Journal of Educational Psychology
2015, Vol. 107, No. 4, 954–970

0022-0663/15/$12.00 http://dx.doi.org/10.1037/edu0000022

Direct Instruction of Metacognition Benefits Adolescent Science Learning, Transfer, and Motivation: An In Vivo Study

Cristina D. Zepeda and J. Elizabeth Richey
University of Pittsburgh

Paul Ronevich
Pittsburgh Science and Technology Academy

Timothy J. Nokes-Malach
University of Pittsburgh

Prior studies have not tested whether an instructional intervention aimed at improving metacognitive skills results in changes to student metacognition, motivation, learning, and future learning in the classroom. We examined whether a 6-hr intervention designed to teach the declarative and procedural components of planning, monitoring, and evaluation could increase students' metacognition, motivation, learning, and preparation for future learning for middle school science. Forty-six eighth-grade students were randomly assigned to either a control group, which received extensive problem-solving practice, or an experimental group, which received more limited problem-solving practice along with metacognitive instruction and training. Results revealed that those who received the metacognitive instruction and training were less biased when making metacognitive judgments, $p = .03$, $d = 0.65$, endorsed higher levels of motivation after instruction (e.g., there was a large effect on task value, $p = .006$, $d = 0.87$), performed better on a conceptual physics test, $p = .03$, $d = 0.64$, and performed better on a novel self-guided learning activity, $p = .007$, $d = 0.87$. This study demonstrates that metacognitive instruction can lead to better self-regulated learning outcomes during adolescence, a period in which students' academic achievement and motivation often decline.

Keywords: instruction, learning, metacognition, motivation, transfer

Supplemental materials: http://dx.doi.org/10.1037/edu0000022.supp

A student's ability to adapt his or her problem-solving behaviors to different types of academic tasks and feedback is critical for successful learning and academic achievement. Educational psychologists have referred to this ability as self-regulated learning (SRL) and define it as a set of interrelated skills and motivations to control learning. Most theories of SRL hypothesize that both metacognitive skills (e.g., planning, monitoring, and evaluation) and student motivation (e.g., beliefs, goals, and dispositions) interact to determine learning outcomes (Boekaerts & Corno, 2005; Efklides, 2011; Winne, 1995; Zimmerman, 2001, 2011). However, relatively few studies have tested whether a metacognitive instructional intervention designed to improve students' metacognitive knowledge and skills can also improve student motivation more broadly. Moreover, we know of no work that has examined metacognitive, learning, transfer, and motivational outcomes together in a single study.

To address these issues, we conducted an in vivo classroom experiment with students randomly assigned to either a metacognitive instruction and training condition or a problem-solving practice condition. We examined whether students given the metacognitive instruction acquired knowledge and skills about metacognition and whether those skills improved their learning of the target instructional content (physics concepts and problem-solving procedures) as well as new material given later in the semester (experimental design, control of variables strategy). We also examined whether the intervention affected a wide range of motivational constructs specified in Zimmerman's (2011) sociocognitive SRL theory. In the following sections, we review relevant literature that guided the development and theoretical framing of this study.

Metacognition and Motivation in SRL

We focus on the metacognitive skills of planning, monitoring, and evaluation that occur in Zimmerman's (2000, 2011; Zimmerman & Campillo, 2003) SRL phases of forethought, performance, and self-reflection, respectively (see Figure 1). We define planning as identifying the goal of the problem, the critical features, and a set of strategies to move toward that goal; monitoring as keeping track of one's current state and progress moving toward the goal; and evalu-

---

Cristina D. Zepeda and J. Elizabeth Richey, Department of Psychology, Learning Research and Development Center, University of Pittsburgh; Paul Ronevich, Pittsburgh Science and Technology Academy, Pittsburgh, Pennsylvania; Timothy J. Nokes-Malach, Department of Psychology, Learning Research and Development Center, University of Pittsburgh.

This research was supported by Grant SBE 0836012 from the National Science Foundation to the Pittsburgh Science of Learning Center. No endorsement should be inferred. We thank research assistant Sarah Honsaker for her help in coding the data.

Correspondence concerning this article should be addressed to Cristina D. Zepeda, Learning Research and Development Center, 3939 O'Hara Street, University of Pittsburgh, Pittsburgh, PA 15260. E-mail: cdz7@pitt.edu

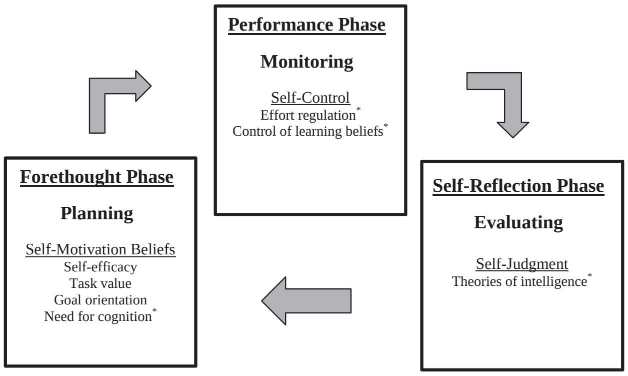

Figure 1. Phases and subprocesses of self-regulated learning (SRL). Asterisks denote constructs we have added to the model based on related literature. Adapted from Zimmerman and Campillo's "Motivating self-regulated problem solvers," in *The psychology of problem solving* (p. 239) edited by J. Davidson & R. J. Sternberg, 2003, New York: Cambridge University Press. Adapted with permission.

ating as assessing one's solution and determining whether it satisfies the goal as well as reviewing which strategies worked best. Zimmerman's model provides a theoretical framework for determining which motivational constructs might be affected by a metacognitive intervention targeting these skills, including aspects of self-motivation beliefs, self-control, and self-judgment. In Figure 1 we have emphasized the metacognitive skills hypothesized to occur during each SRL phase (in bold) and the hypothesized motivational constructs associated with each phase. In addition to Zimmerman's self-motivation beliefs we have added motivational constructs hypothesized to be closely related to the later phases of the model. For example, students' beliefs about effort regulation and control of learning are likely to be associated with their self-control while students' theories of intelligence may affect their self-judgment.

Although a number of studies have explored possible relationships between metacognition and motivation, the majority of empirical investigations exploring those connections have used correlational and quasi-experimental designs (e.g., Ford, Smith, Weissbein, Gully, & Salas, 1998; Somuncuoglu & Yildirim, 1999). The experiments that have used metacognitive interventions to test the relationship typically assess only one or two motivational outcomes, making it difficult to determine whether there are wide-ranging changes in student motivation or just changes to a few constructs.

Our investigation addresses these issues in multiple ways. First, by conducting an experiment with a randomized control, we can test the causal effects of a metacognitive intervention on motivational outcomes. Second, because our intervention targets metacognitive skills used across all three phases of SRL, we expect it to have a broad impact on student motivation, as the hypothesized and cyclical relationships in Zimmerman's SRL model would suggest. To assess the breadth of motivational changes produced by metacognitive training, we included a number of motivational constructs including self-efficacy (Pintrich, Smith, Garcia, & McKeachie, 1991), task value (Pintrich et al., 1991), achievement goal orientations (Elliot & Murayama, 2008), need for cognition (Cacioppo, Petty, Feinstein, & Jarvis, 1996), effort regulation and control of learning beliefs (Pintrich et al., 1991), and theories of intelligence (Dweck, 1999).

We briefly discuss the way (or ways) in which each motivational construct might be affected by the intervention, based on the metacognitive skill most closely associated with it. Some hypothesized effects have more prior empirical support whereas others are more exploratory. While this work contributes to the literature by testing the relationship between a metacognitive intervention and a number of specific motivational outcomes, we do not test the relationship between distinct components of metacognition and different motivational constructs. Zimmerman's phases are thought to be cyclical, and therefore, it would be difficult to identify which metacognitive skill led to a specific motivational outcome. Future work exploring these relationships will be needed to test the underlying mechanisms at play.

Planning

Planning skills serve as domain-general knowledge that can be applied to solve new problems. Having knowledge of these skills should increase students' self-efficacy, defined as confidence in their capabilities to solve such problems, because it suggests a set of strategies to apply when students might otherwise feel unsure about how to approach a new problem (Pajares, 2008; Zimmerman, Bonner, & Kovach, 1996). Students' self-efficacy is thought to relate to their value of a particular task, defined as the degree to which they believe that the task is interesting, important, and useful, as past work has identified positive correlations between perceived confidence and task value (Jacobs, Lanza, Osgood, Eccles, & Wigfield, 2002). Some prior research also suggests that value is positively related to the use of constructive strategies, cognitive engagement, and mastery-approach goal orientation, defined as aiming to fully understand material (Meece, Blumenfeld, & Hoyle, 1988; Nolen, 1988). If students can more successfully solve new problems and come to value such tasks, this could lead them to a have a have a higher need for cognition, defined as a desire to experience more complex or challenging thinking (Cacioppo et al., 1996).

Pintrich (2000) argues that achievement goal orientations play an important role in self-regulated learning because learners' goals serve as criteria by which they can evaluate and regulate their progress. Students with a mastery-approach goal orientation engage in competence-related activities to improve their understanding (Elliot & Murayama, 2008), and this type of goal has been related to successful self-regulated learning (Pintrich, 1999). Since metacognitive interventions teach students different skills to improve their understanding, students who learn about metacognition might be more likely to adopt and endorse mastery-approach goals to create consistency between their goals, knowledge, and behaviors.

Monitoring

If students are able to monitor their progress toward a goal, they may also be better equipped to make decisions about how to manage their available resources, including their effort. Being aware of their progress may also make them more willing to apply effort, which can be measured by the effort regulation scale (Pintrich et al., 1991). Zimmerman and Martinez Pons (1988) found that students' self-regulated learning strategies were related to their efforts to learn. Relatedly, improving students' ability to monitor their task performance might also make them more aware of their own control of learning, which can be measured by the control of learning beliefs scale (Pintrich et al., 1991).

Evaluating

Part of the evaluation process involves making causal attributions for success or failure on a task. Learners with poor evaluation skills may be more likely to attribute success or failure to their personal characteristics (e.g., having ability in a certain domain or not), while learners with more accurate evaluation skills and awareness may be better equipped to identify learning behaviors that led to a given outcome. If students' beliefs of intelligence are informed by their task evaluation skills, then improving those skills might promote endorsement of incremental theories of intelligence—the belief that intelligence is malleable and not fixed (Dweck, 1999). It is also possible that students might adopt an incremental view of intelligence after learning that evaluation skills can be changed with instruction and practice.

Metacognitive Interventions: Transfer and Future Learning Measures

Metacognitive skills have also been hypothesized to facilitate knowledge transfer and preparation for future learning (Schraw, Dunkle, Bendixen & Roedel, 1995; Schraw & Nietfeld, 1998; Veenman, Elshout & Meijer, 1997; Veenman & Verheij, 2001; Veenman, Wilhelm, & Beishuizen, 2004; Wolters & Pintrich, 1998). However, little empirical work has tested these hypotheses. Through the use of metacognitive skills, students can acquire both procedural (Berardi-Coletta, Buyer, Dominowski, & Rellinger, 1995) and declarative knowledge (Cross & Paris, 1988). The type of knowledge a learner acquires is important as it has implications for when and how that knowledge transfers (Nokes, 2009; Nokes-Malach & Mestre, 2013). Proceduralized knowledge (i.e., knowing how to do a task) typically facilitates near transfer to problems that have surface features and structures identical or similar to prior problems, whereas declarative knowledge (i.e., knowing descriptive information about a task) can support performance on far-transfer problems that have different surface features but similar structures (Nokes, 2009; Nokes-Malach & Mestre, 2013; Nokes & Ohlsson, 2005). Adopting Barnett and Ceci's (2002) transfer framework, we define near transfer of content knowledge as the execution of prior procedures or the recall and application of prior concepts to familiar problem features and far transfer of content as the recall and application of prior concepts and principles to new problem features.

We also distinguish between two sources of knowledge that may support transfer from metacognitive interventions: the domain-relevant instruction such as physics content given over the course of the intervention and instruction about metacognitive skills themselves. These two sources of knowledge have different implications for transfer. Domain-relevant instruction should support transfer to the degree that the knowledge acquired is abstract and applicable to new problems or questions (Barnett & Ceci, 2002). Metacognitive instruction should support different types of transfer, including near, far, and preparation for future learning, through the application of domain-general metacognitive skills to new problems or learning opportunities. This suggests two important questions about the types of transfer supported by metacognitive interventions: what type of knowledge transfers and what is the source of the knowledge being transferred?

Many past metacognitive interventions have not assessed far-transfer outcomes, and few have taken a rigorous approach to defining levels and types of transfer. However, some studies have shown promising results for both near and far transfer (e.g., Brand, Reimer, & Opwis, 2003; Palincsar & Brown, 1984; Lin & Lehman, 1999). Lin and Lehman (1999) found that students who received metacognitive prompts performed better than other prompt conditions and a control condition on near- and far-transfer assessments. In order to solve the far-transfer problems, the students needed to adapt their previous conceptual understanding to the new problem features.

In the current work, we employed multiple transfer assessments to distinguish between near transfer (structurally similar problems included at the end of each instructional packet) and far transfer (questions included on a conceptual test given after a delay). We also include assessments to differentiate between transfer of domain-relevant content covered in the instructional packets (the end-of-packet transfer problems and questions) and the transfer of domain-general metacognitive skills to new learning opportunities (a preparation for future learning, or PFL, measure). In contrast to the more classical conceptualization of transfer that focuses on whether or not one can use knowledge acquired from instruction to solve novel problems (Barnett & Ceci, 2002), the PFL measure focuses on whether the initial instruction affects what one learns *from subsequent instruction* and how that knowledge is then used to solve new problems (Bransford & Schwartz, 1999). In this study we examined whether students could utilize metacognitive skills in a self-guided learning activity on a novel topic two weeks after the intervention. We discuss the details of these transfer measures in the methods section below.

Present Study

Given this theoretical and empirical backdrop, we tested a self-guided metacognitive intervention that targeted planning, monitoring, and evaluating skills and highlighted how those skills support adaptive problem solving. Meta-analyses of metacognitive interventions (Dignath & Büttner, 2008; Dignath, Buettner, & Langfeldt, 2008; Hattie, Biggs, & Purdie, 1996) were particularly helpful in guiding our design and implementation of the intervention, including the decision to provide instruction on all three skills along with conditional, interactive process knowledge of how they work together. We view this as a critical component of the intervention because if students know the interrelations between planning, monitoring, and evaluating, then they can use that knowledge to effectively adapt their behavior. Without such knowledge, they might become confused about when or how to use the skills or just focus on one skill.

Hypotheses

This study had three central goals. The first goal was to create an easy-to-implement instructional intervention designed to increase students' metacognitive knowledge and skills (i.e., awareness, accuracy, and use). The second was to examine whether the metacognitive intervention affects different motivational constructs hypothesized to relate to metacognition and self-regulated learning. The third was to test whether the intervention results in greater learning and transfer. Figure 2 represents our hypotheses as they relate to each goal.

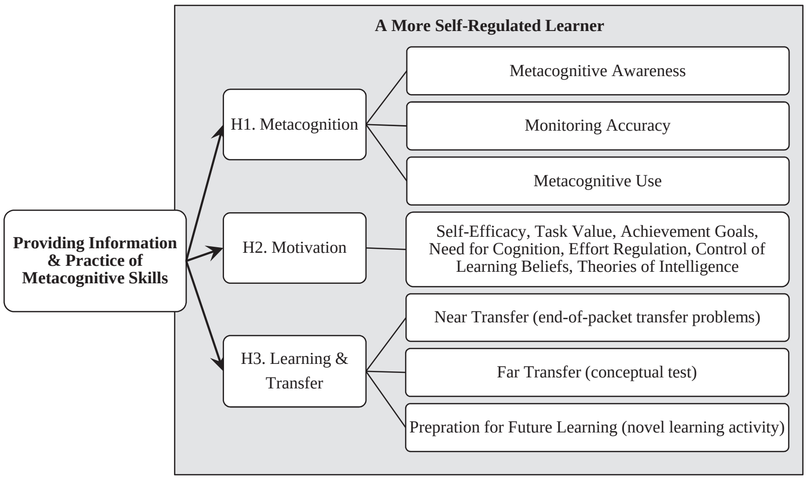

Figure 2. Graphical representation of the hypotheses.

Method

Participants. Forty-nine students from two eighth-grade physics classes at the same urban, public middle school participated in the experiment (23 students in one class and 26 in the other). Three students were dropped from the study because they did not complete the majority of the assessments. Of the remaining 46 students (13 females, 33 males), 44% self-identified as White, 41% as African American, 13% as mixed race, and 2% as Latino(a)/Hispanic.

Design. We used a between-subjects, preposttest design with students randomly assigned to either the experimental ($n = 23$) or control condition ($n = 23$). There were no differences between the conditions across any of the demographic variables including students' gender, $\chi^2(1, N = 46) = .11, p = .74$, or race, $\chi^2(3, N = 46) = 1.05, p = .79$. All students had the same teacher, who was blind to the condition assignment of the students. Students participated as a regular part of their classroom instruction and received participation points for completing the various activities. The teacher distributed all materials in packet form and made sure that students were not looking at each other's packets, working together, or asking each other for help. The packets had a noninformative group label on the cover page so the teacher could distribute them without learning the condition assignments. We provided a script for the teacher to follow and met regularly with him to ensure treatment fidelity.

Due to an implementation error, one student from each condition completed instructional materials from the opposite condition for one of the eight learning packets. We decided to include these students in the data analysis as they received the majority of training materials for their conditions and contribute to increasing power, albeit with a weaker dosage of the intervention.[1] Figure 3 shows an overview of the experiment design, materials, and procedure.

Intervention materials. We used puzzle problems for the first round of the intervention to avoid the potentially distracting knowledge demands that science content might have introduced. This enabled us to emphasize problem-solving and metacognitive skills over the problem content. The puzzles consisted of spatial and verbal insight problems, riddles, rebus word problems, and simple math problems (see online Supplemental Materials for an example of each problem type). All students were given the same initial problem, followed by a hint, another opportunity to solve it, and the solution. At the end of each packet, all students were given a transfer problem with a similar structure to that of the initial problem. For example, the initial and transfer problems in the first packet were spatial insight problems that focused on the manipulation of shapes. By using the same problem type, we could see if training improved immediate performance on near-transfer problems. Following the transfer problem, students in both conditions received a packet quiz that assessed their declarative knowledge of the targeted metacognitive skill. The quiz consisted of an open-ended question and multiple-choice questions (see online Supplemental Materials).

Packets 1–4: experimental. In the first packet, students studied an explanation of planning, reviewed worked examples of plans, responded to questions about their own planning activities, and created a plan to solve a new problem (see Table 1). In the second packet, students studied an explanation of monitoring, reviewed and analyzed fictional students' attempts to solve problems, and responded to questions about their own monitoring activities (see Table 1). In the third packet, students studied an explanation of evaluating and responded to prompts to evaluate their solutions (see Table 1). In the fourth packet, students reviewed descriptions of planning, monitoring, and evaluating, read about how to integrate the three skills when problem solving, and responded to prompts and questions targeting all three skills (see Table 1). Unlike the other packets, Packet 4 concluded with two transfer problems instead of one. See the online Supplemental Materials for the detailed definitions presented in each of the four packets. On average the experimental condition completed 93% of the packet materials ($SD = .07$).

---

[1] The same general pattern of results was observed when we remove these students from the analyses. When we exclude the two students the only difference is that the effect of condition on the Force Concept Inventory (FCI) becomes marginal.

Week(s)		Control Condition	Experimental Condition
1		**Pretests**	
	Day 1	• Conceptual Knowledge of Physics: Force Concept Inventory (45 minutes; Hestenes et al., 1992)	
6			
	Day 2	• Metacognition: Metacognitive Awareness Inventory (20 minutes; Schraw & Dennison, 1994)	
	Day 3	• Task Value, Control of Learning Beliefs, Self-Efficacy and Effort Regulation: Motivated Strategies for Learning Questionnaire (20 minutes; Pintrich et al., 1991)	
	Day 4	• Achievement Goals: Achievement Goal Questionnaire – Revised (Elliot & Murayama, 2008) • Beliefs about Intelligence: Theories of Intelligence (Dweck, 1999) – 20 minutes	
	Day 5	• Engagement in Thinking: Need for Cognition (20 minutes; Cacioppo et al., 1996)	
13-15		**Round One: Puzzles**	
		• More puzzle problems • Received a worked example of the initial problem • No direct instruction or worked examples of metacognition • Practiced solving problems without metacognitive prompts • Received solutions to all problems	• Fewer puzzle problems • Received a worked example of the initial problem • Direct instruction and worked examples of metacognition • Practiced solving problems with metacognitive prompts • Received solutions to all problems
	Day 6	Problem Solving 1 (45 minutes)	**Planning** and Problem Solving 1 (45 minutes)
	Day 7	Problem Solving 2 (45 minutes)	**Monitoring** and Problem Solving 2 (45 minutes)
	Day 8	Problem Solving 3 (45 minutes)	**Evaluating** and Problem Solving 3 (45 minutes)
	Day 9	Problem Solving 4 (45 minutes)	**Integration** and Problem Solving 4 (45 minutes)
15-17		**Round Two: Physics**	
		• More physics problems • Received a worked example of the initial problem • No direct instruction or worked examples of metacognition • Practiced solving problems without metacognitive prompts • Received solutions to all problems	• Fewer physics problems • Received a worked example of the initial problem • Direct instruction and worked examples of metacognition • Practiced solving problems with metacognitive prompts • Received solutions to all problems
	Day 10	Problem Solving 5 (45 minutes)	**Planning** and Problem Solving 5 (45 minutes)
	Day 11	Problem Solving 6 (45 minutes)	**Monitoring** and Problem Solving 6 (45 minutes)
	Day 12	Problem Solving 7 (45 minutes)	**Evaluating** and Problem Solving 7 (45 minutes)
	Day 13	Problem Solving 8 (45 minutes)	**Integration** and Problem Solving 8 (45 minutes)
18-19		**Posttests**	
	Day 14	• Metacognition: Metacognitive Awareness Inventory (20 minutes; Schraw & Dennison, 1994)	
	Day 15	• Task Value, Control of Learning Beliefs, Self-Efficacy and Effort Regulation: Motivated Strategies for Learning Questionnaire (20 minutes; Pintrich et al., 1991)	
	Day 16	• Achievement Goals: Achievement Goal Questionnaire – Revised (Elliot & Murayama, 2008) • Beliefs about Intelligence: Theories of Intelligence (Dweck, 1999) – 20 minutes	
	Day 17	• Engagement in Thinking: Need for Cognition (20 minutes; Cacioppo et al., 1996)	
20			
	Day 18	• PFL Task: Control of Variables Strategy Activity (45 minutes)	
22			
	Day 19	• Conceptual Knowledge of Physics and Metacognitive Monitoring: Force Concept Inventory (50 minutes; Hestenes et al., 1992) with Confidence Ratings	
29		**Delayed Posttests**	
	Day 20	• Preparation for Future Learning Task – Retention Rate: Delayed Transfer Test of the Control of Variables Strategy Activity (25 minutes)	
30			
	Day 21	• Metacognition Reflection: Declarative Knowledge and Utility Defined (30 minutes)	

Figure 3. Outline of the procedure by condition as indicated by the first row.

Packets 1–4: control. The control materials also consisted of puzzle problems (see online Supplemental Materials for examples). The packets did not include any instruction on planning, monitoring, or evaluating, but instead instructed students that they could improve their general problem-solving skills by working through the packets. The initial problem for each packet was the same as in the experimental materials. Following the initial problem, packets were divided into sections of problems. At the end of each section, students were given solutions and encouraged to check their answers before continuing on to the next section of problem solving. Piloting work revealed that the problems within each packet differed in the amount of time it took to complete them; consequently, the first packet had seven problems, the second had 13, the third had 16, and the fourth had seven. We gave the control condition a sufficient amount of problems to ensure they did not finish before the experimental condition. The control packets concluded with the same transfer problems and packet quizzes as the experimental packets. On average the control condition completed 87% of the packet materials ($SD = .12$).

In round two of the intervention, we integrated the instruction from round one into a series of physics problems that were adapted from the students' physics textbook (Hsu, 2005). Each packet focused on different physics concepts about which the students had previously received instruction. The first packet consisted of problems that required students to calculate the average speeds of two objects. The second packet contained conservation of momentum problems. The third packet contained problems that required students to apply Newton's second law to calculate speed, acceleration, and distance for a single falling object. The fourth packet required students to apply Newton's second law

Table 1
Definition, Questions, and Practice Procedure for the Experimental Group

Packet	Definition	Example Prompting Questions	Practice Procedure
Planning	Understanding the problem, identifying the goal, and strategizing to create a plan.	Does the problem make sense? Are there similar problems you have encountered that can help you decide which strategies to use?	Before starting to solve a problem, students wrote how they planned to solve the problem.
Monitoring	Thinking about where one is on the path to solving the problem in order to monitor progress toward the goal.	Are you close to the solution? What step are you on? Are your reasons for taking each step moving you closer to the goal?	While problem solving, students stopped to check their progress, looking for any errors, and answered a progress check: 1. Do you know what step you are on? 2. Why are you on that step? 3. How close to the solution do you think you are? 4. Is your approach working? 5. How do you know you are on the right path? 6. Are you going to start over with a different strategy?
Evaluating	Comparing the answer to the problem's goal, and looking to see which strategies worked best to evaluate the solution.	Does my solution reach the problem's goal(s)? Does my solution make sense?	After students solved the problem, they checked their solution to make sure it made sense by answering an evaluation check: 1. What was the goal(s) of the problem? 2. How does your solution meet the goal(s)? 3. Is your solution correct? 4. Why?

to calculate weight, acceleration, and time for pairs of falling objects.

Similar to the first round, each of the packets began with an initial problem (see Table 2), followed by a hint with an additional opportunity to solve that problem, followed by a solution and explanation. The control group then received several sets of isomorphic physics problems with solutions given at the end of each set, while the experimental group received the same direct instruction and practice of metacognitive skills from round one. Again, students in the control group received sufficient problems to make sure they would not finish before the experimental condition. The first packet had five problems, the second had seven, the third had nine, and the fourth had four. For the experimental condition, the first packet emphasized planning, the second monitoring, the third evaluation, and the fourth integrated the three skills. The instructional materials were identical to the first round of the intervention, except that the planning packet for the experimental condition required students to create a plan *before* solving the initial problem. At the end of each packet, both conditions received a packet quiz identical to the one given in the first round of the intervention to assess their declarative knowledge of the targeted metacognitive skill. On average the control condition completed 76% of the packet materials (*SD* = .18) and the experimental condition completed 85% of the packet materials (*SD* = .08). Although the control condition completed proportionally less of the packet materials, on average they solved four more problems per packet than the experimental condition.

Scoring of packets. The initial and transfer problems within the intervention packets were used to evaluate how well the students were able to transfer what they had learned in the packets to problems that required similar problem-solving procedures. To investigate performance on these near-transfer problems, we scored each transfer problem from the first round as either incorrect (0) or correct (1). We scored each transfer problem from the second round as either incorrect (0), partially correct if students had either the correct number or the correct unit (.5), or completely correct (1) if they had both the correct number and unit. Since the integration packet for each round had two transfer problems, we

Table 2
An Example of the Initial, Isomorphic, and Near-Transfer Problems for the First Packet, Round 2

Initial	Isomorphic	Transfer
Sam and Mitch wanted to know who would win if they raced against each other. They know that during a 6.2-mile race, Sam runs at a constant speed of 6 miles per hr. They also know that during a 10-kilometer race, Mitch finishes in 50 min. (1 kilometer = .62 mile) Who would win if they raced against each other? Explain how you arrived at your answer.	Marcus can swim 40 laps in 50 min, while John can swim at a constant rate of 1.75 miles per hr. (32 laps = 1 mile) If Marcus and John race in a swim meet, who will win?	Sarah rides her bicycle at 10 miles per hr for 3 hr. She stops to rest for 1 hr and then continues her ride. For the next hour, she rides at a speed of 20 miles per hr. During this ride, what was the rider's average speed for the 5-hr period?

calculated accuracy on those packets by dividing students' scores by two. Transfer problems from round two also had multiple questions, and each of these questions was scored for accuracy and then divided by the total by the number of questions per problem. After these calculations, the scores for each round were summed for a total of four points each (1 point for each packet) and divided by four to give a proportion of accuracy. The transfer scores ranged from 0 to 0.88 for both rounds.

Scoring of packet quizzes. For each of the eight packets, we scored the quiz to evaluate students' declarative knowledge about the metacognitive skills. Within the quiz there were two types of questions, open-ended and multiple-choice. The open-ended questions asked students to state important steps to take while problem solving. We gave one point for every key term or phrase that matched or was synonymous to a metacognitive concept from the corresponding experimental packet (see Table 3). For example, if a student wrote the words "plan" and "create a goal" on the planning packet then she would receive two points. If a student used a synonymous phrase such as "think about how to approach the problem" and "figure out what the problem is asking for," she would also receive two points. We calculated the number of points each participant received across all four packets for each round and divided by the total number of points possible (five for planning, four for monitoring, four for evaluating, and 13 for integrating). Two raters independently coded all the open-ended responses and reached adequate reliability ($\kappa > .70$). The two raters discussed and resolved all disagreements. The open-ended scores ranged from 0 to 0.42 for the first round and 0 to 0.38 for the second.

There were also two types of multiple-choice questions. The first type had a single best answer and was scored as correct (1), partially correct (.5), or incorrect (0). The second type had multiple correct responses (checklist) and was scored as the proportion of correctly selected and correctly rejected alternatives out of the total number of alternatives (scores ranged between 0 and 1 per problem). To calculate the accuracy score per round, we divided the sums by the total number of points possible for that round (14). The multiple-choice scores ranged from 0.16 to 0.77 for the first round and 0.20 to 0.71 for the second.

Pretests and Posttests

Conceptual knowledge test. Students' conceptual understanding of Newtonian physics was measured with Hestenes, Wells, and Swackhamer's (1992) Force Concept Inventory (FCI). The FCI consists of 29 multiple-choice questions, each describing an applied situation or scenario in which the student must qualitatively reason about the relevant physics concepts to determine the correct answer ($\alpha = .69$). Some of the concepts targeted by the FCI were covered in the packet problems, while others were not. This assessment provides a strong test of conceptual understanding as students had to distinguish the correct answer from four alternatives that capture common physics misconceptions. The FCI was given before and after the intervention, with both versions identical except that the posttest asked students to rate their confidence after each answer on a scale, ranging from one (*not confident at all*) to five (*very confident*). This confidence rating enabled us to examine whether the intervention improved students' metacognitive accuracy in assessing their physics knowledge on a difficult conceptual test.

Scoring. The FCI was analyzed to evaluate whether the intervention affected how well students performed on a far-transfer task. We scored each question as either correct or incorrect, and then calculated the proportion of questions students answered correctly by dividing the sum of correct responses by the total number of items. On the pretest, scores ranged from 0.14 to 0.53 and on the posttest scores ranged from 0.18 to 0.86.

We also used the confidence ratings ($\alpha = .92$) and accuracy scores from the posttest to assess metacognitive monitoring using absolute accuracy and two forms of relative accuracy (for a distinction see Schraw, 2009, Table 4). Absolute accuracy (sometimes referred to as calibration) measures the difference between a student's confidence judgment and the corresponding accuracy

Table 3
Scoring for the Open-Ended Questions

Instructional Packet	Phrases Awarded Points
Planning	1. Mentions the process of creating a plan 2. Mentions creating a goal, using goals, goal(s) 3. Mentions making sure they understand the problem/what is being asked of them 4. Mentions using strategies or processes 5. Mentions the rules of the problem, the limitations of the problem
Monitoring	1. Mentions monitoring or staying on track 2. Specifies monitoring in terms of the solution, goal, or understanding 3. Mentions providing reasoning 4. Mentions understanding why you complete a step or the logic behind the step
Evaluating	1. Mentions checking and comparing or assessing answers 2. Mentions checking to make sure your solution makes sense 3. Mentions that there are multiple solutions and/or strategies and you have to figure out the best one 4. Mentions helping them in the future
Integrating	All of the above

Table 4

Formulas Used to Calculate Absolute Accuracy, Relative Accuracy: Gamma and Relative Accuracy: Discrimination

Type of Equation	Equation
Absolute accuracy	$\frac{1}{N}\sum_{i=1}^{N}(c_i - p_i)^2$
Relative accuracy: Gamma	$\frac{N_S - N_D}{N_S + N_D}$ where N_S is the number of concordant pairs and N_D is the number of discordant pairs
Relative accuracy: Discrimination	$\frac{1}{N}\left[\sum_{i=1}^{N_c}(c_{i\text{correct}}) - \sum_{i=1}^{N_i}(c_{i\text{incorrect}})\right]$

score. This measure assesses metacognitive precision. A smaller deviation indicates greater accuracy. Gamma is a form of relative accuracy (sometimes referred to as resolution) and measures the agreement between a student's confidence judgments and accuracy on each item in relationship to other items (Goodman & Kruskal, 1954). This measure assesses whether a student is consistent with her metacognitive judgments regardless of absolute accuracy. Discrimination is also a form of relative accuracy, but this measure assesses whether a student can differentiate between correct versus incorrect performance. Positive discrimination scores indicate that a student has higher confidence in correct versus incorrect items, whereas negative discrimination scores indicate that a student has higher confidence in incorrect versus correct items. Using multiple measures can provide greater insight into the monitoring process because different measures could be differentially affected by underlying mechanisms such as feedback and motivation (Schraw, Kuch, & Gutierrez, 2013). For instance, prior research has shown that absolute accuracy and relative accuracy are unrelated and that absolute accuracy tends to be sensitive to task difficulty and student ability whereas Gamma is not (Maki, Shields, Wheeler, & Zacchilli, 2005; Kelemen, Frost, & Weaver, 2000).

Metacognition. Students' beliefs about metacognition were assessed with three components of the Metacognitive Awareness Inventory (MAI) including seven items on planning ("I set specific goals before I begin a task"), seven items on monitoring ("I ask myself questions about how well I am doing while I am learning something new"), and six on evaluating ("I ask myself how well I accomplish my goals once I'm finished"; Schraw & Dennison, 1994). Students rated how characteristic they thought each statement was of themselves on a five-point Likert scale from one (*extremely uncharacteristic*) to five (*extremely characteristic*).[2] The MAI scores had high internal reliability at pretest ($\alpha = .85$) and posttest ($\alpha = .92$).

Self-efficacy, value, effort regulation, and control of learning beliefs. Students completed four components of the Motivated Strategies for Learning Questionnaire (MSLQ; Pintrich et al., 1991) including: (1) seven items for self-efficacy, for example, "I'm confident I can understand the basic concepts taught in this class," (2) six items for task value, for example, "I think the material in this class is useful for me to learn," (3) four items for strategies of effort regulation, for example, "I work hard to do well in this class even if I don't like what we're doing," and (4) three items for control of learning beliefs, for example, "If I don't understand the class material, it is because I didn't try hard enough." Students rated how much they agreed or disagreed with each of the items on a seven-point Likert scale from one (*strongly disagree*) to seven (*strongly agree*). Scores from the measures of self-efficacy ($\alpha = .87$) and task value ($\alpha = .78$) had adequate reliability at the pretest whereas scores from the measures of effort-regulation ($\alpha = .65$) and control of learning beliefs ($\alpha = .45$) had low reliability. At posttest, scores from the self-efficacy ($\alpha = .87$), task value ($\alpha = .82$), and effort-regulation ($\alpha = .71$) measures had adequate reliability, but scores from the control of learning beliefs measure ($\alpha = .63$) had low reliability. Due to the low reliability of scores from the control of learning beliefs and effort regulation measures, we do not analyze these results further.

Goal orientations. Students' achievement goal orientations were assessed using the Achievement Goal Questionnaire – Revised (AGQ-R; Elliot & Murayama, 2008). The questionnaire evaluates students' endorsement of mastery-approach ("My aim is to completely master the material presented in class"), mastery-avoidance ("My aim is to avoid learning less than I possibly could"), performance-approach ("My aim is to perform well relative to other students"), and performance-avoidance goals ("My aim is to avoid doing worse than other students") with three items per orientation. Students rated how much they agreed or disagreed with each item on a seven-point Likert scale from one (*strongly disagree*) to seven (*strongly agree*). The scores from the mastery-approach, performance-approach, and performance-avoidance measures had high reliability at pretest (ranging from $\alpha = .89$ to $\alpha = .91$) and posttest (ranging from $\alpha = .81$ to $\alpha = .90$). Mastery-avoidance scores had low reliability at pretest ($\alpha = .38$) and high reliability at posttest ($\alpha = .81$), and it was therefore excluded from further analyses.

Theories of intelligence. We used the six-item Implicit Theories of Intelligence Scale for Children – Self-Form to assess students' beliefs about intelligence (Dweck, 1999). This survey measures the degree to which students think intelligence is malleable (incremental) or fixed (entity). Students rated how much they agreed or disagreed with three entity theory and three incremental theory statements on a seven-point Likert scale from one

---

[2] Although the original survey comprised several other metacognitive components, we assessed only planning, monitoring, and evaluation because our intervention explicitly targeted these three components.

(*strongly disagree*) to seven (*strongly agree*; pretest: $\alpha = .94$, posttest: $\alpha = .89$). Entity questions were reverse scored to calculate a single score across the six items, with low values indicating greater endorsement of entity beliefs and higher values indicating greater endorsement of incremental beliefs, for example, "No matter who you are, you can change your intelligence a lot."

Need for cognition. Students' enjoyment of and desire to engage in cognitive activities that require large amounts of thinking and understanding were assessed by the 18-item Need for Cognition Scale, for example, "I like to have the responsibility of handling a situation that requires a lot of thinking" (NCS; Cacioppo et al., 1996). Students rated how characteristic or uncharacteristic each item was of themselves using a five-point Likert scale from one (*extremely uncharacteristic*) to five (*extremely characteristic*). The NCS scores had high reliability at pretest ($\alpha = .90$) and posttest ($\alpha = .91$).

Posttests and Delayed Posttests

Preparation for future learning—Control of variables strategy. We created a preparation for future learning activity based on Siler, Klahr, and Price's (2012) "control of variables strategy" (CVS) instructional materials. These materials are designed to teach students how to create valid experiments, which is critical for reasoning about variables and confounds when designing experiments and interpreting results (Chen & Klahr, 1999). Although CVS is a cornerstone of experimental design, many students struggle to understand and use it in science class. For our purposes, this activity served as a preparation for future learning measure as the students could apply their newly acquired metacognitive knowledge and skills to learn from self-guided materials about CVS. Critically, the PFL activity differed from problems covered in the intervention and class in terms of both content and structure, which allowed us to gauge how the intervention prepared students to learn a novel concept and skill.

During the PFL activity, students completed pretest, learning, and posttest packets. The pretest asked students to design four valid experiments and evaluate whether an experiment was correctly designed. Following the pretest packet, students worked through a nine-page learning packet that explained how to create a valid experiment through an analysis of a well-designed and poorly designed pair of experiments. Students were given a series of questions that asked them to explain why the experiments were well or poorly designed, and they were shown how to correct the confounded experiment.

After the learning packet, the students completed a posttest packet with the same four design questions from the pretest, two new questions asking them to evaluate an experimental design, and three new questions asking them to generate an experiment testing whether a specific variable affected an outcome (see online Supplemental Materials). After each of the new questions, students were asked to explain their answers, with the option of saying they guessed after each of the "generate" questions.

Approximately 2 months after completing the PFL activity, students received a delayed transfer test on creating valid experiments. Similar to the posttest, the delayed transfer test contained two new questions asking students to evaluate an experimental design and two new questions asking them to generate an experiment testing a specific variable. Students were again prompted to explain their answers, with the option of saying that they guessed after each of the "generate" questions.

Scoring. To investigate the effect of condition on performance on the PFL activity, we calculated the number of questions answered correctly for the pretest, learning, posttest and delayed transfer packets as well as the number of times they guessed during the activity. The five pretest items were coded as either correct or incorrect. Pretest scores ($\alpha = .92$) were summed for a maximum of five points and divided by five to calculate the proportion of accuracy, ranging from 0 to 1. We combined the scores on the learning packet (two questions), the posttest packet (nine questions), and the delayed transfer packet (four questions) to provide a simple measure of PFL and to have enough data points for parametric analyses. PFL scores ($\alpha = .86$) were summed for a maximum of 15 points and divided by 15 to give a proportion of accuracy, ranging from 0.13 to 1. Students also had six opportunities to say they guessed during the activity (two per phase during learning, posttest, and transfer phases). We calculated the number of times they reported guessing across the six problems and divided it by six to calculate the proportion of reported guessing, which ranged from 0 to 0.67.

Reflection survey. Students received an opportunity to reflect on their understanding and the utility of the training intervention a week after completing the delayed PFL transfer test. Both groups answered a multiple-choice question about important steps of problem solving (i.e., metacognitive skills) and a question about whether they found the problem-solving activities useful. The metacognitive skills question was scored dichotomously as either correct or incorrect and the usefulness of the activities was scored as useful (1) or not (0). They also answered questions using a Likert scale from one to five about the frequency (never, rarely, sometimes, a lot, all the time) with which they generally used metacognitive skills across different contexts as well as the frequency with which they used each skill during problem solving.

Procedure

This experiment took place over 30 weeks throughout the academic school year with 21 days of activities. The procedure is divided into three time frames: before the intervention (pretest), the intervention, and after the intervention (posttest). For an overview of the experimental procedure, see Figure 3. The teacher covered motion and forces in the first part of the year and had completed all relevant physics content before the intervention began.

Before the intervention. Twelve weeks before the intervention, students were given the FCI (Hestenes et al., 1992) to measure their conceptual knowledge in physics. Students were told to try their best and were given 45 min to complete the test. At the time the pretest was given, the teacher had covered some of the content on the FCI (e.g., basic motion and Newton's First Law) but not all of it. This timing was selected so that students would have some context for understanding the questions but also to identify misconceptions from everyday experience. Once the FCI was completed, the teacher covered more of the material from the FCI (e.g., Newton's Second and Third Laws). Nine weeks later students completed a series of self-report surveys designed to assess key aspects of their metacognition, learning strategies, and motivation during a 4-day period, taking approximately 20 min each day. To ensure that the students took the survey materials seriously, the teacher told them he would use the survey materials to better understand how they were learning to improve his teaching.

Intervention. After the pretests, the teacher administered the instructional intervention over the course of 4 weeks. The intervention consisted of eight learning packets for both the experimental and control conditions. Each week students completed one packet every Tuesday and Thursday. Students were given 45 min in class to work on each packet. The packets were self-guided but had a few subcomponents that were timed by the teacher (see online Supplemental Materials for packet timings). In total, students spent 6 hours on the intervention materials.

After the intervention. The week after the intervention, students were given the same set of self-report surveys and the PFL scientific reasoning task. Similar to the pretest procedure, students were given the surveys in the same order across 4 days, with 20 min each day to complete them. On the day after the last set of posttest surveys, students were given the PFL activity. This activity was administered in packet form and included pretest, learning, and posttest packets. Students were given 45 min total with 15-min blocks for each packet.

A week and a half later, students were given the FCI posttest. Students were given 50 min to complete the test in class. Seven weeks after that, students spent approximately 20 min completing a delayed transfer test on the PFL scientific reasoning task. A week later, students were given 20 min to complete a short-answer and multiple-choice assessment of their use of metacognition including planning, monitoring, and evaluation.

Results

To test the effect of instructional condition on posttest performance, we conducted analyses of covariance (ANCOVAs) controlling for students' pretest scores. This approach provides a test of instructional condition on metacognition, learning, transfer, and motivational outcomes while controlling for students' initial scores on the construct of interest. We implemented this approach for all analyses with the exception of the packet quizzes, confidence ratings, and metacognitive reflections, as we did not have any pretest measures for these materials. Instead we used analyses of variance (ANOVAs) and chi-square tests, where appropriate. We set the alpha level at .05 and report effects for *p* values less than .05 (Keppel & Wickens, 2004). Assumptions underlying ANCOVAs and ANOVAs were tested. For cases in which the assumptions were not met, we note them below. For all effects, we report effect sizes (Cohen's *d* or partial eta squared, η_p^2). We interpret effects as small when $\eta_p^2 < 0.06$ or $d < 0.2$, medium when $0.06 < \eta_p^2 < 0.14$ or $0.3 < d < 0.8$, and large when $\eta_p^2 > 0.14$ or $d > 0.8$ (see Cohen, 1988; Olejnik & Algina, 2000). See Table 5 for a summary of all the results with descriptives and effect sizes.

Performance on Learning Packets

Problem solving. To examine whether there was an effect of condition on near-transfer problems we used a 2 × 2 mixed-design ANOVA with a within-subjects factor of round (1 vs. 2) and a between-subjects factor of condition (experimental vs. control). Analyses revealed a medium effect of condition, $F(1, 44) = 6.13$, $p = .02$, $\eta_p^2 = .12$, with the experimental condition performing better than the control condition across both rounds. There was no effect of round, $F(1, 44) = 0.43$, $p = .52$, $\eta_p^2 = .01$, nor was there an interaction between round and condition, $F(1, 44) = 0.01$, $p = .92$, $\eta_p^2 = .00$ (see Figure 4).

Packet quizzes. To investigate the effect of condition on the open-ended questions for rounds one and two, we conducted a 2 × 2 mixed-design ANOVA with round of the intervention as a within-subjects factor and condition as a between-subjects factor. Analyses revealed a large effect of condition, $F(1, 44) = 11.48$, $p = .001$, $\eta_p^2 = .21$, with the experimental condition receiving more points than the control condition on the open-ended questions across both rounds. There was no effect of round, $F(1, 44) = 0.001$, $p = .97$, $\eta_p^2 = .00$, nor was there an interaction between round and condition, $F(1, 44) = 0.01$, $p = .92$, $\eta_p^2 = .00$ (see Figure 4).

We also tested the effect of condition on the multiple-choice questions for rounds one and two, using a 2 × 2 mixed-design ANOVA with a within-subjects factor of round and a between-subjects factor of condition. Results revealed a large effect of condition, $F(1, 44) = 10.35$, $p = .002$, $\eta_p^2 = .19$, with the experimental condition performing better than the control condition across both rounds. There was no effect of round, $F(1, 44) = 0.19$, $p = .68$, $\eta_p^2 = .00$, nor was there an interaction between round and condition, $F(1, 44) = 0.83$, $p = .37$, $\eta_p^2 = .02$ (see Figure 4). These results provide evidence that the experimental group acquired declarative knowledge of the metacognitive skills presented in the learning packets.

Self-Report Surveys

Metacognitive Awareness Inventory. The MAI measured students' metacognitive awareness on a scale from one to five, with higher scores indicating higher levels of metacognitive awareness. An ANCOVA was conducted to examine the effect of condition on metacognitive awareness. Analyses revealed a large effect of the covariate, $F(1, 43) = 75.08$, $p < .001$, $\eta_p^2 = .64$, showing that students' MAI pretest scores predicted their posttest scores. There was no effect of condition, $F(1, 43) = 0.83$, $p = .37$, $\eta_p^2 = .02$.[3] Contrary to our predictions, these results show that the metacognitive intervention did not enhance participants' metacognitive awareness as measured by this survey compared to the control.

Task value. Task value measured how much students valued classroom tasks. Testing the effect of condition while controlling for prior task value, an ANCOVA revealed that there was a large effect of the covariate, $F(1, 43) = 48.98$, $p < .001$, $\eta_p^2 = .53$, which indicates that students' pretest scores predicted their posttest scores. There was also a large effect of condition, $F(1, 43) = 8.40$, $p = .006$, $\eta_p^2 = .16$, with the experimental condition placing a higher value on learning the class materials than the control group.

Self-efficacy. Self-efficacy measured how much students believed they were able to accomplish classroom tasks. An ANCOVA revealed that there was a large effect of the covariate, $F(1, 43) = 46.32$, $p < .001$, $\eta_p^2 = .52$, showing that students' pretest scores predicted their posttest scores. There was also a large effect of condition, $F(1, 43) = 8.22$, $p = .006$, $\eta_p^2 = .16$, with the experi-

---

[3] When the planning, monitoring, and evaluation components of the MAI were analyzed separately there was no effect for any subcomponent.

Table 5
Summary of Results

Measure	Targeted Outcome	Experimental			Control			Effect Size	Result Direction
		n	*M* (*SD*)	95% CI	*n*	*M* (*SD*)	95% CI		
Packets: Transfer accuracy (Round 1, Round 2)	Near Transfer	23	0.41 (0.21)	[0.33, 0.50]	23	0.28 (0.21)	[0.19, 0.37]	Medium, $d = 0.72$	E > C
		23	0.43 (0.21)	[0.34, 0.52]	23	0.31 (0.21)	[0.22, 0.39]		
Packets: Open-ended accuracy (Round 1, Round 2)	Metacognition	23	0.14 (0.09)	[0.10, 0.18]	23	0.06 (0.09)	[0.03, 0.10]	Large, $d = 1.0$	E > C
		23	0.14 (0.09)	[0.11, 0.18]	23	0.06 (0.09)	[0.03, 0.10]		
Packets: Multiple-choice accuracy (Round 1, Round 2)	Metacognition	23	0.56 (0.14)	[0.50, 0.62]	23	0.44 (0.14)	[0.38, 0.50]	Large, $d = 0.91$	E > C
		23	0.53 (0.12)	[0.48, 0.58]	23	0.45 (0.12)	[0.40, 0.50]		
MAI	Metacognition	23	3.76 (0.46)	[3.57, 3.96]	23	3.64 (0.46)	[3.45, 3.83]	No effect, $d = 0.26$	N/A
Task value	Motivation	23	5.70 (0.68)	[5.41, 5.98]	23	5.11 (0.68)	[4.83, 5.40]	Large, $d = 0.87$	E > C
Self-efficacy	Motivation	23	5.75 (0.75)	[5.43, 6.06]	23	5.11 (0.75)	[4.79, 5.42]	Large, $d = 0.85$	E > C
Mastery-approach goals	Motivation	23	6.17 (0.74)	[5.85, 6.49]	22	5.11 (1.49)	[4.45, 5.78]	Large, $d = 0.83$	E > C
Performance-approach goals	Motivation	23	5.52 (1.03)	[5.08, 5.95]	22	5.02 (1.03)	[4.58, 5.47]	No effect, $d = 0.49$	N/A
Performance-avoidance goals	Motivation	23	5.26 (1.39)	[4.68, 5.84]	22	5.12 (1.38)	[4.52, 5.71]	No effect, $d = 0.10$	N/A
TOI	Motivation	23	5.81 (1.04)	[5.37, 6.24]	22	5.01 (1.04)	[4.56, 5.45]	Medium, $d = 0.77$	E > C incremental
NCS	Motivation	22	4.42 (0.38)	[4.26, 4.58]	22	4.30 (0.38)	[4.14, 4.47]	No effect, $d = 0.32$	N/A
FCI: Accuracy	Far Transfer	23	0.45 (0.11)	[0.41, 0.50]	19	0.38 (0.11)	[0.33, 0.43]	Medium, $d = 0.64$	E > C
FCI: Absolute accuracy	Metacognition	23	0.31 (0.07)	[0.28, 0.34]	21	0.33 (0.08)	[0.30, 0.37]	No effect, $d = 0.27$	N/A
FCI: Gamma	Metacognition	23	0.38 (0.37)	[0.22, 0.54]	21	0.21 (0.37)	[0.05, 0.38]	No effect, $d = 0.46$	N/A
FCI: Discrimination	Metacognition	23	0.01 (0.20)	[−0.07, 0.10]	21	−0.13 (0.23)	[−0.22, −0.03]	Medium, $d = 0.65$	E < C discriminatory
CVS: Accuracy	PFL	20	0.94 (0.10)	[0.84, 1.05]	21	0.74 (0.31)	[0.63, 0.84]	Large, $d = 0.87$	E > C
CVS: Guessing	PFL	20	0.02 (0.05)	[−0.01, 0.04]	21	0.16 (0.19)	[0.07, 0.25]	Large, $d = 1.01$	E < C
Metacognitive reflection: Knowledge of skills	Metacognition	22	0.45 (0.51)	[0.23, 0.68]	20	0.25 (0.44)	[0.04, 0.46]	No effect, $d = 0.44$	N/A
Metacognitive reflection: Usefulness of activities	Metacognition	22	0.36 (0.49)	[0.15, 0.58]	20	0.40 (0.50)	[0.16, 0.64]	No effect, $d = 0.08$	N/A
Metacognitive reflection: Frequency across contexts	Metacognition	22	2.45 (0.83)	[2.09, 2.82]	19	2.61 (1.08)	[2.09, 3.14]	No effect, $d = 0.17$	N/A
Metacognitive reflection: Frequency of skill use	Metacognition	21	3.10 (0.77)	[2.75, 3.45]	20	3.20 (0.83)	[2.81, 3.59]	No effect, $d = 0.12$	N/A
(Plan, Monitor, Evaluate)		21	3.43 (0.93)	[3.01, 3.85]	20	2.90 (0.79)	[2.53, 3.27]	No effect, $d = 0.61$	E > C
		21	3.95 (1.02)	[3.49, 4.42]	20	3.70 (1.08)	[3.19, 4.21]	No effect, $d = 0.24$	N/A

Note. E = experimental condition; C = control condition; MAI = Metacognitive Awareness Inventory; TOI = Theories of Intelligence; NCS = Need for Cognition Scale; FCI = Force Concept Inventory; CVS = Control of Variables Strategy.

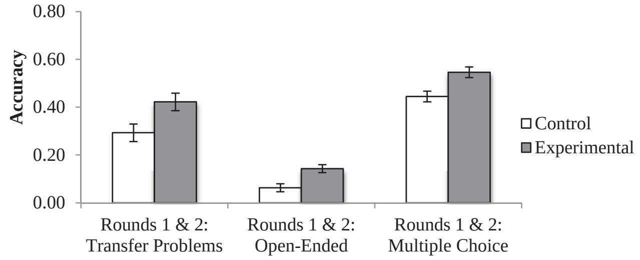

Figure 4. Proportion of correctly solved transfer problems, correctly recalled metacognitive concepts, and correctly answered multiple-choice questions.

mental condition reporting greater self-efficacy to complete classroom-related tasks than the control.

Achievement goal orientations. A preliminary analysis evaluating the homogeneity-of-regression assumption indicated an interaction between condition and pretest scores for only mastery-approach goals, $F(1, 41) = 3.94$, $p = .05$, $\eta_p^2 = .09$, rendering an ANCOVA unfit for interpretation. Therefore, we evaluated the effect of condition on the mastery-approach goals using a standard ANOVA. Since there was also a violation of equal variances indicated by the Levene's Test of Homogeneity of Variances, $F(1, 43) = 12.26$, $p = .001$, we report Welch's adjusted F-ratio for mastery-approach goals, *Welch's F* (1, 30.38) = 7.67, $p = .005$, est. $\omega^2 = .15$, with the experimental condition more strongly endorsing those goals than the control. One-way ANCOVAs showed large effects for the covariates for performance-approach, $F(1, 42) = 26.06$, $p < .001$, $\eta_p^2 = .38$, and performance-avoidance goals, $F(1, 42) = 18.17$, $p < .001$, $\eta_p^2 = .30$. This indicates that students' AGQ-R pretest scores predicted their posttest scores for each performance goal. There was no effect of condition on performance-approach, $F(1, 42) = 2.49$, $p = .12$, $\eta_p^2 = .06$, or performance-avoidance goals, $F(1, 42) = 0.12$, $p = .73$, $\eta_p^2 = .003$. These results provide evidence that training in metacognition facilitates mastery-approach goal adoption more than problem-solving practice.

Theories of intelligence. A one-way ANCOVA revealed a large effect of the covariate on the theory of intelligence measure, $F(1, 42) = 11.48$, $p = .002$, $\eta_p^2 = .22$, indicating that students' TOI pretest scores predicted their posttest scores. Analyses also revealed a medium effect of condition, $F(1, 42) = 6.42$, $p = .02$, $\eta_p^2 = .13$, with the experimental condition more strongly endorsing the belief that intelligence can be changed compared to the control.

Need for cognition. An ANCOVA revealed a large effect of the covariate, $F(1, 41) = 121.66$, $p < .001$, $\eta_p^2 = .79$, indicating that students' NCS pretest scores predicted their posttest scores. However, there was no effect of condition, $F(1, 41) = 1.02$, $p = .32$, $\eta_p^2 = .02$, with both conditions reporting similar levels of enjoyment for demanding cognitive tasks.

Summary of Self-Reports

The intervention had a large effect on students' self-reported motivation, with the experimental condition more strongly endorsing beliefs in self-efficacy and task value compared to the control. Additionally, the intervention led to higher endorsements of mastery-approach goals and incremental theories of intelligence. There was no effect on students' need for cognition, performance-approach goals, or performance-avoidance goals.

Conceptual Knowledge Test

Performance accuracy. A one-way ANCOVA revealed a large effect of the covariate, $F(1, 39) = 26.58$, $p < .001$, $\eta_p^2 = .41$, indicating that students' FCI pretest scores predicted their posttest scores. Analyses revealed a medium effect of condition on their conceptual understanding of physics, $F(1, 39) = 4.89$, $p = .03$, $\eta_p^2 = .11$, with the experimental group outperforming the control.[4] This shows that the experimental group had a better conceptual understanding of Newtonian physics than the control condition after the intervention.

Metacognitive monitoring accuracy. One-way ANOVAs revealed that there was no effect of condition on absolute accuracy, $F(1, 42) = 0.56$, $p = .46$, $\eta_p^2 = .01$, or Gamma, $F(1, 42) = 2.23$, $p = .14$ $\eta_p^2 = .05$. However, there was a medium effect on discrimination, $F(1, 42) = 5.20$, $p = .03$, $\eta_p^2 = .11$, in which the experimental group had higher discrimination scores than the control group. This shows that although there were no differences in the absolute accuracy or Gamma, the experimental group was unable to discriminate between what they knew and what they did not know, as indicated by the near-zero discrimination score, whereas the control believed they knew the answers more often to questions answered incorrectly (see Figure 5).

Preparation for Future Learning – Learning to Control the Variables

A preliminary analysis evaluating the homogeneity-of-regression assumption indicated an interaction between condition and pretest performance, $F(1, 37) = 9.27$, $p = .004$, $\eta_p^2 = .20$, indicating a violation and rendering an ANCOVA unfit for interpretation. Therefore, we evaluated the effect of condition on the PFL CVS using a standard ANOVA without controlling for the pretest as a covariate. A one-way ANOVA revealed a large effect of condition on PFL CVS scores, $F(1, 39) = 8.16$, $p = .007$, $\eta_p^2 = .17$, in which the experimental condition outperformed the control condition. A one-way ANOVA also revealed a large effect of condition on the proportion of guessing during the activity, $F(1, 39) = 10.10$, $p = .003$, $\eta_p^2 = .21$, with students in the control condition saying they guessed more often than the experimental condition.

Metacognitive Reflection

The reflection survey was analyzed for three types of questions: knowledge of the metacognitive skills, perceived usefulness, and frequency of use. A chi-square test showed no effect of condition on students' knowledge about metacognitive skills, $\chi^2(1, N = 42) = 1.91$, $p = .17$ or perceived usefulness, $\chi^2(1, N = 42) = 0.06$, $p = .81$. A one-way ANOVA revealed no effect of condition on frequency of using the skills, $F(1, 39) = 0.52$, $p = .26$, $\eta_p^2 = .03$. There was also no effect of condition when students were specifically asked how frequently they planned, $F(1, 39) = 0.18$, $p =$

---

[4] The same results occurred with a between-subjects ANOVA without controlling for the pretest.

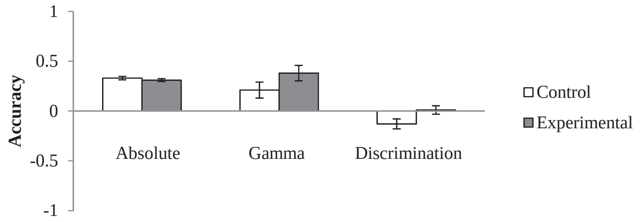

Figure 5. Average scores for absolute accuracy, gamma, and discrimination by condition.

.68, $\eta_p^2 = .004$, monitored, $F(1, 39) = 3.86$, $p = .06$, $\eta_p^2 = .09$, or evaluated their solutions, $F(1, 39) = 0.59$, $p = .45$, $\eta_p^2 = .02$.

Discussion

The results of this study highlight the importance of metacognitive skills as they demonstrate that direct instruction and practice of multiple metacognitive skills can improve metacognitive monitoring, learning, transfer, and motivational outcomes in a middle school science class. First we discuss the results pertaining to each of these outcomes, and then we discuss the particular strengths of the intervention and its implications, limitations, and future directions.

Metacognition

Much past work has examined the effect of metacognitive training or support for a single academic topic or task (e.g., linear equations), and then assessed the learning outcomes for that same topic. However, training the metacognitive skills in a *single* topic or task may limit the generalizability of those skills as they are specialized to that task. In contrast, we implemented our training across two domains (puzzles and physics) that enabled us to use the exact same metacognitive instruction in two different contexts with the aim of facilitating the acquisition of domain-general skills. This approach is consistent with research on analogy and worked examples that has shown that solving problems with the same underlying structure in multiple problem contexts can increase the likelihood of abstraction and transfer (Alfieri, Nokes-Malach, & Schunn, 2013; Gick & Holyoak, 1983). First students learned about metacognitive skills when solving content-light puzzle problems and then transitioned to using those skills when solving more content-rich physics problems. We used this approach with the central goal of increasing students' declarative and procedural knowledge of metacognition through instruction and practice.

As expected, we found that students in the experimental condition demonstrated greater declarative knowledge of what it means to plan, monitor, and evaluate, as assessed in students' responses to multiple-choice and open-ended questions in the eight packet quizzes. These results show that the metacognitive training was effective in facilitating the immediate acquisition of declarative knowledge of metacognition. Although absolute gains in declarative knowledge of metacognitive skills were relatively modest, these results show that metacognitive knowledge is not self-evident to the average middle school student.

To assess students' awareness of their use of metacognitive skills, we administered the self-report MAI questionnaire. The results did not support our hypothesis that the metacognitive training would increase students' reported use of planning, monitoring, and evaluating, despite being given explicit instruction on the nature, development, and application of these skills. One explanation for this finding is that students may not have been aware of their use of metacognitive skills, or their recollection of their use may not have been accurate. It is also possible that it takes some time after training for students to develop and notice habits of using metacognitive skills. Consistent with this explanation, there was a trend in which the experimental condition reported engaging in more metacognitive monitoring than the control condition on an assessment administered 13 weeks after the intervention.

While students' self-reports of metacognition showed very few differences between the experimental and control conditions, the metacognitive intervention still had an effect on students' behaviors. We assessed students' metacognitive monitoring abilities by prompting them to rate their confidence on each of their answers on a conceptual knowledge assessment. These confidence judgments supported our hypothesis that the experimental condition demonstrated less of a confidence bias when discriminating between correct and incorrect answers. We view this as a first step toward a more accurate assessment of one's knowledge. However, contrary to our expectations, the intervention did not improve students' absolute accuracy. Perhaps with more experience and practice with these skills or through the use of an easier test (students on average answered fewer than half of the FCI items correctly), the students would have increased their absolute accuracy as well.

Motivation

Little prior work has examined the effects of metacognitive instruction on multiple motivational outcomes; therefore, we included a number of constructs that have been hypothesized to relate to the metacognitive skills of planning, monitoring, and evaluating (Zimmerman, 2011), as well as the multiple motivational pathways hypothesized to support students' learning (Pintrich, 2003). As predicted, the intervention increased students' endorsements of a number of these measures. First, we found that students in the metacognitive intervention reported greater task value for the material covered in their science class. One possible explanation for this effect is that the intervention materials emphasized the importance of understanding and provided strategies to improve students' learning. This may have helped students to realize the value of mastering and learning the course material. Increases in task value may also be related to increases in self-efficacy. Past work has shown a positive relationship between students' competence beliefs and task value, although the direction of this relationship has not been tested (Jacobs et al., 2002).

Since the instructional materials focused on understanding (e.g., the planning packet focused on understanding the problem, the problem's goal, and relevant prior knowledge), we also expected students to be more likely to adopt a mastery (or intrapersonal) goal over a performance (or normative) goal. Results confirmed these hypotheses as the intervention led to higher endorsements of mastery-approach goals while it did not affect students' endorsements of performance-approach or performance-avoidance goals. This result makes sense given that the intervention did not highlight normative measures of learning. Both findings regarding mastery-approach goals and task value are consistent with prior work showing a relationship between mastery-approach goals and interest (Harackiewicz, Barron, Tauer, Carter, & Elliot, 2000), though we cannot speculate on whether one caused the other.

The intervention's focus on understanding could have also led to the experimental condition's stronger endorsement of incremental theories of intelligence and beliefs of self-efficacy. Students' incremental theories of intelligence may have been affected in several possible ways. First, to the degree that students viewed themselves as acquiring tools to change their problem-solving ability and engaged in evaluation after arriving at solutions, they may have come to attribute their success or failure to their problem-solving decisions and effort. Thus they might have come to regard success, and by extension intelligence, as something that can be changed through effort and learning. Second, past work has related theories of intelligence to achievement goals, with incremental theories having a positive correlation with mastery-approach goals (Blackwell, Trzesniewski, & Dweck, 2007; Dweck & Leggett, 1988; Dweck, 1999). Therefore changing mastery-approach goals might have led to changes in theories of intelligence. Third, by creating instruction that stated that it is okay to learn from mistakes, students might see that mistakes are important for learning and part of improving their problem solving. As a result, students might be less likely to view mistakes as a sign that they are simply not "intelligent." Students' higher levels of self-efficacy may also result from several factors. Students might have felt more confident about their ability to learn in the class as they acquired more domain knowledge through the intervention. They might have also felt more confident in their learning because they had metacognitive skills to help them identify their learning progress.

Three of the measures, mastery-avoidance goals, control of learning beliefs, and effort regulation, were dropped because scores on those measures had low reliability, and there was no effect of the intervention on students' need for cognition. Although prior work has shown that those who endorse a higher need for cognition have higher levels of metacognition (Coutinho, 2006), it is not clear that the relationship is causal. Alternatively, increasing students' need for cognition may require more time and experience than our intervention provided. Cacioppo and colleagues (1996) argue that the internal desire to perform on cognitively effortful endeavors originates from past experiences, memories, and behavioral histories, as data from their studies indicated that need for cognition changes over long periods of time. Another explanation for this finding is that students may not have felt that metacognition was more effortful than their original ways of thinking. It is clear that the declarative knowledge and practice using metacognitive skills led to consistent effects across competence-related motivational assessments, resulting in greater endorsement of task value, mastery-approach goals, self-efficacy, and incremental theories of intelligence.

Learning and Knowledge Transfer

Having evidence that the students in the experimental condition acquired some declarative knowledge of metacognitive skills, changed at least one of their metacognitive behaviors, and demonstrated motivational benefits, we evaluated whether the intervention affected their domain learning and transfer outcomes. These results contribute to the limited research on metacognition's impact on learning and the degree to which it improves future learning.

Results from the learning packets revealed that students in the experimental condition performed better than those in the control condition on the learning packet near-transfer problems. This effect is particularly surprising given that the control condition received significantly more practice on similar problems while completing the learning packets. The majority of the experimental condition's time on the packets was spent learning metacognitive content, while the control condition practiced problems that were directly relevant to the transfer problem. These results show the power of metacognitive instruction for optimizing students' learning and making a small amount of practice more effective, especially since both conditions spent the same amount of time working through the learning packets.

Students' conceptual understanding of the physics content being covered in class at the time of the intervention as well as content covered in the second round of learning packets was assessed by the FCI. The results from this assessment were consistent with our hypothesis that the intervention would increase students' conceptual knowledge and reduce their misconceptions. This is an impressive result given that the control group received more practice problems than the experimental condition, which shows that metacognitive training promotes deeper learning than that which occurs by simply completing additional practice. We acknowledge that our pretest was an imprecise measure of the students' knowledge before the intervention because students were instructed on some of the content after the pretest assessment but before the intervention. However when excluding the pretest we still see the same posttest results with students in the experimental condition performing better than students in the control.

Students in the experimental condition also showed better performance on the PFL assessment, which tested students' ability to learn from a new self-guided science activity. The learning and assessment results showed that students in the experimental condition learned more from the new materials than the control condition. Students who received metacognitive training also reported less guessing, suggesting that they had either learned the new material better and therefore did not need to guess or were more likely to find other productive problem-solving strategies. These results are consistent with the interpretation that students given metacognitive instruction and training were able to apply those skills to a new task.

Our use of multiple assessments allowed us to examine the impact of metacognitive instruction on the transfer of domain-relevant content as well as domain-general metacognitive skills. We were able to see the immediate effects of metacognition instruction on near-transfer problems at the end of the learning packets (domain-relevant content). This benefit could have been supported by declarative or procedural knowledge gained from the earlier problems in the packet, the application of metacognitive skills to the transfer problems, or both. The results from the FCI show the metacognitive instruction's impact on the acquisition of declarative, conceptual knowledge acquired from the learning packets (transfer of the domain-relevant content), its impact on students' abilities to reason through challenging conceptual questions (transfer of the domain-general metacognitive skills), or both. This type of performance is classified as far transfer because the test was delayed and the training problems and classroom instruction had different problem features than the FCI, so students had to adapt concepts to new features on the test. We also saw the intervention's benefits on preparation for future learning as the domain-general metacognitive skills transferred to a novel learning activity with new content (i.e., CVS) as evidenced by better learning outcomes.

Implications

The wide-ranging effects of this study provide evidence for understanding the nature of metacognition within four strands of research: SRL, motivation, metacognitive instruction, and transfer. First, this study shows that metacognitive training can improve both cognitive and motivational aspects of learning. This provides support to SRL hypotheses that enhancing one SRL component (metacognition) leads to benefits in another (motivation). This study also demonstrates how metacognition can impact near, far, and PFL transfer and highlights the need for more research to explore the degree to which metacognition influences various levels, types, and sources of transfer.

There were two design features of the intervention that may have helped to contribute to its effectiveness in the classroom. First, the instruction was delivered through paper packets that students completed individually. This instructional technique has the potential to reduce variability in the implementation across teachers and contexts as the instruction and practice is embedded into the class materials and activities. This may be especially critical as discipline-area teachers often have difficulty providing direct instruction on self-regulated learning processes (Zohar, 1999). Second, students' responses in the experimental condition were scaffolded throughout the packets, with instruction of metacognitive skills interwoven with prompts to practice executing those skills. This level of scaffolding appears to be effective, as students responded to the prompts and questions regarding the metacognitive skills on their own. We view this as another strength of our study as students utilized the skills once they learned more about them. Evidence for the use of these skills comes from the performance differences on the manipulation checks as well as the learning benefits in the preparation for future learning measure and FCI. From a practical perspective, this work demonstrates that students' knowledge of metacognitive skills can be improved through direct instruction and practice. Furthermore, the framework of the intervention (i.e., direct instruction, worked examples, scaffolded practice) could be utilized as a model for teachers when instructing other content.

Limitations and Future Research

While this study has several positive results, there are also limitations and questions for future research to address. Although the intervention was effective for an adolescent population, further research is required to generalize these results to younger or older learners. Metacognition continues to develop into adulthood, so it is unclear whether this intervention would be as effective for older populations such as college students. Prior work has shown that variations in metacognitive ability tend to decrease with age, so a metacognitive intervention might not benefit older populations that may have reached a plateau (Weil et al., 2013). Adolescence is also a critical period in which motivation tends to decline, which could be why the intervention impacted student motivation. This was also a small-scale study implemented in two science classrooms with a single teacher. Future research is necessary to generalize these results across class topics and contexts.

Future research should also investigate which component of the intervention, direct instruction or practice, was most effective. How does each instructional component relate to the different learning, transfer, and motivational outcomes? Much literature on learning and instruction suggests that a combination of instruction and practice is generally most effective (Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013). However, understanding how each component affected the outcomes could guide efforts to improve or streamline the materials. Future work should also tease apart which metacognitive skills caused the different effects in this study. For example, perhaps learning the skill of monitoring is the primary reason students were able to perform better on the PFL task, as students could pinpoint what they did and did not understand and, therefore, fill in the gaps in their understanding. There is also the possibility that the types of problems we used within the intervention packets differentially influenced the use of the metacognitive skills; for example, one problem type might have encouraged the use of monitoring when it was meant to scaffold planning. Although all three skills could be applied to every problem, some may have been better suited for the targeted skill than others.

We encourage future research to use multiple measures to assess both metacognitive and motivational outcomes. We attempted to address this limitation by incorporating both metacognitive questionnaires and metacognitive monitoring judgments. However, future work should also evaluate SRL behaviors of volitional strategies (e.g., likelihood to complete homework), motivation (e.g., likelihood to pursue science courses), and metacognition (e.g., likelihood to create a plan and quality of the plan).

Lastly, there are important questions to explore regarding the interactive nature and timing of the metacognitive and motivational processes as they relate to learning outcomes. For example, how much of the PFL effect is due to the changed motivation of the students, the metacognitive skills themselves, or the combination? Since incremental theories of intelligence are related to better achievement, especially for underrepresented students (e.g., Blackwell et al., 2007; Good, Aronson, & Inzlicht, 2003), it would also be interesting to examine whether it was learning about metacognition, practicing the skills, increasing other motivational constructs, or a combination that increased students' incremental theories. The current study shows that an intervention that brings together complementary theoretical frameworks (e.g., SRL, metacognitive, and motivation) can have powerful effects across an array of student learning and motivation outcomes. Future work should further examine the promise of multiple-component interventions targeting several constructs simultaneously.

References

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

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

Berardi-Coletta, B., Buyer, L. S., Dominowski, R. L., & Rellinger, E. R. (1995). Metacognition and problem solving: A process-oriented approach. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 205–223. http://dx.doi.org/10.1037/0278-7393.21.1.205

Blackwell, L. S., Trzesniewski, K. H., & Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development, 78,* 246–263. http://dx.doi.org/10.1111/j.1467-8624.2007.00995.x

Boekaerts, M., & Corno, L. (2005). Self-Regulation in the classroom: A perspective on assessment and intervention. *Applied Psychology, 54,* 199–231. http://dx.doi.org/10.1111/j.1464-0597.2005.00205.x

Brand, S., Reimer, T., & Opwis, K. (2003). Effects of metacognitive thinking and knowledge acquisition in dyads on individual problem solving and transfer performance. *Swiss Journal of Psychology, 62,* 251–261. http://dx.doi.org/10.1024/1421-0185.62.4.251

Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple implications. *Review of Research in Education, 24,* 61–100.

Cacioppo, J. T., Petty, R. E., Feinstein, J. A., & Jarvis, W. B. G. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin, 119,* 197–253. http://dx.doi.org/10.1037/0033-2909.119.2.197

Chen, Z., & Klahr, D. (1999). All other things being equal: Acquisition and transfer of the control of variables strategy. *Child Development, 70,* 1098–1120. http://dx.doi.org/10.1111/1467-8624.00081

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum.

Coutinho, S. A. (2006). The relationship between the need for cognition, metacognition, and intellectual task performance. *Educational Research Review, 1,* 162–164.

Cross, D. R., & Paris, S. G. (1988). Developmental and instructional analyses of children's metacognition and reading comprehension. *Journal of Educational Psychology, 80,* 131–142. http://dx.doi.org/10.1037/0022-0663.80.2.131

Dignath, C., Büttner, G., & Langfeldt, H. P. (2008). How can primary school students learn self-regulated learning strategies most effectively? *Educational Research Review, 3,* 101–129. http://dx.doi.org/10.1016/j.edurev.2008.02.003

Dignath, C., & Büttner, G. (2008). Components of fostering self-regulated learning among students. A meta-analysis on intervention studies at primary and secondary school level. *Metacognition and Learning, 3,* 231–264. http://dx.doi.org/10.1007/s11409-008-9029-x

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, 14,* 4–58. http://dx.doi.org/10.1177/1529100612453266

Dweck, C. S. (1999). *Self-theories: Their Role in Motivation, Personality, and Development*. Philadelphia: Psychology Press.

Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review, 95,* 256–273. http://dx.doi.org/10.1037/0033-295X.95.2.256

Efklides, A. (2011). Interactions of metacognition with motivation and affect in self-regulated learning: The MASRL model. *Educational Psychologist, 46,* 6–25. http://dx.doi.org/10.1080/00461520.2011.538645

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

Ford, J. K., Smith, E. M., Weissbein, D. A., Gully, S. M., & Salas, E. (1998). Relationships of goal orientation, metacognitive activity, and practice strategies with learning outcomes and transfer. *Journal of Applied Psychology, 83,* 218–233. http://dx.doi.org/10.1037/0021-9010.83.2.218

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

Good, C., Aronson, J., & Inzlicht, M. (2003). Improving adolescents' standardized test performance: An intervention to reduce the effects of stereotype threat. *Journal of Applied Developmental Psychology, 24,* 645–662. http://dx.doi.org/10.1016/j.appdev.2003.09.002

Goodman, L. A., & Kruskal, W. H. (1954). Measures of association for cross classification. *Journal of The American Statistical Association, 49,* 732e764.

Harackiewicz, J. M., Barron, K. E., Tauer, J. M., Carter, S. M., & Elliot, A. J. (2000). Short-term and long-term consequences of achievement goals: Predicting interest and performance over time. *Journal of Educational Psychology, 92,* 316–330. http://dx.doi.org/10.1037/0022-0663.92.2.316

Hattie, J., Biggs, J., & Purdie, N. (1996). Effects of learning skills interventions on student learning: A meta-analysis. *Review of Educational Research, 66,* 99–136. http://dx.doi.org/10.3102/00346543066002099

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force Concept Inventory. *The Physics Teacher, 30,* 141–158. http://dx.doi.org/10.1119/1.2343497

Hsu, T. (2005). *Foundations of Physical Science* (2nd ed.). Peabody, MA: CPO Science.

Jacobs, J. E., Lanza, S., Osgood, D. W., Eccles, J. S., & Wigfield, A. (2002). Changes in children's self-competence and values: Gender and domain differences across grades one through twelve. *Child Development, 73,* 509–527. http://dx.doi.org/10.1111/1467-8624.00421

Kelemen, W. L., Frost, P. J., & Weaver, C. A., III. (2000). Individual differences in metacognition: Evidence against a general metacognitive ability. *Memory & Cognition, 28,* 92–107. http://www.ncbi.nlm.nih.gov/pubmed/10714142. http://dx.doi.org/10.3758/BF03211579

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

Lin, X., & Lehman, J. D. (1999). Supporting learning of variable control in a computer-based biology environment: Effects of prompting college students to reflect on their own thinking. *Journal of Research in Science Teaching, 36,* 837–858. http://dx.doi.org/10.1002/(SICI)1098-2736(199909)36:7<837::AID-TEA6>3.0.CO;2-U

Maki, R. H., Shields, M., Wheeler, A. E., & Zacchilli, T. L. (2005). Individual differences in absolute and relative metacomprehension accuracy. *Journal of Educational Psychology, 97,* 723–731. http://dx.doi.org/10.1037/0022-0663.97.4.723

Meece, J. L., Blumenfeld, P. C., & Hoyle, R. H. (1988). Students' goal orientations and cognitive engagement in classroom activities. *Journal of Educational Psychology, 80,* 514–523. http://dx.doi.org/10.1037/0022-0663.80.4.514

Nokes, T. J. (2009). Mechanisms of knowledge transfer. *Thinking & Reasoning, 15,* 1–36. http://dx.doi.org/10.1080/13546780802490186

Nokes, T. J., & Ohlsson, S. (2005). Comparing multiple paths to mastery: What is learned? *Cognitive Science, 29,* 769–796. http://dx.doi.org/10.1207/s15516709cog0000_32

Nokes-Malach, T. J., & Mestre, J. (2013). Toward a model of transfer as sense-making. *Educational Psychologist, 48,* 184–207. http://dx.doi.org/10.1080/00461520.2013.807556

Nolen, S. (1988). Reasons for studying: Motivational orientations and study strategies. *Cognition and Instruction, 5,* 269–287. http://dx.doi.org/10.1207/s1532690xci0504_2

Olejnik, S., & Algina, J. (2000). Measures of effect size for comparative studies: Applications, interpretations, and limitations. *Contemporary Educational Psychology, 25,* 241–286. http://dx.doi.org/10.1006/ceps.2000.1040

Pajares, F. (2008). Motivational role of self-efficacy beliefs in self-regulated learning. In D. H. Schunk & B. J. Zimmerman (Eds.), *Motivation and self-regulated learning: Theory, research, and applications* (pp. 111–139). Mahwah, NJ: Lawrence Erlbaum.

Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and monitoring activities. *Cognition and Instruction, 1,* 117–175. http://dx.doi.org/10.1207/s1532690xci0102_1

Pintrich, P. R. (1999). The role of motivation in promoting and sustaining self-regulated learning. *International Journal of Educational Research, 31,* 459–470. http://dx.doi.org/10.1016/S0883-0355(99)00015-4

Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 451–502). San Diego, CA: Academic Press. http://dx.doi.org/10.1016/B978-012109890-2/50043-3

Pintrich, P. R. (2003). A motivational science perspective on the role of student motivation in learning and teaching contexts. *Journal of Educational Psychology, 95,* 667–686. http://dx.doi.org/10.1037/0022-0663.95.4.667

Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. J. (1991). *A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ). The University of Michigan*. Ann Arbor, MI: National Center for Research to Improve Post-Secondary Teaching.

Schraw, G. (2009). A conceptual analysis of five measures of metacognitive monitoring. *Metacognition and Learning, 4,* 33–45. http://dx.doi.org/10.1007/s11409-008-9031-3

Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology, 19,* 460–475. http://dx.doi.org/10.1006/ceps.1994.1033

Schraw, G., Dunkle, M. E., Bendixen, L. D., & Roedel, T. D. (1995). Does a general monitoring skill exist? *Journal of Educational Psychology, 87,* 433–444. http://dx.doi.org/10.1037/0022-0663.87.3.433

Schraw, G., Kuch, F., & Gutierrez, A. P. (2013). Measure for measure: Calibrating ten commonly used calibration scores. *Learning and Instruction, 24,* 48–57. http://dx.doi.org/10.1016/j.learninstruc.2012.08.007

Schraw, G., & Nietfeld, J. (1998). A further test of the general monitoring skill hypothesis. *Journal of Educational Psychology, 90,* 236–248. http://dx.doi.org/10.1037/0022-0663.90.2.236

Siler, S. A., Klahr, D., & Price, N. (2012). Investigating the mechanisms of learning from a constrained preparation for future learning activity. *Instructional Science*. Advance online publication.

Somuncuoglu, Y., & Yildirim, A. (1999). Relationship between achievement goal orientation and use of learning strategies. *The Journal of Educational Research, 92,* 267–277. http://dx.doi.org/10.1080/00220679909597606

Veenman, M. V., Elshout, J. J., & Meijer, J. (1997). The generality vs domain-specificity of metacognitive skills in novice learning across domains. *Learning and Instruction, 7,* 187–209. http://dx.doi.org/10.1016/S0959-4752(96)00025-4

Veenman, M. V. J., & Verheij, J. (2001). Identifying technical students at risk: Relating general versus specific metacognitive skills to study success. *Learning and Individual Differences, 13,* 259–272. http://dx.doi.org/10.1016/S1041-6080(02)00094-8

Veenman, M. V. J., Wilhelm, P., & Beishuizen, J. J. (2004). The relation between intellectual and metacognitive skills from a developmental perspective. *Learning and Instruction, 14,* 89–109. http://dx.doi.org/10.1016/j.learninstruc.2003.10.004

Weil, L. G., Fleming, S. M., Dumontheil, I., Kilford, E. J., Weil, R. S., Rees, G., . . . Blakemore, S. J. (2013). The development of metacognitive ability in adolescence. *Consciousness and Cognition, 22,* 264–271. http://dx.doi.org/10.1016/j.concog.2013.01.004

Winne, P. H. (1995). Inherent details in self-regulated learning. *Educational Psychologist, 30,* 173–187. http://dx.doi.org/10.1207/s15326985ep3004_2

Wolters, C. A., & Pintrich, P. R. (1998). Contextual differences in student motivation and self- regulated learning in mathematics, English, and social studies classrooms. *Instructional Science, 26* (1–2):2 7–47. http://dx.doi.org/10.1023/A:1003035929216

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13–39). San Diego, CA: Academic Press. http://dx.doi.org/10.1016/B978-012109890-2/50031-7

Zimmerman, B. J. (2001). Theories of self-regulated learning and academic achievement: An overview and analysis. In B. Zimmerman & D. Schunk (Eds.), *Self-regulated learning and academic achievement: Theoretical perspectives* (pp. 1–37). Mahwah, NJ: Erlbaum.

Zimmerman, B. J. (2011). Motivational sources and outcomes of self-regulated learning and performance. In B. J. Zimmerman & D. H. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 49–64). New York: Routledge.

Zimmerman, B. J., Bonner, S., & Kovach, R. (1996). *Developing self-regulated learners; Beyond achievement to self-efficacy*. Washington, DC: American Psychological Association. http://dx.doi.org/10.1037/10213-000

Zimmerman, B. J., & Campillo, M. (2003). Motivating self-regulated problem solvers. In J. Davidson & R. J. Sternberg (Eds.), *The psychology of problem solving* (pp. 233–262). New York: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511615771.009

Zimmerman, B. J., & Martinez Pons, M. (1988). Construct validation of a strategy model of student self-regulated learning. *Journal of Educational Psychology, 80,* 284–290. http://dx.doi.org/10.1037/0022-0663.80.3.284

Zohar, A. (1999). Teachers' metacognitive knowledge and the instruction of higher order thinking. *Teaching and Teacher Education, 15,* 413–429. http://dx.doi.org/10.1016/S0742-051X(98)00063-8

Received January 2, 2014
Revision received October 27, 2014
Accepted November 27, 2014 ■