Learning and Instruction 25 (2013) 104–124

Contents lists available at SciVerse ScienceDirect

Learning and Instruction

journal homepage: www.elsevier.com/locate/learninstruc

How much is too much? Learning and motivation effects of adding instructional explanations to worked examples

J. Elizabeth Richey*, Timothy J. Nokes-Malach

Department of Psychology, Learning Research and Development Center, University of Pittsburgh, United States

A R T I C L E I N F O

Article history:
Received 7 November 2011
Received in revised form
21 November 2012
Accepted 26 November 2012

Keywords:
Achievement goals
Learning
Motivation
Assistance
Instructional explanations

A B S T R A C T

A central goal of the learning sciences is to discover principles that determine the optimal amount of instructional assistance to support robust learning (Koedinger & Aleven, 2007). We examined learning outcomes from providing and withholding stepwise instructional explanations as students studied worked examples and solved physics problems. We hypothesized that students would acquire more conceptual knowledge from withholding instructional explanations because they would be more likely to engage in constructive cognitive activities to understand the problem-solving steps, whereas providing instructional explanations might suppress such activities. Furthermore, we examined the roles of prior knowledge and student motivation in determining learning outcomes. Across three experiments, students in the withholding conditions showed greater conceptual learning than students in the providing conditions. Additionally, achievement goal orientations were more predictive of learning for the withholding conditions than the providing conditions. We discuss how the interactions between prior knowledge, motivation, and instruction can support learning and transfer.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

Instructors in every domain face a common challenge in determining when to provide students with explanations and when to have them generate their own. This challenge creates a pedagogical dilemma for choosing between the intuitive merits of two instructional approaches. On one hand, providing detailed examples and instructional explanations can help a learner obtain an accurate understanding of a topic in a relatively quick, efficient manner by focusing attention on appropriate solution paths and key features while discouraging the use of inefficient or inaccurate strategies. On the other hand, leaving a learner to figure out a problem on her own can promote constructive cognitive activities such as self-explanation, which can facilitate a deep understanding of the materials (Chi, Slotta, & de Leeuw, 1994; Renkl, 1997). Although the challenge of finding an appropriate level of instructional assistance arises in a number of learning situations, it is especially salient in the domains of math and science instruction, where common instructional approaches range from solving open-ended problems to studying highly scaffolded worked examples that incorporate instructional explanations.

The advantages and disadvantages of providing versus withholding information have been explored from a number of perspectives in the learning sciences, including research on desirable difficulties (Bjork, 1994; Schmidt & Bjork, 1992), worked examples (Renkl, Atkinson, & Maier, 2000), and the assistance dilemma in intelligent tutoring systems (Koedinger & Aleven, 2007). Recently, Wittwer and Renkl (2010) published a meta-analysis showing that providing instructional explanations in worked examples (i.e., explanations of either the principles or operators applied in accompanying worked examples) had a positive effect on the acquisition of conceptual knowledge, but not on problem-solving skills. However, the effects were found only in the domain of math, not in science or learning science disciplines, and they disappeared when compared to worked examples that encouraged self-explanation. These findings show that there are important moderating factors on the effectiveness of providing instructional explanations, and that there may be particular situations in which withholding such explanations would be beneficial for learning and transfer. To further investigate this issue, we compared learning outcomes from providing versus withholding instructional explanations as students studied worked examples and solved practice problems in electricity. To determine what was learned, we measured conceptual reasoning, problem-solving

* Corresponding author. University of Pittsburgh, Learning Research and Development Center, 3939 O'Hara Street, Pittsburgh, PA 15260, United States. Tel.: +1 412 648 1109; fax: +1 412 624 9149.

E-mail address: jes1235@pitt.edu (J.E. Richey).

0959-4752/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.learninstruc.2012.11.006

performance, and preparation for future learning. We also explored the role of two factors hypothesized to be particularly important for learning from withholding instructional explanations: the roles of *prior knowledge* and *achievement motivation*.

First, we tested the hypothesis that worked examples and problem-solving activities that withhold instructional explanations promote deeper conceptual learning than activities that provide explanations [Hypothesis 1, Experiments 1, 2, and 3]. Second, we examined whether such learning depends on knowing the relevant ontological categories for the to-be-learned science concepts [Hypothesis 2, Experiment 1]. Third, we tested the hypothesis that students' achievement orientations have a larger effect on learning outcomes when instructional explanations are withheld compared to when they are provided [Hypothesis 3, Experiments 1, 2, and 3]. We expected that withholding explanations would force students to rely more on their personal achievement motivations (e.g., strive for understanding or performance) to regulate their learning activities and behaviors. Furthermore, we examined the possibility that withholding instructional explanations might promote the adoption of mastery goals (i.e., the desire to understand) because the materials put more responsibility on the students to make sense of them, in contrast to telling the students what they needed to know [Hypothesis 4, Experiments 2 and 3].

In the sections that follow, we describe the cognitive and motivational processes that providing and withholding instructional explanations are hypothesized to support (Section 1). We then present three experiments that examine what is learned from withholding or providing instructional explanations in worked examples and problem-solving activities. We also examine the roles of prior knowledge and achievement orientations in that learning (Sections 2–4). We conclude with a discussion of the results and implications for instructional theory (Section 5).

1.1. Balancing withholding and providing information

The question of whether providing or withholding information in worked examples and problem-solving activities leads to better learning outcomes depends on a number of instructional factors including the nature of the information (problem-solving steps versus instructional explanations), the amount of information provided (a little or a lot), and when the information is provided or withheld (early or late in practice). Many experiments have examined a direct comparison of problem solving, an activity that provides little to moderate assistance depending on whether any help is given in addition to the problem, against worked example study, which provides more assistance by illustrating the solution steps and final answer to the problem. The results have generally favored the use of worked examples interleaved with practice problems over problem solving alone to support learning and transfer (e.g., Renkl et al., 2000; Renkl, Atkinson, Maier, & Staley, 2002; Ward & Sweller, 1990).

Providing worked examples along with practice problems improves learning and reduces memory load by eliminating the need for the learner to maintain too many pieces of knowledge in working memory at a given time and instead allowing her to utilize the information provided in the worked example (Paas & Van Merriënboer, 1994; Ward & Sweller, 1990). Additionally, worked examples can support more efficient learning by reducing the pursuit of incorrect solution paths and focusing the student on the correct problem steps. For example, within the context of an intelligent tutoring system, Mclaren, Lim, and Koedinger (2008) found that a group that solved problems with interleaved worked examples achieved mastery in significantly less time than a group that just solved problems.

Providing too much information, however, may come with a cost. Renkl et al. (2000) found that decreasing the amount of information provided across a series of worked examples – a process the authors called "fading" – improved performance on near-transfer problems (i.e., problems with a similar structure to the examples) compared to a condition that continued to receive complete worked examples throughout the sequence. The authors concluded that the process of generating the missing steps gave students in the fading condition a learning advantage. These results suggest that providing some information is fruitful, but withholding information at critical junctures may also facilitate constructive cognitive processes that improve learning and transfer.

In addition to providing or withholding worked examples or steps of worked examples, researchers have also examined providing instructional explanations within worked examples. Instructional explanations typically consist of definitions of the key concepts and principles used in the examples as well as descriptions of the relationships between those concepts (Leinhardt, 2010; Renkl, 2002; Wittwer & Renkl, 2008). They can also include descriptions of the links between goals and operators as well as the application conditions for using those operators (Wittwer & Renkl, 2010; van Gog, Paas, & van Merriënboer, 2008). Given that not all instructional explanations are productive, much recent work has gone into determining what makes instructional explanations effective, both in classroom dialog and in written materials (Leinhardt, 2001; Renkl, 2002; Schworm & Renkl, 2006; Wittwer & Renkl, 2008, 2010). For example, Wittwer and Renkl (2008) reviewed the instructional explanations literature in an effort to identify the key characteristics of explanations that support robust learning. They argued that instructional explanations should be adapted to learners' prior knowledge, focused on principles or conceptual information, and designed to engage learners in constructing or applying knowledge. We used these recommendations to guide the development of the instructional explanations tested in the current studies.

1.2. Effects of inference generation

Evidence from worked example experiments (e.g., Renkl, 1997; Renkl et al., 2000) suggests that inference generation might be the key cognitive process driving benefits of withholding information. To test this hypothesis, Hausmann and VanLehn (2007) compared the learning outcomes of students who were instructed to self-explain worked examples (i.e., engaging in inference generation) to students who were asked to paraphrase those same examples (i.e., suppressing inference generation). Regardless of whether the worked examples were complete or incomplete, the students who self-explained performed with greater accuracy on the learning materials and on both near- and far-transfer homework problems (i.e., problems that had either similar or different structures compared to the learning problems). The results suggest the self-explanation prompts triggered inference generation, which supported greater learning gains than simply paying attention to the provided instructional explanations. Consistent with these findings, Schworm and Renkl (2006) found that self-explanation prompts improved math teachers' learning outcomes, while providing them with instructional explanations reduced spontaneous self-explanations and, in turn, negatively affected their learning. These results indicate that it is not simply the type of information provided in the worked examples that is important but also *how* that information is processed (inference generation or paraphrasing). It suggests that students may benefit from materials that encourage them to self-explain, even if they are not able to generate high-quality explanations.

Renkl (2002) suggests that instructional explanations have the potential to scaffold learning in situations in which students struggle to generate self-explanations. Specifically, he argues that simple instructional explanations may be able to help learners when comprehension issues make it difficult for them to generate self-explanations on their own. However, materials that include stepwise instructional explanations could also suppress inference generation because the explanatory information is already present, thereby encouraging more passive learning activities such as rehearsal and paraphrasing (Chi, 2011). In contrast, materials that withhold explanations may promote more constructive learning activities such as inference generation because the explanations are absent and generating them is helpful for understanding the problem solution. Engaging in self-explanation presents the learner with an opportunity to connect her prior knowledge to the new information she is trying to understand or reason about (Chi, de Leeuw, Chiu, & LaVancher, 1994; Nokes, Hausmann, VanLehn, & Gershman, 2011; Rittle-Johnson, 2006). Instructional strategies such as withholding explanations may increase the likelihood that students will attempt to draw on existing knowledge to self-explain and fill in gaps in their understanding, which creates the opportunity to connect existing knowledge to the target problems (Chi, 2000).

1.3. Role of prior knowledge

Past work on self-explanation suggests that prior knowledge may play a particularly important role in learning from self-explanations, as a learner who lacks sufficient prior knowledge may not be able to construct meaningful explanations. Chi et al. have examined this issue by focusing on the effects of ontological category knowledge (Chi, 2008; Slotta & Chi, 2006). Slotta and Chi (2006) proposed that learning certain notoriously difficult concepts in science first requires students to know the appropriate ontological categories for those concepts. They demonstrated that teaching students about *emergent processes*, in which patterns arise from the actions of independent elements such as cars causing a traffic jam, improved students' later learning of a new concept in the same ontological category, such as electricity.

Given the important role of prior knowledge in constructive learning activities, Slotta and Chi's (2006) ontological training approach could be used to provide students with relevant conceptual background knowledge and facilitate learning from a less structured learning environment. This may be particularly important for young learners who have not been previously exposed to the science topics in question. In Experiment 1, we use Slotta and Chi's (2006) emergent process training in conjunction with materials that withhold stepwise instructional explanations at the problem-solving level to test whether relevant prior knowledge from a different domain can enhance learning from worked examples that withhold instructional explanations.

1.4. Motivation

A number of authors have recently called for new research to examine the role of motivation in an effort to better understand learning and transfer from different types of instruction and training (Engle, 2012; Nokes & Belenky, 2011; Nokes-Malach & Mestre, in press; Perkins & Salomon, 2012; Pugh & Bergin, 2006). Past work on the assistance dilemma has focused exclusively on the instructional and cognitive processes responsible for differential learning effects of providing versus withholding information; however, cognitive processes such as self-explanation are volitional and may require a level of effort that many students are not willing to exert. For example, in a pilot study testing the effect of self-explanation prompts in an interactive, high school math-tutoring program, Aleven and Koedinger (2000) found that less than 10 percent of students offered thorough explanations of problem steps. If the act of generating an explanation drives students' learning outcomes, individual differences in motivation likely play a role in moderating the extent to which students are willing to generate inferences and, in turn, how much they will benefit from the activity.

A separate body of work has studied the effects of students' achievement motivation on learning outcomes (Ames & Archer, 1988; Dweck, 1986; Elliot, McGregor, & Gable, 1999; Grant & Dweck, 2003; Harackiewicz, Barron, Pintrich, Elliot, & Thrash, 2002). This research has examined the relationships between different types of achievement goals and positive or negative learning and motivational outcomes. The dominant theoretical framework for considering these goals depends on two dimensions: how a person *defines* competence and the *valence* for achieving that competence (Elliot & McGregor, 2001; Elliot & Murayama, 2008). *Mastery goals* are based on an internal standard with a focus on developing understanding, whereas *performance goals* are based on a normative standard with a focus on demonstrating ability in comparison to other individuals. For the valence dimension, *approach goals* have a positive valence and are characterized by trying to succeed, while *avoidance goals* have a negative valence and are characterized by trying to avoid failure. Crossing the two dimensions results in four separate goals: mastery-approach (e.g., *My goal is to learn as much as possible*), mastery-avoidance (e.g., *My goal is to avoid learning less than it is possible to learn*), performance-approach (e.g., *I strive to do well compared to other students*), and performance-avoidance (e.g., *I strive to avoid performing worse than others*).

Mastery-approach goals have been associated with deep processing, a preference for challenging tasks, greater interest, effective learning strategies, and coping in the face of failure (Ames & Archer, 1988; Elliot et al., 1999; Grant & Dweck, 2003). In contrast, performance-approach goals have led to more mixed results including positive associations such as increased persistence, effort, and positive achievement outcomes, and negative associations such as surface processing and reduced interest (Elliot & Harackiewicz, 1996; Elliot et al., 1999). Performance-avoidance goals have been associated with uniformly negative strategies and outcomes such as surface processing, low interest, and poor performance (Elliot et al., 1999; Hulleman, Durik, Schweigert, & Harackiewicz, 2008; Linnenbrink-Garcia, Tyson, & Patall, 2008). Finally, mastery-avoidance goals are thought to arise when a person seeks to avoid losing competence at a skill or attaining an incomplete understanding of something (Elliot & McGregor, 2001). Elliot and Murayama (2008) found that both the need for achievement and the fear of failure predict mastery-avoidance goals, just as they both predict performance-approach goals. They argue that mastery-avoidance goals may lead to improved outcomes, particularly for students with a strong need for achievement. However, these goals have generally received much less empirical study than the other three goals and thus strong conclusions cannot yet be made about their relation to learning and motivation outcomes. We examine mastery-avoidance goals in Experiments 2 and 3 as we focus on college students learning science concepts (i.e., electricity and electrical circuits) often covered in middle or high school science classes.

Although performance goals may promote achievement when students strive to outperform others, they may also decrease persistence in the face of challenge. For example, when examining children's achievement goals and help-seeking behaviors in an interactive learning environment, Harris, Bonnett, Luckin, Yuill, and Avramides (2009) found that students with performance goals preferred explicit answers from the tutor and moved on to new questions when the hints did not provide clear answers. In contrast, mastery-oriented students used a wider variety of the resources available in the learning environment and tended to prefer hints that did not explicitly provide answers. Based on these results, mastery-oriented students might be expected to respond more positively to withholding materials than performance-oriented students. The approach/avoidance dimension was not examined in the Harris et al. study, so it is unclear how that dimension would affect responses to withholding materials.

It may be that providing or withholding instructional explanations can also alter students' motivation. Providing too much support could make a task seem too easy while not providing enough support could make a task seem too challenging, either of which could reduce a student's motivation to engage in the task. Furthermore, different learning materials may increase or reduce the role of individual differences in predicting learning outcomes. Instructional environments that provide extensive scaffolding may reduce the need for students to regulate their learning and thus decrease the role of motivation in predicting learning outcomes. In contrast, instructional environments that provide little support may force students to rely more on their own motivation to regulate their learning, thus accentuating the impact of individual differences on learning outcomes.

We expect motivation to play a larger role in determining learning outcomes when instructional explanations are withheld compared to when they are provided. In other words, we expect that achievement orientation will moderate performance in the withholding condition but not in the providing condition. Withholding explanations may even trigger the adoption of mastery goals for a given task because students may seek to understand materials by generating their own explanations. This is consistent with recent work by Belenky and Nokes-Malach (2012) showing that invention tasks promote mastery goal adoption, compared to a form of direct instruction that provided a worked example and placed less burden on students to make sense of the materials. In Experiments 2 and 3 we used two different self-report measures to distinguish between the domain-level achievement orientations students reported for science class and the task-level goals they reported during the instructional intervention. For the purposes of clarity, we will refer to the former, more dispositional goals as "achievement orientations" and the latter, more state-based goals as "task goals."

1.5. Measuring learning outcomes

Learning outcomes can be measured in a variety of ways. Prior work on this topic has examined students' learning gains primarily through their performance on new problem-solving tasks. In Wittwer and Renkl's (2010) meta-analysis, only six out of 21 experiments looked at the impact of instructional explanations on conceptual knowledge as measured by questions targeting important principles and concepts. In contrast, 14 measured learning with problem-solving tasks that were structurally similar to those from the learning phase (near transfer), eight measured learning with problems that were structurally different (far transfer), and an additional four measured both but did not separate the two types of problems when reporting results. Problem solving alone does not necessarily provide a robust measure of students' conceptual understanding. For example, in Barnett and Ceci's (2002) taxonomy of transfer tasks, the *content* of what is learned is a critical factor for determining the transfer distance between learning and test. The authors defined near-transfer tasks as those that require the application of familiar problem-solving procedures, such as test problems that have similar structures to the learning problems like those tasks used most often to test learning in Wittwer and Renkl's (2010) meta-analysis. In contrast, they defined far-transfer tasks as those that require the application of concepts or principles from learning, such as targeted conceptual questions or test problems with different structures from the learning problems.

Further evidence for the importance of including multiple dependent measures to assess learning comes from research on physics education. Students frequently achieve problem-solving proficiency without conceptual understanding in physics (Hake, 1998; Hestenes, Wells, & Swackhamer, 1992). For example, Hestenes et al. (1992) demonstrated that students who were able to solve physics problems accurately often lacked the deeper conceptual knowledge needed to explain complex processes or make predictions based on physics principles. Through their development of the Force Concept Inventory, the authors showed that targeted, conceptual assessments were required to measure students' conceptual knowledge. Consequently, research that examines only problem-solving skills may miss important differences in how instructional interventions promote conceptual knowledge acquisition. Work that examines near and far transfer has demonstrated great variety in how different interventions support different kinds of learning, suggesting that a clearer understanding of the effects of withholding and providing instructional explanations will emerge only through an examination of both problem-solving skills and conceptual knowledge (Hausmann & VanLehn, 2007; Lovett, 1992; Nokes, VanLehn, Belenky, Lichtenstein, & Cox, in press). In the current work, we use near-transfer assessments that employ test problems with similar structures to the worked examples, as well as far-transfer assessments that employ multiple-choice and short-answer questions targeting conceptual understanding.

Another way of measuring far transfer and conceptual learning is through Bransford and Schwartz's (1999) "preparation for future learning" paradigm. Preparation for future learning (PFL) tasks examine how instruction prepares a student to learn from new materials and, in turn, how the acquired knowledge transfers to new problem-solving tasks (Bransford & Schwartz, 1999; Schwartz, Bransford, & Sears, 2005). If a student has a conceptual representation of a topic, he should be better prepared to learn from a new resource than if he has only a superficial understanding. By measuring a student's ability to learn from a new resource following instruction on a related topic, we may be able to demonstrate additional conceptual learning benefits from withholding or providing explanations.

2. Experiment 1

The first experiment, situated in middle school science classrooms, compared three conditions: a providing condition, a withholding condition, and a withholding condition that provided training in the relevant ontological category of emergent processes.[1] All three conditions received the same instructional text describing target concepts and principles from the topics of electricity and electrical circuits, and all participants were given the same basic worked examples. The primary difference between the providing and withholding conditions was whether the worked examples included instructional explanations with directions to apply them to problem solving. The only difference between the

[1] A providing with conceptual background condition, which would have made this a 2 × 2 factorial, was not included because we were specifically interested in testing whether or not prior knowledge mattered for learning in a condition hypothesized to require constructive learning activities for students (i.e., withholding instructional explanations).

two withholding conditions was that the students in the withholding with conceptual background condition read materials explaining emergent versus direct processes before beginning the training, while students in the withholding and providing conditions received a control reading. Students' achievement orientations were measured prior to the experiment, and learning was measured with a posttest targeting definition knowledge, problem solving, conceptual reasoning, and preparation for future learning.

We hypothesized that withholding instructional explanations would promote greater conceptual learning and preparation for future learning than providing instructional explanations [Hypothesis 1]. In addition, if prior knowledge of the relevant ontological categories was critical for such learning, we expected the withholding with conceptual background condition to perform better than the other two conditions on problem-solving, conceptual reasoning, and preparation for future learning tasks [Hypothesis 2].

We also explored the role of motivation in learning from these different types of instruction by using students' self-reported achievement orientations to predict learning outcomes. We hypothesized that students' dispositional achievement goals would be more predictive in the withholding conditions, where students had to connect concepts from the instructional text to the worked examples and problem-solving tasks on their own, than in the providing condition, where students were given that information [Hypothesis 3].

2.1. Method

2.1.1. Participants

Four science classes with a total of 97 middle school students from an urban, public school participated in the experiment. Fourteen students were dropped from the study because they did not complete the posttest, and another three were dropped for missing two or more of the ten learning sessions. The remaining 80 students (28 females, 52 males) were enrolled in two sixth grade classes (40 students) and two seventh grade classes (40 students), with a different teacher for each grade. Participation occurred as part of regular classroom activities, with students receiving class participation credit for completing the materials.

2.1.2. Design

The experiment had a between-subjects design and students were randomly assigned to one of three conditions: providing (28 students), withholding (25 students), or withholding with conceptual background (27 students). There were no differences in condition distribution between grades, $X^2(2, N = 80) = .08$, *ns*. Students in all three conditions participated in identical teacher-led demonstrations and received identical learning texts; conditions differed only in the worked examples and practice problems and in whether they received a background reading on emergent and direct processes or a control reading prior to beginning the lessons on electricity. Fig. 1 shows an overview of the design, materials, and procedure.

2.1.3. Materials

There were six types of materials employed in this experiment: conceptual background materials, a conceptual background manipulation check, electricity learning materials, demonstrations, test materials, and motivational assessments.

2.1.3.1. Conceptual background materials. Materials for the withholding with conceptual background condition were based on those used by Slotta and Chi (2006) and presented four examples that provided an explanation and comparison of direct and emergent processes. Descriptions of traffic jams and fish schooling demonstrated the key features of emergent processes, while descriptions of wolf packs and skyscraper construction demonstrated the key features of direct processes. To control for content, students who were not in the withholding with conceptual background condition received science articles about traffic jams, fish, wolf packs, and skyscrapers that did not address emergent or direct processes.

2.1.3.2. Conceptual background manipulation check. To measure the effect of the emergent process learning materials on students' understanding of direct and emergent processes, a test based on questions used by Slotta and Chi (2006) was given to all students. The questions on the test targeted students' understanding of the key features differentiating direct and emergent processes, as well as their ability to identify a specific process as direct or emergent. Accuracy scores were coded as a zero or one (incorrect or correct) for each question. With a total of 13 questions, scores could range from zero to 13 and are reported as percentage correct.

2.1.3.3. Electricity learning materials. Ten sets of learning packets were constructed for the 10 days of electricity learning activities. Learning packets consisted of three parts: instructional text, worked examples, and practice problems. The instructional text used in the learning packets was identical across conditions. Instructional text was taken from a middle school science textbook (Ezrailson, Zike, & Zorn, 2005, p. 138) and averaged two to three pages per day. Edits were made to eliminate ancillary text and alter the sequence of ideas when necessary to support the segmentation of lessons. The descriptions of the principles and concepts in the text were expected to help students construct explanations when studying worked examples or solving practice problems.

Worked examples demonstrating a relevant problem were provided at the end of each day's instructional text. Worked examples differed between the withholding conditions and the providing condition but were identical for the two withholding conditions. The withholding conditions received worked examples identifying solution steps for each problem, while the providing condition received the same steps paired with instructional explanations connecting the steps to the principles introduced in the preceding instructional text (Fig. 2). Consistent with Wittwer and Renkl's (2008) recommendations, the instructional explanations focused on the principles and the defining concepts in relation to other concepts.

Practice problems maintained the contrast between withholding and providing stepwise instructional explanations. Withholding practice problems were unstructured and presented only questions; providing practice problems presented questions with the same two-column solution space for students to fill in. The decision to include solution spaces for instructional explanations in practice problems was based on Wittwer and Renkl's (2008) recommendation that successful instructional explanations should include opportunities for participants to apply instructional explanations to new problems. Early piloting revealed that students in the providing condition took significantly longer to complete the learning materials than the students in the withholding conditions, so to control for time on task, students in the withholding condition were given two isomorphic problems (i.e., the same problem with new quantitative values) for every individual problem the providing condition received.

Learning materials were coded for accuracy on practice problems. Accuracy scores were coded as a zero or one (incorrect or correct) for each problem, except when the correct response involved a numerical value and a unit of measurement, which were coded separately for accuracy and resulted in scores of zero, one, or

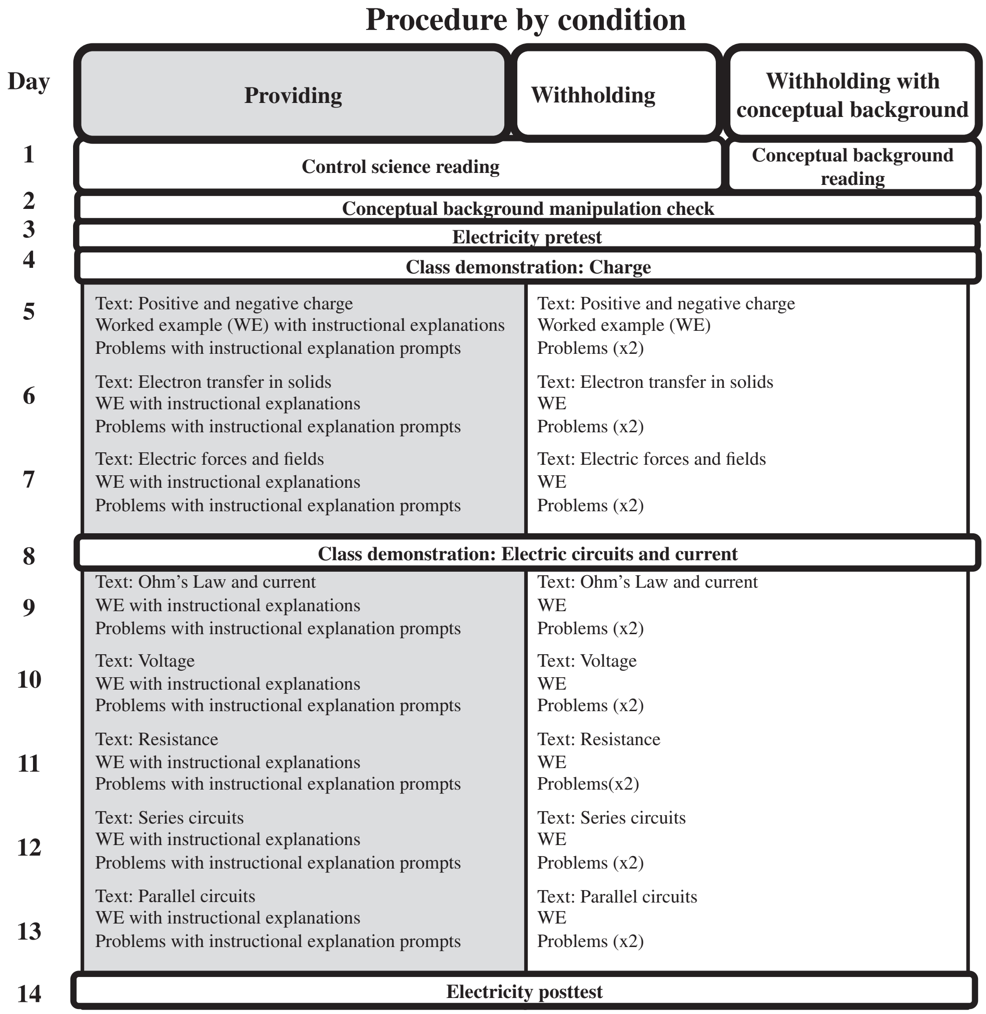

Fig. 1. List of activities and measures by condition, as indicated in top row.

two for a problem. For the providing condition, coding focused only on students' final answers, which were typically found in the bottom, right-hand box.

2.1.3.4. Demonstrations. Teachers facilitated two classroom demonstrations during the course of the intervention to provide more concrete examples of the abstract concepts described in the learning materials. All conditions participated in the same demonstrations, which were part of the curriculum-based instructional activities. Demonstrations introduced the two main topics of the intervention: electric charge and electric circuits. For the first demonstration, the teachers charged various objects through contact, such as rubbing a glass rod with a piece of silk, and held the charged objects near two small metal balls to illustrate the effects of charge. For the second demonstration, pairs of students were given two wires, a battery, and a small light bulb and instructed to find a way to light the bulb, with the instructional goal of demonstrating the importance of constructing a complete circuit. Students in both the providing and withholding conditions answered seven single-response questions related to the in-class demonstrations.

Across all intervention days, students in the withholding conditions responded to a total of 16 single-response practice problems, six value-and-unit practice problems, and seven demonstration-based questions, so scores on the learning materials could range from zero to 35. Students in the providing condition responded to a total of 8 single-response practice problems, three value-and-unit practice problems, and seven demonstration-based questions, so scores could range from zero to 21.

2.1.3.5. Test materials. An eight-question pretest and 25-question posttest on the topic of electricity were administered to measure students' learning gains. All posttest measures were coded for accuracy by two independent coders using a rubric; any differences between the coders were discussed and 100 percent agreement was reached for all questions. Two versions of the pretest contained isomorphic versions of the same questions, and these versions were counterbalanced with the first eight questions of the posttest. These isomorphic questions were included to allow for a direct comparison of performance on problems across pretest and posttest. The pretest contained three types of questions: four definition questions, two problem-solving questions, and two conceptual questions. The posttest contained four types of questions: four definition questions ($\alpha = .48$), four problem-solving questions ($\alpha = .85$), 14 conceptual questions ($\alpha = .60$), and three preparations for future learning (PFL) questions ($\alpha = .78$). PFL questions were based on an additional learning resource on power, a topic that builds on Ohm's law but was not introduced in the learning materials. The resource was provided at the end of the posttest and was followed by the three PFL questions.

Worked Example

1. A circuit has a voltage of 60 V and a resistance of 4 Ω. What is the current in the circuit?

Solution

Write Ohm's law. $I = \frac{V}{R}$
The amount of current is not given in this problem. *This is what we must solve for.*
Voltage (V) is given in this problem: V = 60 V *There are 60 volts present in this circuit.*
Resistance (R) is given in this problem: R = 4 Ω *There are 4 ohms (Ω) of resistance in this circuit.*
Current is already isolated in this equation: $I = \frac{V}{R}$
$I = \frac{V}{R} \rightarrow I = \frac{60\ V}{4\ \Omega} = 15A$ *60 volts (V) divided by 4 ohms (Ω) equals 15 amperes (A)*

Worked Example

1. A circuit has a voltage of 60 V and a resistance of 4 Ω. What is the current in the circuit?

Description of Principle	Application to this problem
General principle applied: Current (I), resistance (R), and voltage (V) are all related through Ohm's law. Ohm's law lets you determine the quantity of one value based on the other two values. Ohm's law defines the relationships between these variables.	*Write Ohm's law.* $I = \frac{V}{R}$
Define values and relations: Electric current is the flow of electric charge. If electricity is flowing in a circuit, there is a current. The higher the current, the more charge flowing.	*The amount of current is not given in this problem.* *This is what we must solve for.*
Define values and relations: In an electric circuit, voltage (V) measures the amount of electric energy provided by the energy source, which is usually a battery or a wall outlet. Voltage (V) is directly related to current (I), or electron flow, and an **increase** in voltage (V) means an **increase** in current (I).	*Voltage (V) is given in this problem: V = 60 V* *There are 60 volts present in this circuit.*
Define values and relations: Resistance (R) is the measure of how difficult it is for electrons to flow through a circuit. It is the opposition to the flow of current (I). An **increase** in resistance (R) leads to a **decrease** in the current (I).	*Resistance (R) is given in this problem: R = 4 Ω* *There are 4 ohms (Ω) of resistance in this circuit.*
Define values and relations: Current (I) increases as voltage (V) increases and decreases as resistance (R) increases. Current (I) is the ratio of voltage (V) to the resistance (R). Isolate the unknown value, current, in the equation.	*Current is already isolated in this equation:* $I = \frac{V}{R}$
Solve based on values and principle: To solve for current, we can divide voltage by resistance.	$I = \frac{V}{R} \rightarrow I = \frac{60\ V}{4\ \Omega} = 15A$ *60 volts (V) divided by 4 ohms (Ω) equals 15 amperes (A)*

Fig. 2. A worked example for the withholding conditions (left) and providing condition (right). Information provided for the withholding conditions is identical to the right-hand column for the providing condition.

The Cronbach's alpha value for the definition and conceptual questions in Experiment 1 showed relatively low internal consistency given typical interpretation guidelines. The low alpha for the definition questions (.48) was likely a result of low variability and few problems. Although the alpha level for conceptual questions was low (.60), the 95 percent confidence intervals of the interclass correlation coefficient include values in the acceptable range in the upper-bound estimate [.46–.72]. Furthermore, item-level analyses and a factor analysis revealed that the test appears to assess multiple dimensions of conceptual knowledge. For the purposes of the current work, however, we are not concerned with different dimensions of conceptual knowledge, so we treat the questions as a single conceptual measure. This approach is consistent with Schmitt (1996), who has argued that assessments with alphas between .5 and .7 can still be useful and informative, especially for new or exploratory measures of a construct.

2.1.3.5.1. Definition. Definition questions required students to provide a term or definition. These questions targeted knowledge that could be memorized from the learning packets, and they did *not* test an understanding of relationships between concepts or quantities (see Table 1a). Any questions targeting relationships, including those that could be understood as formulas, were considered conceptual in nature. Definition questions were all multiple-choice, with one point awarded for each correct response.

2.1.3.5.2. Problem solving. Problem-solving questions were the most similar to the type of problems given in the learning packets; they required knowledge of the appropriate formula, which was not given on the test, and involved quantitative calculations (see Table 1b). This measure assessed near transfer of content according to Barnett and Ceci's (2002) transfer taxonomy; the problems had a similar structure to the problems from the learning materials and required transfer of procedures and equations. Responses were scored for both the value and the units of measurement, with one point awarded for each.

2.1.3.5.3. Conceptual. Conceptual questions required an understanding of relationships between principles or the ability to make inferences based on concepts presented in the learning packets. Many of the conceptual questions also provided an opportunity for students to demonstrate misconceptions about electricity (see Table 1c). This measure assessed far transfer of content according to Barnett and Ceci's (2002) taxonomy, as the questions required the application of concepts and principles from the learning materials. Conceptual questions were a mix of short-answer and multiple-choice, and accuracy scores were coded as either a zero or one (incorrect or correct) for each question.

2.1.3.5.4. Preparation for future learning. The PFL activity presented an additional learning resource at the end of the test in the form of a one-page reading on power, which is generally introduced in texts after students have learned about Ohm's law. Two

Table 1
Samples of the four types of assessment questions.

Question type	Example
A. Definition	What kind of circuit has only one path for current to flow? a. A parallel circuit b. A series circuit c. A single circuit
B. Problem solving	A hairdryer has a current of 1.2 A and a resistance of 100 Ω. What is the voltage of the circuit?
C. Conceptual	In an electric circuit, where do the electrons come from that create the current in the circuit?
D. PFL	A toaster is plugged into a 120-V wall outlet. How much electric power does the toaster use if the current in the toaster is 10A?

conceptual and one problem-solving question about power were presented after the reading (Table 1d). We classify this measure as far transfer because it builds directly on the concepts covered in the learning materials. Accuracy scores for PFL questions were coded as zero or one for conceptual questions and as two points for each problem-solving question.

2.1.3.6. Motivation assessments. A nine-item version of Elliot and Murayama's (2008) Achievement Goal Questionnaire-Revised (AGQ-R) with questions about mastery-avoidance omitted was given to assess individual differences in students' science achievement orientations. Mastery-avoidance is considered a measure of a person's desire to avoid losing a previously attained level of mastery. Given that the middle school students in Experiment 1 had never formally studied physics and were just being introduced to topics like planetary science, earth science, and biology, we did not assess this goal because these students were unlikely to have sufficient prior knowledge to accurately assess their motivation to avoid a loss of competence. Each of the remaining three orientation subscales consisted of three items, and all three subscales were found to be reliable (mastery-approach, $\alpha = .77$, performance-approach, $\alpha = .73$, and performance-avoidance, $\alpha = .80$). Students responded to each item on a seven-point Likert scale.

2.1.4. Procedure

Several months prior to the start of the intervention, all students completed a questionnaire about their achievement orientations for their science class (AGQ-R, Elliot & Murayama, 2008). Students completed the learning materials in small sections over the course of 14 days, with activities lasting approximately 30 min each day. On the first day of the intervention, students in the withholding with conceptual background condition reviewed a packet of materials explaining the nature of emergent and direct processes for several non-electricity topics while students in the other two conditions completed science readings on the same topics that did not address those processes. As a manipulation check, all students completed a test on emergent and direct processes the next day, followed by an electricity pretest on the third day. Students then completed ten days of learning activities. They were allowed to flip back or forward in their daily learning packets, and the teachers allotted enough time for all students to complete materials to their satisfaction. Although students were encouraged to read all materials and solve all problems, they were not forced to do so. An electricity posttest was administered on day 14.

2.2. Results

Analyses focused on testing the effect of learning condition on students' posttest scores across the different question types. The conceptual background process test was analyzed for ontological training effects. Pretest scores were used for assessing overall learning gains and as a covariate in examining the effect of condition on posttest. We also conducted exploratory regression analyses of the relationship between students' achievement orientations and learning outcomes for each condition. We set the alpha level at .05 for main effects, interactions, and planned comparisons, used Bonferroni corrections for post-hoc comparisons, and report marginal effects for p values between .05 and .10 (Keppel & Wickens, 2004). We report effect sizes (Cohen's d or partial eta squared, η_p^2) for all significant main effects, interactions, and planned comparisons, and we interpret effects as small when $\eta_p^2 < .06$ or d is about .2, medium when $.06 < \eta_p^2 < .14$ or d is about .5, and large when $\eta_p^2 > .14$ or d is about .8 (see Cohen, 1988; Olejnik & Algina, 2000). According to the homogeneity-of-regression assumption, the results of an ANCOVA are not meaningful if there is a significant interaction between the covariate and the independent variable, which suggests the relationship between the covariate and the dependent variable differ across conditions (Wilson & Carry, 1969). For all ANCOVAs, we report the covariate by independent variable interaction effects to test the homogeneity-of-regression assumption.

2.2.1. Conceptual background acquisition

A one-way analysis of variance (ANOVA) showed a small effect of condition, $F(1, 78) = 4.58$, $p = .04$, $\eta^2 = .06$, with participants in the conceptual background training condition ($M = .44$, $SD = .24$) outperforming those who received the control materials ($M = .34$, $SD = .14$). This shows that the emergent process training was effective in increasing students' knowledge about emergent and direct processes compared to the control conditions, although performance remained relatively low across conditions with the training group answering more than half the questions incorrectly.

2.2.2. Pretest accuracy

Conditions were equivalent at pretest, with a one-way ANOVA revealing no differences between the providing condition ($M = .20$, $SD = .08$), the withholding condition ($M = .26$, $SD = .19$), and the withholding with conceptual background condition ($M = .25$, $SD = .15$), $F(2, 77) = .91$, $p = .41$. Additionally, there were no differences between conditions on definition questions, $F(2, 77) = 1.17$, $p = .32$, problem-solving questions, $F(2, 77) = .01$, $p = .99$, or conceptual questions, $F(2, 77) = 1.00$, $p = .37$. This shows there were no differences between conditions on any dimension of knowledge about electricity at the start of the experiment. We use the overall pretest scores as a covariate for prior knowledge when examining the effect of condition on the posttest scores below.

2.2.3. Learning materials

Responses in the learning packets varied widely, from copying worked examples verbatim to summarizing briefly to leaving large segments of boxes blank. To investigate the effect of condition on learning performance we conducted a one-way ANOVA on the proportion of practice problems answered correctly across conditions. The analysis revealed a medium effect of condition, $F(2, 77) = 4.81$, $p = .01$, $\eta^2 = .11$, with participants in the withholding ($M = .58$, $SD = .28$) and withholding with conceptual background ($M = .58$, $SD = .27$) conditions attaining higher accuracy than those in the providing condition ($M = .40$, $SD = .21$), $F(1, 51) = 7.65$, $p = .01$, $\eta^2 = .13$ and $F(1, 53) = 7.84$, $p = .01$, $\eta^2 = .13$, respectively. The two withholding conditions were not different from one another, $F(1, 50) = .00$, $p = .95$. This shows that withholding materials better supported performance during the learning phase.

A single-variable regression revealed that accuracy on learning materials was strongly predictive of overall posttest accuracy, $\beta = .84$, $t(78) = 13.65$, $p < .001$, and of all posttest measures. Looking at conditions individually, a single-variable regression revealed that accuracy on learning materials was predictive of overall posttest accuracy for the providing condition, $\beta = .81$, $t(26) = 7.14$, $p < .001$, the withholding condition, $\beta = .80$, $t(23) = 6.44$, $p < .001$, and the withholding with conceptual background condition, $\beta = .88$, $t(25) = 9.17$, $p < .001$. This shows that the learning materials in all conditions were well aligned to the posttest materials, as success on the former predicted success on the latter.

2.2.4. Posttest performance

To test whether the learning materials created an overall learning effect, a paired-samples t-test was conducted to compare students' pretest performances with their scores on the first section of the posttest, which contained isomorphic problems matching the pretest and counterbalanced to control for difficulty. The test indicated that posttest scores ($M = .42$, $SD = .26$) were higher than pretest scores ($M = .24$, $SD = .15$), $t(79) = 6.85$, $p < .001$, $d = .85$. This demonstrates that students overall showed a sizable learning gain, with the effect size indicating an improvement of more than half of a standard deviation.

To provide a stringent test of the effect of condition on posttest scores, we conducted a one-way Analysis of Covariance (ANCOVA) for each posttest using the pretest score as a covariate for prior knowledge. Given the small number of questions within each category on the pretest and the absence of differences between conditions on any subset of the pretest, the overall pretest score was used as the covariate. Preliminary analyses revealed no differences between the withholding and withholding with conceptual background conditions on any posttest measure (Fs < 1.25, ps $> .26$). Given the lack of differences between the withholding conditions on the learning and posttest measures, the two conditions were combined as a single condition for the remainder of Experiment 1. Fig. 3 shows the estimated marginal means and standard error bars for each condition, controlling for pretest performance, on all posttest measures except for the PFL data, which reflects the means without controlling for pretest.

2.2.4.1. Definition. A preliminary analysis evaluating the homogeneity-of-regression assumption indicated no interaction between condition and pretest performance, $F(3, 75) = .87$, $p = .35$. A one-way ANCOVA revealed a small effect of the covariate, $F(2, 76) = 4.99$, $p = .03$, $\eta_p^2 = .06$, showing that students' pretest scores predicted their definition posttest scores. There was no effect of condition, $F(2, 76) = .53$, $p = .47$. This shows that the two groups did not differ in their declarative knowledge of the relevant definitions at posttest.

2.2.4.2. Problem solving. A preliminary analysis evaluating the homogeneity-of-regression assumption indicated no interaction between condition and pretest performance, $F(3, 75) = 1.90$, $p = .17$. A one-way ANCOVA revealed a large effect of the covariate, $F(2, 76) = 18.13$, $p < .001$, $\eta_p^2 = .19$, showing that students' pretest scores predicted their problem-solving posttest scores. There was no effect of learning condition, $F(2, 76) = 2.55$, $p = .11$, showing that the two groups did not differ on problem-solving performance.

2.2.4.3. Conceptual. A preliminary analysis evaluating the homogeneity-of-regression assumption indicated no interaction between condition and pretest performance, $F(3, 75) = .001$, $p = .97$. A one-way ANCOVA revealed a medium effect of the covariate, $F(2, 76) = 7.15$, $p = .01$, $\eta_p^2 = .09$, showing that students' pretest scores predicted their conceptual posttest scores. There was also a marginal effect of condition, $F(2, 76) = 3.37$, $p = .07$, $\eta_p^2 = .04$, with students in the withholding conditions ($M = .36$, $SE = .02$) performing better than those in the providing condition ($M = .30$, $SE = .03$). This shows that the students in the withholding conditions gained greater conceptual understanding than those in the providing condition. We also examined the pattern of mean differences across conceptual items, and the withholding conditions had qualitatively higher mean scores than the providing condition for 12 out of 14 items. This shows a general pattern consistent across items, and not just a few items driving the effect.

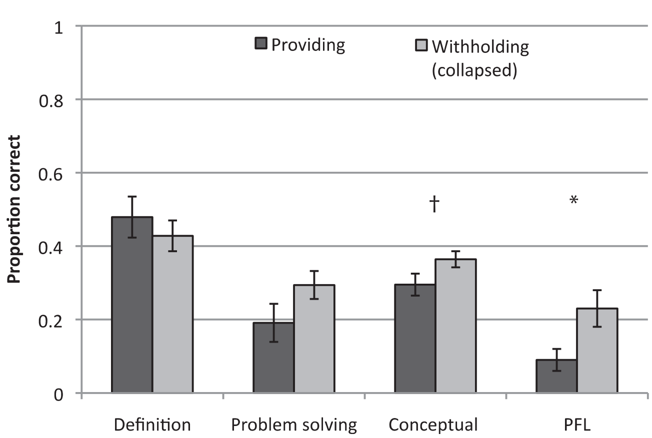

Fig. 3. Analysis of covariance results of learning condition effect on posttest accuracy for all subset measures. * indicates $p < .05$ and † denotes $p < .1$.

2.2.4.4. Preparation for future learning. A preliminary analysis evaluating the homogeneity-of-regression assumption revealed a marginal interaction between condition and pretest performance, $F(3, 75) = 2.84$, $p = .10$, $\eta_p^2 = .04$, indicating a potential violation. Although the effect is weak, it suggests the covariate interacts with the independent variable in different ways across conditions, rendering an ANCOVA unfit for interpretation. Consequently, we evaluated the effect of condition on preparation for future learning using a standard ANOVA without controlling for the covariate. A one-way ANOVA revealed a significant effect of condition, $F(1, 77) = 4.26$, $p = .04$, $\eta_p^2 = .05$, with students in the withholding conditions ($M = .23$, $SE = .05$) performing better than those in the providing condition ($M = .09$, $SE = .03$). This indicates the students in the withholding conditions were better prepared to learn about power than those in the providing condition.

2.2.5. Achievement orientations

A one-way ANOVA revealed a marginal difference between conditions in mastery-approach, $F(1, 69) = 3.48$, $p = .07$, $\eta_p^2 = .05$, with the withholding condition ($M = 6.10$, $SD = 1.29$) reporting higher mastery-approach goals than the providing condition ($M = 5.56$, $SD = .94$). There were no differences between conditions on performance-approach goals, $F(1, 69) = .11$, $p = .74$, or performance-avoidance goals, $F(1, 69) = .17$, $p = .69$. To investigate whether achievement orientations played a different role in determining learning outcomes for participants in the providing versus withholding conditions, a series of hierarchical regression analyses for each orientation were conducted to predict posttest measures separately for each learning condition. We controlled for pretest scores as a proxy for prior knowledge to see whether achievement orientations predict any variance above and beyond that predicted by the pretest scores alone. Pretest score was entered as the first-level predictor to partial out variance due to prior knowledge, and the achievement orientation in question was entered as the second-level predictor.

We are primarily concerned with testing the role of motivation in learning outcomes for which our manipulation had an effect, which for this experiment are the conceptual and PFL measures. For both the providing and withholding conditions, no achievement orientation predicted performance on the overall, conceptual, or PFL measures (βs $< .28$ and $> -.15$, ps $> .14$). These results suggest achievement orientations are unrelated to learning outcomes in this context when controlling for prior knowledge. Another possibility is that the achievement orientations measure may have been administered too far in advance of the intervention to capture the dynamic interplay between the orientations students reported experiencing in science class and the effects of the learning conditions.

2.3. Discussion

The results of Experiment 1 provide preliminary evidence supporting the hypothesis that withholding instructional explanations positively affects learning outcomes [Hypothesis 1]. The withholding conditions performed better than the providing condition on the conceptual and PFL assessments, consistent with the hypothesis that withholding instructional explanations provides learners with an opportunity to engage in constructive learning activities to facilitate deeper learning of the concepts and far transfer. It is also possible that students in the providing condition may have suppressed self-explanation because the explanations were already given in the materials.

Overall performance across the posttest measures was relatively low. There are several possible explanations for this general result. Students lacked prior exposure to topics in electricity, making it difficult to fully grasp the conceptual space with only brief, self-guided learning materials administered over a short period of time. Furthermore, the learning activities required high self-regulation compared to typical classroom instruction, which may have contributed to poor outcomes. In an experiment that spanned approximately three weeks, failure to pay close attention on even a few days could disrupt students' overall learning of the materials. In particular, the providing condition required students to write out extensive explanations, which could have prompted students to disengage. A more scaffolded set of providing materials may reduce the burden on students in the providing condition to remain attentive while completing long sequences of writing. Testing the effects of providing and withholding instructional explanations in worked examples with a population that has previous exposure to these topics, yet still demonstrates inaccuracies in knowledge about electricity, may yield clearer results. Additionally, the design of this intervention may be more effective in a population that is better accustomed to self-guided learning. Therefore, in Experiments 2 and 3 we used a college population that had prior exposure to physics concepts in high school.

The fact that the two withholding conditions did not differ from each other on the posttests shows that prior conceptual knowledge had a minimal effect on the learning outcomes from withholding instructional explanations [Hypothesis 2]. However, this null effect may be due to the relatively weak training effects observed for the students given the conceptual background training. Although these students performed better than the non-training groups on the conceptual background test, they still performed with less than 50 percent accuracy. This result shows that their overall understanding of emergent processes remained poor even after training. One possible explanation for the minimal impact of our training on the conceptual background measure was the abbreviated duration of our training regimen compared to past work. Slotta and Chi (2006) noted that the materials were extremely difficult when deployed as a 2-h, computer-based training program for undergraduate college students; given classroom constraints we were limited to approximately 1 h of training meaning the materials were reduced by more than 50 percent. For such training materials to be effective, middle school students may require much more instruction and practice.

In contrast to our hypotheses, we found no relationships between achievement orientations, type of instruction, and learning [Hypothesis 3]. We failed to replicate past findings connecting achievement orientations with learning outcomes, which may be a result of the achievement orientation measure having been administered several months before the intervention. Past achievement orientation research has suggested that students' orientations decline over the course of the school year (Chouinard & Roy, 2008). Several months, therefore, might be enough time for students' goals to change at different rates or in different ways, thus reducing the overall predictive power of achievement orientations for learning outcomes. In the next two experiments we measure achievement orientations and learning outcomes across a much shorter timeframe.

Another possibility is that the learning activities were driven less by an effect of general achievement orientations and more by task goals promoted by different learning materials. Specifically, it may be that the providing condition suppressed mastery goals during task completion, which in turn reduced learning. This suggests there may be merit in measuring students' task goals as well as dispositions (for a similar approach see Belenky & Nokes, 2010; Belenky & Nokes-Malach, 2012). Given the research showing that a task can promote specific goals based on its framing and evaluative measures (Ames, 1992; Belenky & Nokes-Malach, 2012), Experiments 2 and 3 measure task goals immediately after completion of the learning materials in addition to general achievement orientations for science.

3. Experiment 2

For Experiment 2 we dropped the withholding with conceptual background training condition because Experiment 1 showed no differences between the two withholding conditions on any posttest measures. Based on Slotta and Chi's (2006) prior work and our results from Experiment 1, we concluded that to create a robust effect, the conceptual background training would require a more elaborate training regimen beyond the scope of the current work. Therefore, in this experiment we focused on comparing the withholding and providing conditions using the same topic but with college students. Students completed all materials during a 2.25-h laboratory session. In contrast to the first experiment, students in the providing condition were also given the instructional explanations (i.e., the principles) for each step of the practice problems, so their task was to fill in the solution steps and final answer for each problem. This change was intended to reduce the cognitive load of students in the providing condition. In addition to measuring students' achievement orientations for science class, we measured students' task goals immediately following their completion of the learning materials. The posttest consisted of the same four types of assessments.

As with Experiment 1, we expected students in the withholding condition to show greater conceptual learning than those in the providing condition. We also expected students in the withholding condition to perform better on the PFL questions. We explored the relationship between achievement goals and learning outcomes by measuring both general achievement orientations and specific task goals that might be formed as a result of the materials. We expected general achievement orientations to predict learning outcomes more strongly for students in the withholding condition than in the providing condition because they must rely on their individual motivations to engage in constructive cognitive processing [Hypothesis 2]. For students in the withholding condition, we expected achievement orientations and task goals to be highly correlated, as the relatively sparse scaffolding provided to students in the withholding condition was expected to have a minimal effect on the goals the students formed regarding the task at hand. In the providing condition, however, task goals were expected to be less highly correlated with general orientations, as the instructional scaffolding might reduce students' reliance on their general science orientations and possibly suppress mastery orientations in relation to the learning task [Hypothesis 4].

3.1. Method

3.1.1. Participants

Eighty-four college students (57 females, 27 males) enrolled in an introductory psychology course at the University of Pittsburgh participated in the study. Sixty-four students were freshmen, twelve were sophomores, two were juniors, three were seniors, and three were not traditional undergraduates. Participants received three credits toward a research participation requirement associated with the course. On a survey administered at the end of the experiment, 87 percent of students reported having taken a physics course in high school.

3.1.2. Design

The experiment was a between-subject design, with participants randomly assigned to one of two conditions: a providing or a withholding condition, with 42 participants in each condition. There were no differences in condition distribution across class years, $X^2(4, N = 84) = 2.25$, *ns*. The experiment consisted of a single session lasting approximately 2.25 h, and a maximum of four participants were allowed to participate in a session by working independently at separate workstations.

3.1.3. Materials

There were three types of materials employed in this experiment: electricity learning materials, test materials, and motivational assessments.

3.1.3.1. Learning materials. Learning materials were largely the same as those used in Experiment 1, with some alterations. Materials were combined into two parts instead of 10, although they still followed the pattern of multiple sequences of instructional text, worked examples, and practice problems. Minor additions from a more technical middle school textbook (Hsu, 2005) were used to replace some of the simpler language used in Experiment 1, and some information on basic topics was omitted to accommodate time constraints.

In Experiment 1, the providing condition was less accurate than the withholding conditions on the practice problems completed during the learning phase. We hypothesized that this occurred because the participants in the providing condition had to write out the stepwise instructional explanations as well as the solution steps, thereby overloading working memory and leading to poor problem-solving performance. To address this possibility, we modified the practice problems for the providing condition so they included the same instructional explanations as those from the worked example (see the left-hand column of the providing condition's worked example in Fig. 2, Section 2.1.3.3, for an example of the stepwise instructional explanations). This modification meant that participants did not have to write out the explanation for each step of the problem, but instead had them immediately available while attempting to solve the problem by filling in the stepwise problem solutions in the right-hand column. Practice problems for the withholding condition were unchanged from Experiment 1.

Similar to Experiment 1, pilot testing revealed that students in the providing condition took longer to complete the material, even though instructional explanations were provided in the practice problems to reduce writing time and cognitive load. Therefore, to control for time on task, students in the withholding condition were given a pair of isomorphic problems for every individual problem the providing condition received.

Learning materials were coded for accuracy on practice problems. Accuracy scores were coded as a zero or one (incorrect or correct) for each problem, except when the correct response involved a numerical value and a unit of measurement, which were coded separately for accuracy and resulted in scores of zero, one, or two. Participants in the withholding conditions responded to a total of eight single-response problems and four value-and-unit problems, so scores could range from zero to 16. Students in the providing condition responded to a total of four single-response problems and two value-and-unit problems, so scores could range from zero to eight.

3.1.3.2. Test materials. Test materials were also largely the same as those used in Experiment 1, with a few alterations. New conceptual questions were added to increase the difficulty of the test and create more opportunities for participants to demonstrate different levels of understanding, and multiple-choice options were removed for all definition questions. An 11-question pretest and 30-question posttest on the topic of electricity were administered to measure students' learning. As in Experiment 1, two versions of the pretest contained isomorphic versions of the same questions, and these versions were counterbalanced with the first 11 questions of the posttest to control for pre- and posttest difficulty. The pretest contained three problem-solving questions, four definition questions, and four conceptual questions. The posttest contained four definition items ($\alpha = .26$), five problem-solving items ($\alpha = .54$), 16 conceptual items ($\alpha = .63$), and five PFL items ($\alpha = .52$). The low alpha for definition items resulted from low variance in performance, as participants were near ceiling on two of the four items. As in Experiment 1, PFL questions required the use of an additional learning resource provided to all participants at the end of the test in the form of a one-page reading on power. All posttest measures were coded for accuracy by two independent coders using a rubric. Any differences between the coders were discussed and 100 percent agreement was reached for all questions.

3.1.3.3. Achievement orientation and task goal questionnaires. A 12-item version of Elliot and Murayama's (2008) Achievement Goal Questionnaire-Revised (AGQ-R) was given to assess individual differences in participants' achievement orientations. All four subscales consisted of three items and were found to be reliable (mastery-approach, $\alpha = .80$, mastery-avoidance, $\alpha = .80$, performance-approach, $\alpha = .83$, and performance-avoidance, $\alpha = .83$). We included mastery-avoidance questions for this experiment because mastery-avoidance is considered a measure of a person's desire to maintain a previously attained level of competence or to avoid a loss of competence; as the majority of students were expected to have some prior physics exposure, the construct was considered relevant in Experiments 2 and 3 where it was not relevant in Experiment 1.

A nine-item task-based version of the AGQ-R was created to frame statements around the learning task instead of general views toward science. For example, while a performance-approach question on the general AGQ-R reads, "*I strive to do well compared to other students*," the item was modified on the task goal questionnaire to read, "*During this activity, I was striving to do well compared to other people who complete this activity*." All three task goal subscales consisted of three items and were found to be reliable (mastery-approach, $\alpha = .85$, performance-approach, $\alpha = .93$, and performance-avoidance, $\alpha = .98$). These measures were expected to indicate the state of students' goals as they worked through the learning materials, as opposed to the goals they generally experienced in relation to their science classes.

Mastery-avoidance was omitted from the task questionnaire because students were not expected to have sufficient experience with these particular activities to be able to accurately judge whether or not they could learn less than they possibly could. To our knowledge there is no work examining the amount of prior knowledge necessary to facilitate learners' adoption of mastery-avoidance goals, suggesting that this might be a fruitful topic for future work. Although we expected that students would be familiar with the general physics concepts, we did not expect them to be familiar with our particular learning activities or materials and therefore expected them to have difficulty determining whether they had previously mastered such materials. Additionally, the task framing was considered incongruent with the language used for mastery-avoidance items.

3.1.4. Procedure

Participants first completed an electricity pretest before proceeding to two self-paced learning booklets containing a series of instructional texts, worked examples, and practice problems. Participants could flip back or ahead within each learning booklet, but they could not go back to the first booklet after moving on to the second. Students were given 45 min to complete each booklet; only four students (two in each condition) required the entire time to complete the first booklet, and all students completed the second booklet in less than 45 min. Upon completing the learning booklets, participants responded to an activity goals questionnaire. They then completed a two-part posttest, followed by an achievement orientations questionnaire and a demographic questionnaire.

3.2. Results

As with Experiment 1, analyses focused on testing the condition effect across the different posttest measures. We used pretest scores as a measure of prior knowledge and a covariate when assessing the effect of the learning condition on posttest. In addition, we explored the relationship between students' achievement orientations and task goals in relation to the learning outcomes.

3.2.1. Pretest accuracy

Conditions were equivalent at pretest, with a one-way ANOVA revealing no differences between providing ($M = .58$, $SD = .20$) and withholding ($M = .59$, $SD = .24$), $F(1, 82) = .06$, $p = .81$. There were also no differences on the definition questions, $F(1, 82) = .42$, $p = .52$, problem-solving questions, $F(1, 82) = .03$, $p = .86$, or conceptual questions, $F(1, 82) = 2.37$, $p = .13$. This shows there were no differences between conditions in any component of knowledge about electricity at the start of the experiment.

3.2.2. Learning materials

Participants were generally very thorough in completing learning materials. Students in the providing condition tended to complete all boxes, with responses often paraphrasing the information in the worked examples. Responses from students in the withholding condition varied more widely, with some simply writing the answer and others more closely copying the steps contained in the worked examples or providing brief explanations of their own. Participants in both the providing condition ($M = .85$, $SD = .13$) and the withholding condition ($M = .88$, $SD = .09$) showed very high accuracy in their performance on the learning materials, and a one-way ANOVA comparing accuracy between learning conditions found no difference, $F(1, 82) = 2.29$, $p = .13$. These results differ from Experiment 1, suggesting that any detrimental effects of condition format on learning materials accuracy in Experiment 1 (e.g., perhaps the format of providing materials distracted students from critical content) was eliminated in Experiment 2, either by the addition of one completed column or by the recruitment of older participants.

A single-variable regression revealed that accuracy on learning materials was strongly predictive of overall posttest accuracy, $\beta = .43$, $t(82) = 4.37$, $p < .001$, and of all posttest measures with the exception of the definition questions, for which it was marginal, $\beta = .20$, $t(82) = 1.80$, $p = .08$. This shows that completing materials accurately was associated with better posttest performance.

3.2.3. Posttest performance

To test whether the learning materials created an overall learning effect, a paired-samples *t*-test was conducted to compare students' pretest performances with their scores on the first section of the posttest, which contained isomorphic problems matching the pretest and counterbalanced to control for difficulty. The test indicated that posttest scores ($M = .82$, $SD = .13$) were higher than pretest scores ($M = .58$, $SD = .22$), $t(83) = 11.81$, $p < .001$, $d = 1.44$. This demonstrates that participants showed a sizable learning gain, with the effect size indicating an improvement of more than a standard deviation. As with Experiment 1, pretest scores were used as a covariate in all posttest analyses to control for any differences in prior knowledge at the outset of the experiment. Fig. 4 shows the estimated marginal means and standard error bars for each condition across each posttest measure.

3.2.3.1. Definition. A preliminary analysis evaluating the homogeneity-of-regression assumption indicated no interaction between condition and pretest performance, $F(3, 80) = .74$, $p = .39$. A one-way ANCOVA revealed a revealed a medium effect of the covariate, $F(2, 81) = 11.84$, $p = .001$, $\eta_p^2 = .13$, showing that students' pretest scores predicted their definition posttest scores. There was no effect of condition on definition scores, $F(2, 81) = .26$, $p = .61$.

3.2.3.2. Problem solving. A preliminary analysis evaluating the homogeneity-of-regression assumption indicated no interaction between condition and pretest performance, $F(3, 80) = .05$, $p = .82$. A one-way ANCOVA revealed a medium effect of the covariate, $F(2, 81) = 8.60$, $p = .004$, $\eta_p^2 = .10$, showing that students' pretest scores predicted their problem-solving posttest scores. There was no effect of condition on problem-solving questions, $F(2, 81) = .35$, $p = .55$.

3.2.3.3. Conceptual. A preliminary analysis evaluating the homogeneity-of-regression assumption indicated no interaction between condition and pretest performance, $F(3, 80) = 1.39$, $p = .24$. A one-way ANCOVA revealed a large effect of the covariate, $F(2, 81) = 27.94$, $p < .001$, $\eta_p^2 = .26$, showing that students' pretest scores predicted their conceptual posttest scores. There was also a medium effect of condition on conceptual questions, $F(2, 81) = 5.60$, $p = .02$, $\eta_p^2 = .07$, with the withholding condition ($M = .57$, $SE = .02$) attaining a higher score than the providing condition ($M = .51$, $SE = .02$). This is consistent with the prediction that withholding explanations would promote greater conceptual learning than providing explanations.

3.2.3.4. Preparation for future learning. A preliminary analysis evaluating the homogeneity-of-regression assumption indicated no interaction between condition and pretest performance, $F(3, 80) = 1.35$, $p = .25$. A one-way ANCOVA revealed a large effect of the covariate, $F(2, 81) = 47.67$, $p < .001$, $\eta_p^2 = .37$, showing that students' pretest scores predicted their PFL posttest scores. There was no effect of condition on PFL questions, $F(2, 81) = .03$, $p = .87$.

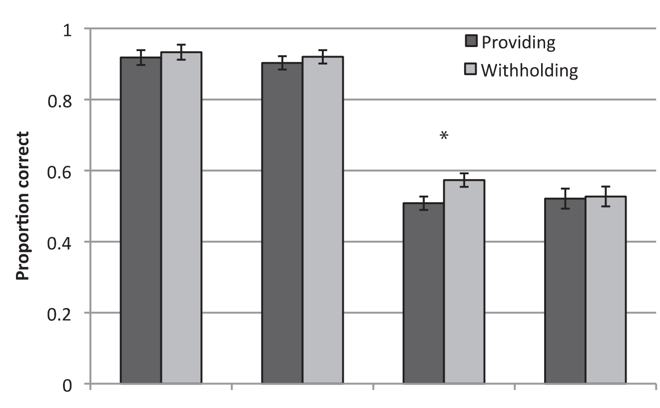

Fig. 4. Learning condition effect on posttest accuracy. * denotes $p < .05$.

3.2.4. Achievement orientations

A one-way ANOVA found no differences across conditions in mastery-approach, $F(1, 82) = 2.23$, $p = .14$, mastery-avoidance, $F(1, 82) = 2.67$, $p = .11$, performance-approach, $F(1, 82) = .43$, $p = .51$, or performance-avoidance, $F(1, 82) = 1.01$, $p = .32$. This was consistent with predictions, as we did not expect the manipulation to change participants' general achievement orientations for science class. Next, we investigated differences in the role of achievement orientations in determining learning outcomes for participants in the providing and withholding conditions. We conducted a series of hierarchical regression analyses that used each orientation and controlled for pretest performance to predict posttest measures separately for each learning condition. As with Experiment 1, we focus on the overall posttest accuracy and any specific posttest measures that showed a condition effect, which for this experiment is the conceptual measure. For the providing condition, no achievement orientations were predictive of overall or conceptual learning outcomes (all βs $< .11$ and $> -.11$, ps $> .45$). In contrast, for the withholding condition, mastery-avoidance was predictive of overall accuracy and marginally predictive of conceptual accuracy, $\beta = .35$, $t(38) = 2.87$, $p = .01$, and $\beta = .25$, $t(38) = 1.93$, $p = .06$, respectively. No other achievement orientations were predictive of these learning measures (all βs $< .14$, ps $> .28$).

These results are consistent with predictions that achievement orientations would predict posttest performance for participants receiving withholding materials but not those who received providing materials. Furthermore, they show that mastery-avoidance orientation predicts learning as students strive to avoid losing previous mastery of a topic. This analysis provides a strong test of the effect of achievement orientation on learning because it shows that orientation scores predict variation in learning outcomes above and beyond that accounted for by prior knowledge as measured by the pretest.

3.2.5. Task goals

A one-way ANOVA found a medium condition effect on mastery-approach task goals, $F(1, 82) = 5.89$, $p = .02$, $\eta^2 = .07$, with the withholding condition ($M = 5.85$, $SD = 1.04$) reporting higher mastery-approach task goals than the providing condition ($M = 5.28$, $SD = 1.10$). This is consistent with predictions that the providing condition would suppress mastery-approach task goals compared to the withholding condition. There were no differences across conditions in performance-approach, $F(1, 82) = 1.08$, $p = .30$, or performance-avoidance, $F(1, 82) = .16$, $p = .69$.

To investigate differences in how well task goals predicted learning outcomes across conditions, a series of hierarchical regression analyses that used each goal and controlled for pretest performance were conducted to predict posttest measures separately for each learning condition. Again, we focus on overall accuracy and the posttest measures that showed a significant learning condition effect. For the providing condition, task performance-approach marginally predicted overall learning, $\beta = .23$, $t(39) = 1.67$, $p = .10$, and conceptual learning, $\beta = .26$, $t(39) = 1.84$, $p = .07$. No other task goals were predictive of overall or conceptual learning outcomes for the providing condition (all βs $< .23$, ps $> .11$). For the withholding condition, no task goals were predictive of overall or conceptual learning outcomes (all βs $< .17$, ps $> .20$).

3.2.6. Task goals and achievement orientations

Although task goals and achievement orientations were predictive of different posttest outcomes, the two measures were correlated. Mastery-approach task goals and achievement orientations were correlated, $r = .38$, $p < .001$, as were performance-approach task goals and achievement orientations, $r = .53$, $p < .001$, and performance-avoidance task goals and achievement orientations, $r = .46$, $p < .001$.

A condition effect on correlations of mastery-approach emerged, with mastery-approach task goals and achievement orientations correlated in the withholding condition, $r = .53$, $p < .001$, but not in the providing condition, $r = .22$, $p = .16$. This suggests participants' mastery-approach task goals in the withholding condition were consistent with their levels of mastery-approach achievement orientations for science. In the providing condition, it seems participants' existing mastery-approach achievement orientations were not expressed in the form of task goals, suggesting that the providing materials may have altered these goals for the task at hand. Taken together with the main effect of learning condition on task goals, it may be that providing materials disrupted students' achievement goals expressed as task mastery.

3.3. Discussion

Similar to Experiment 1, the withholding condition demonstrated greater performance on the conceptual reasoning far-transfer measure [Hypothesis 1]. There are several plausible explanations for this result. Consistent with our hypothesis, one possibility is that the lack of instructional explanations prompted participants in the withholding condition to generate self-explanations, thus engaging in deeper processing and more constructive activities than those in the providing condition. Although the intervention provided no explicit prompts to self-explain, some students may have self-explained on their own. Specifically, without instructional explanations, they may have spent more time coordinating information in the instructional text and worked examples, filling in gaps, and abstracting that information to apply it to the practice problems. Participants in the providing condition, who received instructional explanations of each problem-solving step, may have suppressed their tendencies to self-explain because explicit explanations were readily available.

Chi (2009) proposed an instructional framework that distinguishes between *active* learning activities that involve physically doing something and *constructive* learning activities that involve producing new ideas and outputs that go beyond the information provided in the instructional materials. According to this framework, *active* learning activities such as looking, gazing, pointing, summarizing, or copying are associated with cognitive processes such as searching and activating existing knowledge and encoding, assimilating, and storing new knowledge. *Constructive* activities such as explaining, connecting, reflecting, and predicting are associated with cognitive processes such as reorganizing and repairing existing knowledge, inferring new knowledge, and integrating new ideas into existing knowledge. To aid in interpreting our results, we apply this framework to the activities used in each instructional condition.

By receiving instructional explanations in the learning materials, students are likely to have engaged in "active" learning activities such as reading, paraphrasing, or summarizing those explanations during problem-solving practice while copying the solution steps from the worked examples. If students in the providing condition engaged in these types of activities, we would expect them to acquire knowledge of the concept definitions and problem-solving procedures but not deeper conceptual knowledge. This knowledge would then facilitate performance on the definition questions and the near-transfer problems but not on the far-transfer assessments. In contrast, in the withholding condition students had to figure out how to solve the practice problems without the aid of instructional explanations. This may have led to more constructive activities such as self-explanation because students had to understand the worked examples in order to solve the practice problems. By self-explaining the worked examples, students could generate inferences connecting the principles and concepts embedded in the worked examples to the practice problem features. This knowledge could then facilitate performance on the definition questions, problem-solving tasks, and conceptual assessments on the test.

Furthermore, if withholding instructional explanations places a larger burden on the student to construct explanations in order to solve the practice problems, then we would also expect that a particular student's motivation for learning and achievement would affect how successful she would be in generating those explanations. Consistent with this hypothesis, we found that students' achievement orientations predicted learning outcomes for the withholding condition but not the providing condition. Specifically, the more that students in the withholding condition endorsed mastery-avoidance goals, the more likely they were to perform well on the overall test and conceptual questions. Those who did not strongly endorse those goals were less likely to be successful. In contrast, students' achievement orientations in the providing condition were not related to their learning outcomes. These results provide some of the first evidence that we know of showing that mastery-avoidance goals are positively related to far-transfer outcomes for particular types of instruction.

Another possible explanation for the withholding condition advantage is that students in this condition received twice as many practice problems as the students in the providing condition to control for time. These additional practice problems were isomorphic to the shared problems and required no structural modifications to the equations; instead, they involved a substitution of different values and contained no new conceptual information. It is nevertheless possible that the difference in conceptual understanding resulted from more practice. The isomorphic problems may have provided an additional opportunity for conceptual insights and inference generation, but they also provided additional rote practice. If more practice on problem-solving activities were responsible for the conceptual posttest differences, we would expect to see as strong or stronger differences in accuracy on the problem-solving questions, as these questions more closely resembled the practice problems (i.e., near transfer). For the problem-solving measure, however, there was no effect of condition on accuracy.

Although comparing multiple examples or problems can promote conceptual learning (Gentner, Loewenstein, & Thompson, 2003; Kurtz, Miao, & Gentner, 2001; Nokes & Ross, 2007), prior work has shown that students are unlikely to *spontaneously* compare problems that are presented sequentially without explicit instructions and scaffolding to do so (Alfieri, Nokes-Malach, & Schunn, in press; Gentner et al., 2003; Gick & Holyoak, 1983; Rittle-Johnson & Star, 2007). As our isomorphic practice problems were presented sequentially with no instructions to compare, we expected little conceptual learning from solving additional problems alone. In Experiment 3, we address this potential confound by providing students in both conditions the same number of practice problems (Section 4.1).

Finally, a third possibility is that the difference in task mastery-approach goals between conditions may have promoted a difference. Evidence that achievement and task mastery-approach were highly correlated in the withholding condition but not in the providing condition suggests that participants in the providing condition may have viewed the task as a matter of rote completion and thus suppressed any mastery-approach orientation (i.e., their desire to learn and understand), while participants in the withholding condition may have maintained their general mastery-approach orientation when forming specific task-related goals. This is consistent with the fact that students in the providing condition reported lower task mastery-approach goals than students in the withholding condition [Hypothesis 4].

Interestingly, the performance-approach task goals were marginally predictive of one learning outcome measure for the providing condition but none for the withholding condition. As our task goal measure was exploratory, we did not have strong predictions about how the role of task goals would compare between conditions; however, we believe these results are generally consistent with our hypotheses. Participants in the withholding condition rely on their dispositional achievement orientations (mastery-avoidance) more than participants in the providing condition. In the presence of the additional support provided by the instructional explanations, participants in the providing condition may rely more on their task goals, which in turn play a more direct role in determining their learning outcomes. We regard these results as very preliminary and test these relationships again in Experiment 3.

As expected, both the effects of learning conditions and the predictive power of achievement orientations were stronger in Experiment 2 than in Experiment 1. This is consistent with expectations for an experiment conducted in a more controlled, laboratory environment compared to a classroom. It lends support to the possibility that worked examples of this nature push the limits of middle school students and are more appropriate for college students, who generally have more experience working through materials on their own. It also suggests that worked examples that withhold instructional explanations are most effective when students have some prior exposure to the topic, which might equip them with more knowledge to use when generating self-explanations and attempting to identify and address gaps in their understanding. Additionally, it may be possible that the effects of the experimental manipulation in Experiment 1 were muted by the additional factors supporting learning in a classroom environment, such as student discussions and teacher demonstrations.

4. Experiment 3

In Experiment 3, we tested the effect of providing or withholding instructional explanations while controlling for the number of practice problems students solved. We also sought to more closely examine the hypothesis that the providing condition might reduce learning by suppressing self-explanation tendencies; this was done by creating a second providing condition that prompted participants to generate statements of the principles relevant to each problem while the problem steps were provided. Wittwer and Renkl (2008) argue that instructional explanations may be effectively included in worked examples so long as learners are also given the opportunity to engage actively with the explanations, for example, by applying them to a task. Through the creation of our third condition, we systematically manipulate whether the providing condition also receives instructional explanations with practice problems or whether participants are prompted to generate them on their own.

It may be that the reduced learning effects experienced by students in the providing condition in Experiment 2 were not a result of receiving instructional explanations in the worked examples, but rather a result of not having the opportunity to apply those explanations to problems. In other words, by providing a restatement of principles in the practice problems and having participants generate problem solutions, the practice problems might be emphasizing the wrong practice opportunities. Participants in Experiment 1 were responsible for generating both the instructional explanations and solutions for practice problems; however, we hypothesized that asking participants to do both parts may have created too much cognitive load and resulted in worse learning outcomes. The providing-solutions condition in Experiment 3 is more consistent with Wittwer and Renkl's (2008) criteria for effective instructional explanations because it provides opportunities for learners to generate the instructional explanations portion of the practice problems, while still minimizing cognitive load by providing the solutions for the participants.

4.1. Methods

4.1.1. Participants

Ninety-two college students (48 females, 44 males) enrolled in an introductory psychology course at the University of Pittsburgh participated in the study. Sixty-four students were freshmen, 16 were sophomores, six were juniors, three were seniors, and two were not traditional undergraduates. Participants received two credits toward a research participation requirement associated with the course. On a survey administered at the end of the experiment, 81 percent of students reported having taken a physics course in high school.

4.1.2. Design

The experiment was a between-subject design, with participants randomly assigned to one of three conditions: a providing-explanations condition (30 participants), a providing-solutions condition (31 participants), or a withholding condition (31 participants). There were no differences in condition distribution across class years, X^2 (8, N = 91) = 7.71, *ns.*[2] The withholding and providing-explanations conditions were identical to those in Experiment 2, while the providing-solutions condition was new. The experiment consisted of a single session lasting approximately 2 h, and a maximum of five participants were allowed to participate in a session by working independently at separate workstations.

4.1.3. Materials

There were three types of materials employed in this experiment: electricity learning materials, test materials, and motivational assessments.

4.1.3.1. Learning materials. Learning materials were nearly identical in content to those used in Experiment 2; however, the booklet was not broken into two parts and did not include an opportunity for participants to take a break midway. This was done for logistical purposes as few participants opted to use the break during Experiment 2, and the elimination of the break and consolidation of the booklet reduced the total time needed for the experiment to 2 h.

A new learning booklet, created for the providing-solutions condition, was identical to the existing providing-explanations materials except for one change. Both providing conditions still contained instructional explanations for each worked example; however, to address the possibility that participants were ignoring the provided instructional explanations when solving the practice problems, the problems in the providing-solutions condition were modified to include the right-hand solution steps, leaving participants to fill in the left-hand instructional explanations (see the right-hand column of the providing condition's worked example in Fig. 2, Section 2.1.3.3, for an example of the solution steps). This condition complemented the providing-explanations condition, which included the left-hand instructional explanations on practice problems and instructed participants to complete the right-hand solution steps. To address the potential confound of students in the withholding condition receiving additional isomorphic problems, participants across all conditions received the same number of identical problems.

Learning materials for the providing-explanations and withholding conditions were coded for accuracy on practice problems. Accuracy scores were coded as a zero or one (incorrect or correct) for each problem, except when the correct response involved a numerical value and a unit of measurement, which were coded separately for accuracy and resulted in scores of zero, one, or two. Participants in all conditions responded to a total of four single-response problems and two value-and-unit problems, so scores could range from zero to eight.

4.1.3.2. Test materials. Test materials were very similar to those used in Experiment 2, with several additions and modifications to better capture different elements of participants' learning. A 10-question pretest and 30-question posttest on the topic of electricity were administered to measure students' learning. Two conceptual questions from Experiment 2 were eliminated because a large number of the participants provided correct information about related concepts but failed to address the intended questions. The questions were replaced with two new questions, one conceptual and one PFL. Two versions of the pretest contained isomorphic versions of the same questions, and these versions were counterbalanced with the first 10 questions of the posttest to control for pre- and posttest difficulty. The pretest contained three problem-solving questions, four definition questions, and three conceptual questions. The posttest contained four definition items ($\alpha = .27$), five problem-solving items ($\alpha = .78$), 15 conceptual items ($\alpha = .74$), and six PFL items ($\alpha = .69$). Coding for all problem types was identical to Experiments 1 and 2. As with Experiment 2, the low alpha for definition items resulted from low variance in performance, as participants were near ceiling on one of the four items. As in both prior experiments, PFL questions required the use of an additional learning resource provided to all participants at the end of the test in the form of a two-page reading on power.

4.1.3.3. Achievement orientation and task goal questionnaires. A 12-item version of Elliot and Murayama's (2008) Achievement Goal Questionnaire-Revised (AGQ-R) was given to assess individual differences in participants' achievement orientations. Three of the four three-item subscales from the achievement orientations questionnaire were found to be reliable, including mastery-approach (α = .82), performance-approach (α = .79), and performance-avoidance ($\alpha = .85$). Although the reliability for the mastery-avoidance subscale was somewhat low (α = .52), we report results for this construct because it has been well validated in past work (Elliot & Murayama, 2008). The same nine-item task-based version of the AGQ-R, with questions about mastery-avoidance omitted, was used from Experiment 2. All of the three-item subscales were found to be reliable, including mastery-approach (α = .74), performance-approach (α = .79), and performance-avoidance ($\alpha = .95$). Both questionnaires were identical to the ones used in Experiment 2.

4.1.4. Procedure

Prior to the start of the experiment, participants completed an achievement orientations questionnaire framed around science class. The questionnaire was moved to the beginning of the experiment to eliminate the possibility that responses on the general orientations questionnaire might be affected by condition assignment. Participants then completed an electricity pretest before proceeding to a single self-paced learning booklet containing a series of instructional texts, worked examples, and practice

[2] One participant in the withholding condition did not complete a demographic questionnaire and is therefore excluded from the summary of participants' years in school and the analysis of condition assignment across years.

problems. They could flip back or ahead within the learning booklet, and they were given 55 min to complete the entire booklet. Overall time was reduced because most participants finished early in Experiment 2. Twelve students required the entire allotted time to complete the book (four in the providing-explanations condition, five in the providing-solutions condition, and three in the withholding condition), and three of them did not get to the final problem before time expired (two in the providing-explanations condition and one in the providing-solutions condition). To address the possibility that these three students might lower the performance of both providing groups, we have excluded them from all further analyses.

Upon completing the instructional texts and problems, participants responded to a task goals questionnaire, completed a two-part posttest, and then responded to a demographic questionnaire. Aside from the change in placement of the general achievement orientations questionnaire, the procedure was identical to that of Experiment 2.

4.2. Results

As with Experiments 1 and 2, we examined the effect of learning condition on posttest outcomes. We also examined whether the same pattern of relationships observed in Experiment 2 between learning, achievement orientations, and task goals would hold for this experiment.

4.2.1. Pretest accuracy

Conditions were equivalent at pretest, with a one-way ANOVA revealing no differences between the providing-explanations condition ($M = .60$, $SD = .27$), providing-solutions condition ($M = .64$, $SD = .18$), and withholding condition ($M = .56$, $SD = .24$) on pretest performance, $F(2, 86) = .95$, $p = .39$. There were also no differences on the definition questions, $F(2, 86) = 1.18$, $p = .31$, problem-solving questions, $F(2, 86) = 2.13$, $p = .13$, or conceptual questions, $F(2, 86) = .65$, $p = .52$.

4.2.2. Learning materials

As with Experiment 2, participants were generally very thorough in completing learning materials. Students in the providing conditions tended to complete all boxes and either copied or summarized the information in the worked examples, while responses from students in the withholding condition ranged from simply writing the answer to copying the steps contained in the worked examples to providing brief explanations. Participants in both the providing-explanations condition ($M = .90$, $SD = .13$) and the withholding condition ($M = .88$, $SD = .16$) showed very high accuracy in their performance on the learning materials, and a one-way ANOVA comparing accuracy between learning conditions found no effects, $F(1, 57) = .52$, $p = .47$. These results are consistent with Experiment 2. Participants in the providing-solutions condition could not be assessed on accuracy because they were given the solutions, but they demonstrated high completion rates, with participants completing most ($M = .85$, $SD = .27$) of the explanation boxes available in their practice problems.

A single-variable regression revealed that accuracy on learning materials was strongly predictive of overall posttest accuracy, $\beta = .58$, $t(57) = 5.38$, $p < .001$, and of all posttest measures. This shows that completing materials accurately was associated with better posttest performance and suggests that the test materials and learning materials were well aligned.

4.2.3. Posttest performance

To test whether the learning materials created an overall learning effect, a paired-samples *t*-test was conducted to compare students' pretest performances with their scores on the first section of the posttest, which contained isomorphic problems matching the pretest and counterbalanced to control for difficulty. The test indicated that posttest scores ($M = .86$, $SD = .16$) were higher than pretest scores ($M = .60$, $SD = .23$), $t(88) = 12.97$, $p < .001$, $d = 1.31$. This demonstrates that students overall showed a sizable learning gain, with the effect size indicating improvement of more than a standard deviation. To provide a stringent test of learning condition on posttest scores, we conducted one-way ANCOVAs for each posttest measure using the pretest score as a covariate. Fig. 5 shows the estimated marginal means and standard error bars for each condition, controlling for pretest performance, on all posttest measures except for the definition and problem-solving data, which reflect the means without controlling for pretest.

4.2.3.1. Definition. A preliminary analysis evaluating the homogeneity-of-regression assumption revealed a medium interaction between condition and pretest performance, $F(5, 83) = 4.57$, $p = .01$, $\eta_p^2 = .10$. This indicates the covariate interacts with the independent variable in different ways across conditions, rendering an ANCOVA unfit for interpretation. Consequently, we evaluated the effect of condition on definition performance using a standard analysis of variance without controlling for the covariate. A one-way ANOVA revealed no effect of condition on definition scores, $F(2, 86) = 1.38$, $p = .26$.

4.2.3.2. Problem solving. A preliminary analysis evaluating the homogeneity-of-regression assumption revealed a medium interaction between condition and pretest performance, $F(5, 83) = 3.69$, $p = .03$, $\eta_p^2 = .08$, indicating that an ANCOVA would be unfit for interpretation. Consequently, we evaluated the effect of condition on problem-solving performance using a standard ANOVA without controlling for the covariate. A one-way ANOVA revealed a marginal effect of condition on problem-solving scores, $F(2, 86) = 2.53$, $p = .09$. Planned comparisons were conducted to assess pairwise differences between the adjusted means for each condition, using the Bonferroni procedure to control for Type 1 error across the three pairwise comparisons ($\alpha' = .05/3 = .017$). Comparisons showed that the withholding condition ($M = .90$, $SE = .03$) and the providing-solutions condition ($M = .77$, $SE = .05$) were not significantly different once the Bonferroni correction was applied, $F(1, 59) = 4.92$, $p = .03$. There were also no differences between the withholding condition and the providing-explanations condition, ($M = .85$, $SE = .04$), $F(1, 57) = .83$, $p = .37$, nor between the providing-explanations and providing-solutions conditions, $F(1, 57) = 1.56$, $p = .22$. This suggests that the difference was not robust.

4.2.3.3. Conceptual. A preliminary analysis evaluating the homogeneity-of-regression assumption revealed no interaction

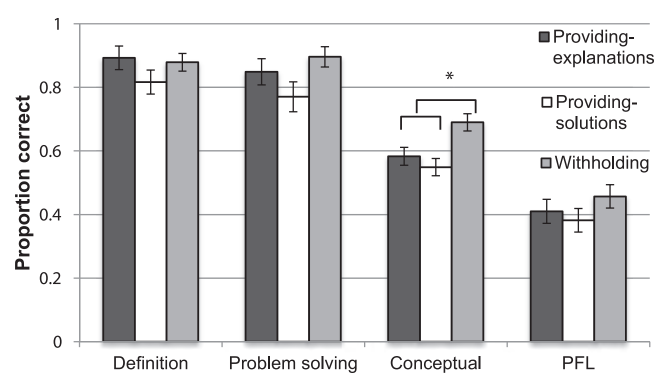

Fig. 5. Learning condition effect on posttest accuracy, controlling for pretest performance. * denotes $p < .05$.

between condition and pretest performance, $F(5, 83) = 1.13$, $p = .33$. A one-way ANCOVA revealed a large effect of the covariate, $F(3, 85) = 37.96$, $p < .001$, $\eta_p^2 = .31$, showing that students' pretest scores predicted their conceptual posttest scores. There was also a large effect of condition on conceptual questions, $F(3, 85) = 7.43$, $p = .001$, $\eta_p^2 = .14$. Planned comparisons were conducted to assess pairwise differences between the adjusted means for each condition, using the Bonferroni procedure to control for Type 1 error across the three pairwise comparisons ($\alpha' = .05/3 = .017$). Comparisons showed that the withholding condition ($M = .68$, $SE = .03$) performed better than the providing-solutions condition ($M = .55$, $SE = .03$), $F(2, 58) = 17.02$, $p < .001$, with a large effect size of $\eta_p^2 = .23$. The withholding condition also performed better than the providing-explanations condition, ($M = .58$, $SE = .03$), $F(2, 56) = 6.84$, $p = .01$, with a medium effect size of $\eta_p^2 = .11$, while there were no differences between the providing-explanations and providing-solutions conditions, $F(2, 55) = .69$, $p = .41$. This is consistent with the hypothesis that participants in the withholding condition would learn conceptual information better than those in either providing condition, and also suggests that there was no difference in conceptual learning between the providing-explanations and providing-solutions conditions.

4.2.3.4. Preparation for future learning. A preliminary analysis evaluating the homogeneity-of-regression assumption revealed no interaction between condition and pretest performance, $F(5, 83) = 1.54$, $p = .22$. A one-way ANCOVA revealed a large effect of the covariate, $F(3, 85) = 46.67$, $p < .001$, $\eta_p^2 = .35$, showing that students' pretest scores predicted their PFL posttest scores. There was no effect of condition on PFL scores, $F(3, 85) = 1.03$, $p = .36$.

4.2.4. Achievement orientations

A one-way ANOVA revealed no differences between conditions, $Fs < 2.28$, $ps > .11$, which was consistent with expectations. To investigate differences between the role of achievement orientations in determining learning outcomes for participants in the providing and withholding conditions, a series of hierarchical regression analyses that used each orientation and controlled for pretest performance were used to predict posttest measures separately for each learning condition. As with both previous experiments, we focus on the overall posttest accuracy and specific measures that showed learning condition effects, which for this experiment was conceptual accuracy. For the providing-explanations condition, no achievement orientations were predictive of the learning outcomes (all β's $< .12$ and $> -.11$, $ps > .43$). For the providing-solutions condition, mastery-avoidance and performance-avoidance were marginal, negative predictors performance on the overall posttest, $\beta = -.24$, $t(27) = -1.86$, $p = .07$, and $\beta = -.23$, $t(27) = -1.82$, $p = .08$. No other achievement orientations were predictive of learning outcomes for the providing-solutions condition (all $\beta s < .09$ and $> -.14$, $ps > .35$).

In contrast, for the withholding condition, mastery-avoidance predicted overall accuracy and marginally predicted conceptual accuracy, $\beta = .34$, $t(28) = 2.26$, $p = .03$, and $\beta = .27$, $t(28) = 1.89$, $p = .07$, respectively. No other achievement orientations were predictive of these learning outcomes (all $\beta s < .26$, $ps > .11$). Together, these results are consistent with predictions that achievement orientations would better predict posttest performance for participants receiving withholding materials than for those who received providing materials. Furthermore, they are consistent with results from Experiment 2 showing that mastery-avoidance orientations affect learning among college students with prior exposure to physics, suggesting that for this task they were striving to avoid losing previous competence in physics.

4.2.5. Task goals

A one-way ANOVA conducted on each goal found no condition effects, $Fs < 1.5$, $ps > .22$. This is contrary to findings in Experiment 2 showing that participants in the withholding condition reported greater mastery-approach task goals than participants in the providing condition. To investigate differences in the role of task goals in determining learning outcomes for participants in the providing and withholding conditions, a hierarchical regression analysis controlling for pretest performance was used again to predict learning. Again, we focus on overall posttest accuracy and any measures that showed a learning condition effect. For the providing-explanations condition, mastery-approach goals predicted overall accuracy, $\beta = .35$, $t(25) = 2.44$, $p = .02$, and marginally predicted conceptual accuracy, $\beta = .31$, $t(25) = 1.69$, $p = .10$. No other goals were significant predictors for the providing-explanations condition (all β's $< .20$, $ps > .18$). For the providing-solutions condition, no task goals predicted posttest performance (all $\beta s < 0$ and $> -.11$, $ps > .44$). For the withholding condition, mastery-approach predicted overall accuracy, $\beta = .33$, $t(28) = 2.15$, $p = .04$. No other task goals were a significant predictor of posttest performance for the withholding condition (all $\beta s < .21$, $ps > .17$).

4.2.6. Task goals and achievement orientations

Although task goals and achievement orientations were predictive of different posttest outcomes, the two measures were correlated. Mastery-approach task goals and achievement orientations were correlated, $r = .58$, $p < .001$, as were performance-approach task goals and achievement orientations, $r = .44$, $p < .001$, and performance-avoidance task goals and achievement orientations, $r = .46$, $p < .001$. A condition effect on the correlations of task goals and achievement orientations emerged. For the providing-solutions condition and the withholding condition, all three measures remained correlated, $rs > .41$, $ps < .03$ and $rs > .48$, $ps < .01$, respectively. For the providing-explanations condition, the two mastery-approach measures remained correlated, $r = .50$, $p = .01$; however performance-approach task goals and achievement orientations were marginally correlated, $r = .37$, $p = .05$, and performance-avoidance task goals and achievement orientations were not correlated, $r = .27$, $p = .17$. This is consistent with the pattern of findings from Experiment 2, which showed that participants' task goals in the withholding condition were closely related with their general achievement orientations toward science, while in the providing condition, some of participants' existing achievement orientations were not expressed in the form of task goals. Taken together, these results suggest the materials for the providing-explanations condition may have altered the expression of some general achievement orientations for the task at hand.

4.3. Discussion

Similar to Experiments 1 and 2, the withholding condition demonstrated greater performance on the conceptual learning measure [Hypothesis 1]. Critically, this effect remained when the number of practice problems was controlled across conditions, thus eliminating one potential explanation of results from Experiments 1 and 2. Results also showed that the detrimental effects of providing instructional explanations emerge even when practice problems prompt students to generate stepwise explanations, as the providing-solutions condition did. This supports our findings from Experiment 1 by showing the condition effect could not have emerged as a result of including instructional explanations in the practice problems.

As in Experiment 2, the achievement orientation results highlight the connection between mastery-avoidance goals and conceptual learning outcomes among college students. These results show that mastery-avoidance dispositions are particularly important for conceptual learning when instructional explanations are not provided [Hypothesis 3]. As previously mentioned (Sections 2.3 and 3.3), the absence of instructional explanations may facilitate more opportunities for students to self-explain, a learning strategy that may be driven by mastery goals. Consequently, those in the withholding condition with higher mastery-avoidance orientations might engage more frequently in self-explanation which, in turn, could lead to greater conceptual learning outcomes.

Experiment 3 did not replicate the Experiment 2 effect of students in the withholding condition reporting higher mastery-approach task goals (Section 3.2.5), suggesting that this may not be a robust effect [Hypothesis 4]. Future work should examine this more closely to better identify the precise conditions under which instructional explanations might affect the task-level achievement goals students adopt. As with Experiment 2 (Section 3.2.5), task goals appeared to play a greater role in predicting learning outcomes for students in the providing-explanations conditions compared to the withholding condition. This is consistent with our hypothesis that dispositional goals may take a more important role in the absence of instructional explanations while goals adopted for a specific task may play a more important role in the presence of instructional explanations (Section 3.3). As with Experiment 2, however, we caution that our examination of task goals is exploratory and results should be further tested in future work.

5. General discussion

Across all three experiments, participants in a condition that withheld instructional explanations for worked examples and practice problems demonstrated greater conceptual accuracy on a posttest than participants in a condition that provided instructional explanations [Hypothesis 1]. This result appears robust, as it emerged across multiple experimental settings and populations and persisted when controlling for the number of practice problems. Neither condition was explicitly prompted to self-explain, suggesting that participants may have spontaneously engaged in self-explanation in the withholding conditions. Even when participants were prompted to generate stepwise explanations for the practice problems after receiving instructional explanations in the worked examples (providing condition in Experiment 1 and providing-solutions condition in Experiment 3), their learning gains still failed to match those of participants who saw no instructional explanations. If participants in the withholding condition were more likely to self-explain, it may be that constructing explanations is more effective than applying explanations that have been previously supplied. Future work should test this hypothesis by collecting think-aloud protocols.

Several past studies and theoretical frameworks support the idea that self-explanation is the mechanism responsible for the withholding condition's superior conceptual learning. Schworm and Renkl (2006) found that providing instructional explanations was less effective than prompting self-explanations. They also found an interaction between providing instructional explanations and prompting self-explanation in which instructional explanations suppressed students' self-explaining when they were prompted, making self-explanation prompts alone more effective than prompts paired with instructional explanations. Schworm and Renkl's (2006) self-explanation with instructional explanations condition is similar to our providing condition in Experiment 1 and our providing-solutions condition in Experiment 3, in that participants in these conditions were given instructional explanations in worked examples and then prompted to fill in their own explanations of the practice problems. Our results are consistent with Schworm and Renkl's (2006) findings and suggest that providing instructional explanations suppresses self-explanation, leading to less conceptual learning than the withholding condition.

One alternative explanation is that the providing conditions created a type of redundancy effect by giving learners more information than necessary once they had successfully acquired the principles described in the instructional explanations from the worked examples. van Gog et al. (2008) found that participants benefitted from stepwise instructional explanations accompanying worked examples when they were first practicing a task, but that the explanations began to hurt performance with more practice. Specifically, they found that providing instructional explanations of solution steps improved efficiency over a set of four examples but suppressed performance and efficiency over a second set of four examples. The authors attributed the effect to extraneous cognitive load created by the instructional explanations once participants had sufficiently mastered the concepts being explained. In our study, however, students saw only one worked example with instructional explanations for each problem type, accompanied by one to two practice problems with instructional explanations. Given that van Gog et al. (2008) failed to find a redundancy effect until after participants had solved more than four isomorphic problems, it seems unlikely that the instructional explanations in our experiment would become redundant on only the second instance of a problem type.

Looking only at participants' performance on the first set of four problems, van Gog et al.'s (2008) results suggested that providing instructional explanations improved participants' efficiency, though it had no effect on accuracy. Although the study tested participants' learning using two near-transfer (same structure) and two far-transfer (different structure - either a novel type of circuit with a familiar issue or a familiar circuit with novel issue) problems, they report only the collapsed performance. As our differences emerged on conceptual questions, it may be that results more similar to ours would have been found after the first four problems by looking at the two types of problems separately. Specifically, on the first four-problem set in van Gog et al. (2008), we might expect to find an instructional explanation advantage on near-transfer problems, which do not require the same kind of deep learning to perform, but an absence of an advantage or a small disadvantage on far-transfer problems, which require the constructive learning processes necessary to develop deep knowledge. Such a disadvantage on far-transfer problems, if it exists, would be consistent with our findings in the present study.

Viewed through Chi's (2009) active-constructive-interactive framework, instructional explanations might promote *active* behaviors such as summarizing and paraphrasing and discourage *constructive* behaviors like generating inferences, connecting ideas, and restructuring knowledge. In the absence of clear instructional explanations, some students are more likely to engage in these constructive activities to make sense of information on their own. As not all students possess sufficient achievement motivation to spontaneously engage in constructive activities, we find that achievement goals act as a moderating variable predicting performance in the withholding condition but not the providing condition.

Mastery-avoidance orientations consistently predicted performance on the overall test and on conceptual measures within the withholding condition for Experiments 2 and 3. These experiments involved a college population that had previous, but not recent, exposure to the topic of physics. Mastery-avoidance is thought to occur among people who have previously mastered a skill or topic and are attempting to avoid losing that mastery. Consequently, mastery-avoidance may have played a more prominent role in Experiments 2 and 3 as students aimed to avoid losing their prior understanding of basic physics concepts. While mastery-approach goals have generally been associated with more constructive learning behaviors and deeper learning outcomes, it remains unclear just how much mastery-avoidance goals resemble mastery-approach goals in terms of their benefits. This work provides an interesting demonstration of the role of mastery-avoidance in academic learning and suggests that future studies should explore this orientation in situations that involve relearning a topic.

A general pattern emerged to suggest that achievement orientations were more important when instructional explanations were absent and task goals were more important when instructional explanations were present [Hypothesis 3]. This suggests the relative stabilities of orientations and task goals may benefit different students depending on the achievement orientations they bring to the task. For example, students with mastery orientations may be more likely to engage in constructive cognitive activities. Students who lack these orientations may need at least some degree of structure, however, so they can rely less on their goals and more on the instructional explanations provided. Findings from Harris et al. (2009) support this idea, suggesting that students with performance goals prefer more structure and guidance than students with mastery goals. Based on our preliminary results, it is possible that providing instructional explanations can alter students' achievement orientations, which may benefit those who lack a mastery orientation. Future research should attempt to capture process data to provide more details into the mechanisms through which different achievement goals might drive learning, as well as how those goals might interact with instructional conditions.

Several published studies on self-explanation might seem contradictory to our result that providing instructional explanations and then prompting students to generate explanations on practice problems would suppress conceptual learning compared to a condition that withheld instructional explanations. For example, several experiments testing the effect of prompting participants to select the appropriate principle for a problem-solving step have reported benefits to accuracy or efficiency compared to an unprompted control condition (Aleven, Popescu, & Koedinger, 2002; Conati & Vanlehn, 2000). We argue that there are several critical factors that differ between our experiments and this past work that can account for the differences in results.

The primary difference is that we provided instructional explanations in the worked examples before students were asked to generate explanations on the practice problems, whereas the prior work typically has not included instructional explanations embedded in worked examples. Having instructional explanations provided in the worked examples may have encouraged participants to simply summarize or paraphrase those explanations – *active* behaviors in Chi's (2009) framework – when prompted to provide their own explanations while solving practice problems. Had those explanations not been provided, participants might have been more likely to attempt more *constructive* behaviors when generating their own instructional explanations in Experiments 1 (providing condition) and 3 (providing-solutions condition).

A second difference is that the past work on self-explanation prompts has almost exclusively measured learning by having participants complete additional problem-solving tasks. While some types of problem solving can assess conceptual knowledge, we argue that relying solely on problem-solving measures is likely to miss complex learning effects, especially in physics domains (Hestenes et al., 1992). Had our present set of experiments measured learning only through problem solving, our results would suggest a null effect of condition. Furthermore, Wittwer and Renkl (2010) found that instructional explanations improved learning in math but not in science. Consequently, it may be that instructional explanations operate through mechanisms that are more effective in some domains than others.

In sum, we believe that differences between our current findings and past work may be attributable to whether or not instructional explanations were included in the worked examples, the types of measures used to assess learning, and the domain in which the intervention was conducted. These findings have important implications for instruction and suggest that one cannot simply combine what appear to be complementary instructional strategies (e.g., instructional explanations and self-explanation prompts) and expect additive learning outcomes. In the current case, combining these two instructional strategies led to decrements in conceptual understanding because students used the former to avoid constructive thinking on the latter by simply paraphrasing those explanations on the practice problems.

A potential future question concerns whether providing a lower dose of instructional explanations – i.e., incorporating instructional explanations in the worked examples but using open-ended practice problems – would mitigate the detrimental effects on conceptual learning. Based on Wittwer and Renkl's (2010) review, we hypothesize that including instructional explanations only in worked examples would not improve learning in a science domain compared to worked examples without instructional explanations, but it is possible that the negative effects found in the current study might be reduced by confining instructional explanation use to only the worked example portion. Future work should test this possibility.

The results across all experiments highlight the importance of multiple measures of learning. Our findings showed consistent advantages of withholding instructional explanations for far transfer as measured by conceptual questions but not for near transfer as measured by definition questions and problem-solving items with similar structures to the worked examples. This is consistent with our idea that providing instructional explanations can interfere with constructive cognitive processes, which are most important for deep, conceptual learning. In fact, the instructional explanations may support learning of definitions and near-transfer problem-solving skills, as we found no condition effects for either in our experiments and participants in Experiments 2 and 3 showed high posttest performance on these two measures. In general, preparation for future learning effects were the most difficult to find and students' low performance across experiments on the PFL measures suggest the brief intervention did not create knowledge robust enough to transfer to new learning opportunities in Experiments 2 and 3.

This work also takes the first steps toward unpacking the complex relationships among withholding and providing instructional explanations, the role of prior knowledge, and the influence of achievement orientations and task goals on learning outcomes. Given the weak learning effects of the conceptual background materials in Experiment 1, we were not able to thoroughly test the effect of ontological training [Hypothesis 2]. Prior knowledge, however, appears to play an important role as we showed larger learning gains for college students, most of whom had previous exposure to physics. These results should be replicated in college classrooms.

This work suggests that providing fewer instructional explanations during worked-example study and problem-solving activities promotes conceptual learning if there are other resources available to support learning (e.g., text, example problems, prior knowledge). It also suggests that there are important individual differences stemming from students' achievement orientations and task goals. Future work should examine whether this effect of additional instructional explanations also holds in different learning environments, such as intelligent tutoring systems, or at different stages in the learning process, such as when students are reading about a topic rather than solving problems. The results also suggest that instructors should resist the intuitive appeal of providing additional explanations whenever possible, as such practices may reduce learning and transfer.

Acknowledgments

This research was supported by Grant SBE00354420 from the National Science Foundation to the Pittsburgh Science of Learning Center (http://www.learnlab.org). We thank research assistants Emily Schmidt, Aleza Wallace, Anh Huynh, and Sarah Honsaker for their help in coding the data. We would also like to thank Michelene Chi and James Slotta for sharing their emergent process training materials. We gratefully acknowledge Christian Schunn and Kevin Crowley for their insightful feedback on this work, as well as Lucia Mason and four anonymous reviewers for their very helpful suggestions and comments on the article.

References

Aleven, V. A. W. M. M., & Koedinger, K. R. (2000). The need for tutorial dialog to support self-explanation. In C. P. Rose, & R. Freedman (Eds.), *Building dialogue systems for tutorial applications, papers of the 2000 AAAI Fall Symposium* (pp. 65–73). Menlo Park, CA: AAAI Press.

Aleven, V. A. W. M. M., Popescu, O., & Koedinger, K. (2002). Pilot-testing a tutorial dialogue system that supports self-explanation. In S. A. Cerri, G. Gouardères, & F. Paraguaçu (Eds.), *Proceedings of the sixth international conference on intelligent tutoring systems, ITS 2002* (pp. 344–354). Berlin: Springer Verlag.

Alfieri, L., Nokes-Malach, T. J., & Schunn, C. D. Learning through case comparisons: a meta-analytic review. Educational Psychologist. Poster session presented at the Fall 2011 conference for the Society for Research in Educational Effectiveness, Washington, D.C., in press.

Ames, C. (1992). Classrooms: goals, structures, and student motivation. *Journal of Educational Psychology, 84*(3), 261–271. http://dx.doi.org/10.1037/0022-0663.84.3.261.

Ames, C., & Archer, J. (1988). Achievement goals in the classroom: students' learning strategies and motivation processes. *Journal of Educational Psychology, 80*(3), 260–267. http://dx.doi.org/10.1037/0022-0663.80.3.260.

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin, 128*(4), 612–637. http://dx.doi.org/10.1037//0033-2909.128.4.612.

Belenky, D. M., & Nokes, T. J. (2010). Optimizing learning environments: an individual difference approach to learning and transfer. In S. Ohlsson, & R. Catrambone (Eds.), *Proceedings of the thirty-second annual conference of the cognitive science society* (pp. 459–464). Portland, OR: Cognitive Science Society.

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

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

Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: a simple proposal with multiple implications. In A. Iran-Nejad, & P. D. Pearson (Eds.). *Review of research in education, Vol. 24* (pp. 61–100). Washington, D.C.: American Educational Research Association.

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, Vol. 5* (pp. 161–238). Lawrence Erlbaum Associates.

Chi, M. T. H. (2008). Three types of conceptual change: belief revision, mental model transformation, and categorical shift. In S. Vosniadou (Ed.), *International handbook of research on conceptual change* (pp. 61–82). Hillsdale, N. J.: Erlbaum.

Chi, M. T. H. (2009). Active-constructive-interactive: a conceptual framework for differentiating learning activities. *Topics in Cognitive Science, 1*, 73–105. http://dx.doi.org/10.1111/j.1756-8765.2008.01005.x.

Chi, M. T. H. (2011). Theoretical perspectives, methodological approaches, and trends in the study of expertise. In Y. Li, & G. Kaiser (Eds.), *Expertise in mathematics instruction: Advancing research and practice from an international perspective* (pp. 17–39). New York: Springer. http://dx.doi.org/10.1007/978-1-4419-7707-6_2.

Chi, M. T. H., de Leeuw, N., Chiu, M.-H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18*(3), 439–477. http://dx.doi.org/10.1207/s15516709cog1803_3.

Chi, M. T. H., Slotta, J. D., & de Leeuw, N. (1994). From things to processes: a theory of conceptual change for learning science concepts. *Science, 4*, 27–43.

Chouinard, R., & Roy, N. (2008). Changes in high-school students' competence beliefs, utility value and achievement goals in mathematics. *British Journal of Educational Psychology, 78*, 31–50. http://dx.doi.org/10.1348/000709907X197993.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, doi:10.1234/12345678.

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*, 389–415.

Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist, 41*(10), 1040–1048. http://dx.doi.org/10.1037//0003-066X.41.10.1040, American Psychological Association.

Elliot, A. J., & Harackiewicz, J. M. (1996). Approach and avoidance achievement goals and intrinsic motivation: a mediational analysis. *Journal of Personality and Social Psychology, 70*(3), 461–475. http://dx.doi.org/10.1037/0022-3514.70.3.461, American Psychological Association.

Elliot, A. J., & McGregor, H. A. (2001). A 2 X 2 achievement goal framework. *Journal of Personality and Social Psychology, 80*(3), 501–519. http://dx.doi.org/10.1037//0022-3514.80.3.501.

Elliot, A. J., McGregor, H. A., & Gable, S. (1999). Achievement goals, study strategies, and exam performance: a mediational analysis. *Journal of Educational Psychology, 91*(3), 549–563.

Elliot, A. J., & Murayama, K. (2008). On the measurement of achievement goals: critique, illustration, and application. *Journal of Educational Psychology, 100*(3), 613–628. http://dx.doi.org/10.1037/0022-0663.100.3.613.

Engle, R. A. (2012). The resurgence of research into transfer: an introduction to the final articles of the transfer strand. *Journal of the Learning Sciences, 21*(3), 347–352. http://dx.doi.org/10.1080/10508406.2012.707994.

Ezrailson, C., Zike, D., & Zorn, M. K. (2005). *Electricity and magnetism*. p. 138 (2nd ed.). Glencoe/McGraw-Hill.

Gentner, D., Loewenstein, J., & Thompson, L. (2003). Learning and transfer: a general role for analogical encoding. *Journal of Educational Psychology, 95*(2), 393–405. http://dx.doi.org/10.1037/0022-0663.95.2.393.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15*(1), 1–38. http://dx.doi.org/10.1016/0010-0285(83)90002-6.

van Gog, T., Paas, F., & van Merriënboer, J. J. G. (2008). Effects of studying sequences of process-oriented and product-oriented worked examples on troubleshooting transfer efficiency. *Learning and Instruction, 18*(3), 211–222. http://dx.doi.org/10.1016/j.learninstruc.2007.03.003.

Grant, H., & Dweck, C. S. (2003). Clarifying achievement goals and their impact. *Journal of Personality and Social Psychology, 85*(3), 541–553.

Hake, R. R. (1998). Interactive-engagement versus traditional methods: a six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics, 66*(1), 64–74. http://dx.doi.org/10.1119/1.18809.

Harackiewicz, J. M., Barron, K. E., Pintrich, P. R., Elliot, A. J., & Thrash, T. M. (2002). Revision of achievement goal theory: necessary and illuminating. *Journal of Educational Psychology, 94*(3), 638–645. http://dx.doi.org/10.1037//0022-0663.94.3.638.

Harris, A., Bonnett, V., Luckin, R., Yuill, N., & Avramides, K. (2009). Scaffolding effective help-seeking behaviour in mastery and performance oriented learners. In V. Dimitrova, R. Mizoguchi, B. du Boulay, & A. C. Graesser (Eds.), *Proceedings of the 14th international conference on artificial intelligence in education* (pp. 425–432). Brighton, UK: IOS Press.

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.). *13th international conference on artificial intelligence in education, Vol. 158* (pp. 417–424). Amsterdam: IOS Press.

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher, 30*(March), 141–158.

Hsu, T. (2005). *Foundations of physical science* (2nd ed.). CPO Science.

Hulleman, C. S., Durik, A. M., Schweigert, S. B., & Harackiewicz, J. M. (2008). Task values, achievement goals, and interest: an integrative analysis. *Journal of Educational Psychology, 100*(2), 398–416. http://dx.doi.org/10.1037/0022-0663.100.2.398, American Psychological Association.

Keppel, G., & Wickens, T. D. (2004). *Design and analysis: A researcher's handbook* (4th ed.). Upper Saddle River, NJ: Pearson Education International.

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

Kurtz, K. J., Miao, C.-H., & Gentner, D. (2001). Learning by analogical bootstrapping. *The Journal of Learning Sciences, 10*(4), 417–446.

Leinhardt, G. (2001). Instructional explanations: a commonplace for teaching and location for contrast. In V. Richardson (Ed.), *Handbook of research on teaching* (4th ed.) (pp. 333–357). Washington, D.C: American Educational Research Association.

Leinhardt, G. (2010). Introduction: explaining instructional explanations. In M. K. Stein, & L. Kucan (Eds.), *Instructional explanations in the disciplines* (pp. 1–5). Boston, MA: Springer US. http://dx.doi.org/10.1007/978-1-4419-0594-9.

Linnenbrink-Garcia, L., Tyson, D. F., & Patall, E. A. (2008). When are achievement goal orientations beneficial for academic achievement? A closer look at main effects and moderating factors. *International Review of Social Psychology, 21*(1–2), 19–70, Presses univ. de Grenoble.

Lovett, M. C. (1992). Learning by problem solving versus by examples: the benefits of generating and receiving information. In *Proceedings of the 14th annual conference of the cognitive science society* (pp. 956–961). Bloomington, IN: Erlbaum.

Mclaren, B. M., Lim, S.-J., & Koedinger, K. R. (2008). When and how often should worked examples be given to students? New results and a summary of the current state of research. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.),

Proceedings of the 30th annual conference of the cognitive science society (pp. 2176–2181). Austin, TX: Cognitive Science Society.
Nokes-Malach, T. J., & Mestre, J. Towards a model of transfer as sense-making. *Educational Psychologist*, in press.
Nokes-Malach, T. J., VanLehn, K. A., Belenky, D. M., Lichtenstein, M., & Cox, G. Coordinating principles and examples through analogy and explanation. *European Journal of Psychology of Education*, in press.
Nokes, T. J., & Belenky, D. M. (2011). Incorporating motivation into a theoretical framework for knowledge transfer. In J. P. Mestre, & B. H. Ross (Eds.). *The psychology of learning and motivation: Cognition and education, Vol. 55* (pp. 109–135). San Diego, CA: Academic Press. http://dx.doi.org/10.1016/B978-0-12-387691-1.00004-1.
Nokes, T. J., Hausmann, R. G. M., VanLehn, K., & Gershman, S. (2011). Testing the instructional fit hypothesis: the case of self-explanation prompts. *Instructional Science, 39*(5), 645–666. http://dx.doi.org/10.1007/s11251-010-9151-4.
Nokes, T. J., & Ross, B. H. (2007). Facilitating conceptual learning through analogy and explanation. In L. Hsu, C. Henderson, & L. McCullough (Eds.), *Physics education research conference* (pp. 7–10). Melville, NY: American Institute of Physics Conference Proceedings. http://dx.doi.org/10.1063/1.2820952.
Olejnik, S., & Algina, J. (2000). Measures of effect size for comparative studies: applications, interpretations, and limitations. *Contemporary Educational Psychology, 25*(3), 241–286. http://dx.doi.org/10.1006/ceps.2000.1040.
Paas, F. G. W. C., & Van Merriënboer, J. J. G. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: a cognitive-load approach. *Journal of Educational Psychology, 86*(1), 122–133. http://dx.doi.org/10.1037/0022-0663.86.1.122.
Perkins, D. N., & Salomon, G. (2012). Knowledge to go: a motivational and dispositional view of transfer. *Educational Psychologist, 47*(3), 248–258. http://dx.doi.org/10.1080/00461520.2012.693354.
Pugh, K. J., & Bergin, D. A. (2006). Motivational influences on transfer. *Educational Psychologist, 41*(3), 147–160. http://dx.doi.org/10.1207/s15326985ep4103_2.
Renkl, A. (1997). Learning from worked-out examples: a study on individual differences. *Cognitive Science, 21*(1), 1–29. http://dx.doi.org/10.1207/s15516709cog2101_1.
Renkl, A. (2002). Worked-out examples: instructional explanations support learning by self-explanations. *Learning and Instruction, 12*(5), 529–556. http://dx.doi.org/10.1016/S0959-4752(01)00030-5.
Renkl, A., Atkinson, R. K., & Maier, U. H. (2000). From studying examples to solving problems: fading worked-out solution steps helps learning. In L. Gleitman, & A. K. Joshi (Eds.), *Proceedings of the 22nd annual conference of the cognitive science society* (pp. 393–398). Mahwah, NJ: Lawrence Erlbaum Associates, http://dx.doi.org/10.1.1.23.6816.
Renkl, A., Atkinson, R. K., Maier, U. H., & Staley, R. (2002). From example study to problem solving: smooth transitions help learning. *The Journal of Experimental Education, 70*(4), 293–315. http://dx.doi.org/10.1080/00220970209599510.
Rittle-Johnson, B. (2006). Promoting transfer: effects of self-explanation and direct instruction. *Child Development, 77*(1), 1–15. http://dx.doi.org/10.1111/j.1467-8624.2006.00852.x.
Rittle-Johnson, B., & Star, J. R. (2007). Does comparing solution methods facilitate conceptual and procedural knowledge? An experimental study on learning to solve equations. *Journal of Educational Psychology, 99*(3), 561–574. http://dx.doi.org/10.1037/0022-0663.99.3.561.
Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: common principles in three paradigms suggest new concepts for training. *Psychological Science, 3*(4), 207–217. http://dx.doi.org/10.1111/j.1467-9280.1992.tb00029.x.
Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment, 8*(4), 350–353. http://dx.doi.org/10.1037//1040-3590.8.4.350.
Schwartz, D. L., Bransford, J. D., & Sears, D. (2005). Efficiency and innovation in transfer. In J. Mestre (Ed.), *Transfer of learning from a modern multidisciplinary perspective* (pp. 1–51). CT: Information Age Publishing.
Schworm, S., & Renkl, A. (2006). Computer-supported example-based learning: when instructional explanations reduce self-explanations. *Computers & Education, 46*(4), 426–445. http://dx.doi.org/10.1016/j.compedu.2004.08.011.
Slotta, J. D., & Chi, M. T. H. (2006). Helping students understand challenging topics in science through ontology training. *Cognition and Instruction, 24*(2), 261–289. http://dx.doi.org/10.1207/s1532690xci2402_3.
Ward, M., & Sweller, J. (1990). Structuring effective worked examples. *Cognition and Instruction, 7*(1), 1–39. http://dx.doi.org/10.1207/s1532690xci0701_1.
Wilson, J. W., & Carry, L. R. (1969). Homogeneity of regression, its rationale, computation and use. *American Educational Research Journal, 6*, 80–90.
Wittwer, J., & Renkl, A. (2008). Why instructional explanations often do not work: a framework for understanding the effectiveness of instructional explanations. *Educational Psychologist, 43*(1), 49–64. http://dx.doi.org/10.1080/00461520701756420.
Wittwer, J., & Renkl, A. (2010). How effective are instructional explanations in example-based learning? A meta-analytic review. *Educational Psychology Review, 22*(4), 393–409. http://dx.doi.org/10.1007/s10648-010-9136-5.