e} = q\mathbf{E}$ $\mathbf{F_w} = m\mathbf{g}$ $\mathbf{F_e} + \mathbf{F_w} = 0$
Example 1	A charged particle ($q = 52.0$ μC) is in equilibrium in a region close to the Earth's surface where there is a uniform upward electric field E of magnitude 120 N/C. The y-component of the net force on the particle near Earth due to the electric field and gravity is zero. What is the mass of the particle?	$\mathbf{F_e} = q\mathbf{E}$ $\mathbf{F_w} = m\mathbf{g}$ $\mathbf{F_e} + \mathbf{F_w} = 0$
Problem 2	An electron ($qe = -1.60\text{e}{-19}$ C; $me = 9.11\text{e}{-31}$ kg) is in a region, between two parallel charged plates, that produce a uniform electric field E of magnitude 2.0e+4 N/C. The electron undergoes a constant acceleration from rest near the negative plate. Find the velocity of the electron after it travels 1.5 cm from the negative plate. In this problem, gravity can be ignored.	$\mathbf{F_e} = q\mathbf{E}$ $\mathbf{F_e} = m\mathbf{a}$ $\mathbf{v_1} = \mathbf{v_0} + \mathbf{at}$ $\mathbf{\Delta x = v_o t + at^2/2}$
Example 2	A proton ($qp = 1.6\text{e}{-19}$ C; $mp = 1.7\text{e}{-27}$ kg) is in a region where there is a uniform electric field E of magnitude 320 N/C, directed along the positive x-axis. The proton accelerates from rest and reaches a speed of 1.20e+5 m/s. How long does it take the proton to reach this speed? In this problem, gravity can be ignored.	$\mathbf{F_e} = q\mathbf{E}$ $\mathbf{F_e} = m\mathbf{a}$ $\mathbf{v_1} = \mathbf{v_0} + \mathbf{at}$
Problem 3	An oil drop ($q = -6.40\text{e}{-19}$ C; $m = 5.0\text{e}{-16}$ kg) is in a region where there is a uniform electric field E of magnitude 5,000 N/C in the negative y-direction. The oil drop is moving in the positive y-direction at an initial velocity of 4.0 m/s. How far will the oil drop travel before it comes to rest? In this problem, please include gravity.	$\mathbf{F_e} = q\mathbf{E}$ $\mathbf{F_w} = m\mathbf{g}$ $\mathbf{F_e} + \mathbf{F_w} = m\mathbf{g}$ $\mathbf{v_1} = \mathbf{v_0} + \mathbf{at}$ $\mathbf{\Delta x = v_o t + at^2/2}$

References

Aleven, V. A. W. M. M., & Koedinger, K. R. (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer-based Cognitive Tutor. *Cognitive Science, 26*, 147–179.

Anderson, J. R. (1993). *Rules of the mind*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *The Journal of the Learning Sciences, 4*(2), 167–207.

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Mahwah, NJ: Lawrence Erlbaum Associates.

Baker, R. S., Corbett, A. T., Koedinger, K. R., & Wagner, A. Z. (2004). Off-task behavior in the Cognitive Tutor classroom: When students "game the system". *Proceedings of ACM CHI 2004: Computer-Human Interaction*, 383–390.

Berthold, K., Eysink, T. H., & Renkl, A. (2009). Assisting self-explanation prompts are more effective than open prompts when learning with multiple representations. *Instructional Science, 37*, 345–363.

Catrambone, R. (1998). The subgoal learning model: Creating better examples so that students can solve novel problems. *Journal of Experimental Psychology: General, 127*(4), 355–376.

Chapin, S., O'Connor, C., & Anderson, N. (2003). *Classroom discussions: Using math talk to help students learn, grades 1–6*. Sausalito, CA: Math Solutions Publications.

Chi, M. T. H. (2000). Self-explaining expository texts: The dual processes of generating inferences and repairing mental models. In R. Glaser (Ed.), *Advances in instructional psychology* (pp. 161–238). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Chi, M. T. H. (2008). Three kinds of conceptual change: Belief revision, mental model transformation, and ontological shift. In S. Vosniadou (Ed.), *Handbook of research on conceptual change* (pp. 61–82). New York, NY: Routledge.

Chi, M. T. H., & Bassok, M. (1989). Learning from examples via self-explanations. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 251–282). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Chi, M. T. H., DeLeeuw, N., Chiu, M.-H., & Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18*, 439–477.

Cohen, J. (1988). *Statistical power analysis of the behavioral sciences* (2nd ed.). New York: Academic Press.

Conati, C., & VanLehn, K. (2000). Toward computer-based support of meta-cognitive skills: A computational framework to coach self-explanation. *International Journal of Artificial Intelligence in Education, 11*, 398–415.

Hausmann, R. G. M., & Chi, M. T. H. (2002). Can a computer interface support self-explaining? *Cognitive Technology, 7*(1), 4–14.

Hausmann, R. G. M., & VanLehn, K. (2007). Explaining self-explaining: A contrast between content and generation. In R. Luckin, K. R. Koedinger, & J. Greer (Eds.), *Artificial intelligence in education: Building technology rich learning contexts that work* (Vol. 158, pp. 417–424). Amsterdam: IOS Press.

Keppel, G. (1991). *Design and analysis: A researcher's guide*. Upper Saddle River, NJ: Prentice Hall.

Koedinger, K. R., Anderson, J. R., Hadley, W. H., & Mark, M. A. (1997). Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education, 8*, 30–43.

Maloney, D. P., O'Kuma, T. L., Hieggelke, C. J., & Van Heuvelen, A. (2001). Surveying students' conceptual knowledge of electricity and magnetism. *American Journal of Physics, 69*(7), S12–S23.

Marshall, S. P. (1995). *Schemas in problem solving*. Cambridge, UK: Cambridge University Press.

McNamara, D. S. (2004). SERT: Self-explanation reading training. *Discourse Processes, 38*(1), 1–30.

Nokes, T. J., Schunn, C. D., & Chi, M. T. H. (2010). Problem solving and human expertise. In P. Peterson, E. Baker, & B. McGraw (Eds.), *International encyclopedia of education* (Vol. 5, pp. 265–272). Oxford: Elsevier.

Olejnik, S., & Algina, J. (2000). Measures for effect size for comparative studies: Applications, interpretations, and limitations. *Contemporary Educational Psychology, 25*, 241–286.

Pashler, H., Bain, P., Bottge, B., Graesser, A., Koedinger, K., McDaniel, M., & Metcalfe, J. (2007). *Organizing Instruction and Study to Improve Student Learning* (NCER 2007–2004). Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education.

Renkl, A. (2002). Learning from worked-out examples: Instructional explanations supplement self-explanations. *Learning & Instruction, 12*, 529–556.

Rittle-Johnson, B. (2006). Promoting transfer: Effects of self-explanation and direct instruction. *Child Development, 77*, 1–15.

Ross, B. H. (1984). Remindings and their effects in learning a cognitive skill. *Cognitive Psychology, 16*, 371–416.

Ross, B. H., & Kilbane, M. C. (1997). Effects of principle explanation and superficial similarity on analogical mapping in problem solving. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 23*(2), 427–440.

Singley, M. K., & Anderson, J. R. (1989). *The transfer of cognitive skill*. Cambridge, MA: Harvard University Press.

VanLehn, K. (1998). Analogy events: How examples are used during problem solving. *Cognitive Science, 22*(3), 347–388.

VanLehn, K., Lynch, C., Schultz, K., Shapiro, J. A., Shelby, R., Taylor, L., et al. (2005). The Andes physics tutoring system: Lessons learned. *International Journal of Artificial Intelligence and Education, 15*(3), 147–204.