

## What works with worked examples: Extending self-explanation and analogical comparison to synthesis problems

Ryan Badeau,<sup>1</sup> Daniel R. White,<sup>1</sup> Bashirah Ibrahim,<sup>2</sup> Lin Ding,<sup>2,\*</sup> and Andrew F. Heckler<sup>1,†</sup>

<sup>1</sup>*Department of Physics, The Ohio State University, Columbus, Ohio 43210, USA*

<sup>2</sup>*Department of Teaching and Learning, The Ohio State University, Columbus, Ohio 43210, USA*

(Received 12 May 2017; published 31 August 2017)

The ability to solve physics problems that require multiple concepts from across the physics curriculum—“synthesis” problems—is often a goal of physics instruction. Three experiments were designed to evaluate the effectiveness of two instructional methods employing worked examples on student performance with synthesis problems; these instructional techniques, analogical comparison and self-explanation, have previously been studied primarily in the context of single-concept problems. Across three experiments with students from introductory calculus-based physics courses, both self-explanation and certain kinds of analogical comparison of worked examples significantly improved student performance on a target synthesis problem, with distinct improvements in recognition of the relevant concepts. More specifically, analogical comparison significantly improved student performance when the comparisons were invoked between worked synthesis examples. In contrast, similar comparisons between corresponding pairs of worked single-concept examples did not significantly improve performance. On a more complicated synthesis problem, self-explanation was significantly more effective than analogical comparison, potentially due to differences in how successfully students encoded the full structure of the worked examples. Finally, we find that the two techniques can be combined for additional benefit, with the trade-off of slightly more time on task.

DOI: 10.1103/PhysRevPhysEducRes.13.020112

### I. INTRODUCTION

Problem solving is a complex and multifaceted process. Accordingly, there has been a significant investment in problem solving research in physics exploring problem solving frameworks, novice vs expert problem solving strategies, and related procedural skills (for reviews, see Refs. [1–3]). However, the vast majority of these studies have typically focused on problems requiring the application of one single, isolated physics concept (e.g., Refs. [4–10]).

The following series of experiments seeks to investigate a specific subclass of physics problem, which we will refer to as a synthesis problem: namely, a question requiring the application of more than one major physics concept, often from disparate parts of the teaching timeline [11]. Synthesis problems are of importance for both theoretical and practical reasons. Practically, synthesis problems are often closer to real world situations in their complexity. As a result, improving student success on synthesis problems is consistent with the goal of better preparing future engineers

and scientists. Within physics education research, synthesis problems are similar to context-rich problems in this pursuit—a topic of ongoing research in both general problem solving [12–14] and computer-aided tutoring [15,16]. Synthesis problems are also of theoretical interest as they provide unique difficulties for students in the recognition and joint application of multiple concepts [17–20]. Our previous studies showed that these distinct challenges extend beyond just the sum of difficulties represented by the individual component concepts. In particular, the recognition of multiple concepts becomes a significant bottleneck in the context of these more complicated problems [20]. This difficulty with synthesis problems is likely exacerbated by end-of-chapter textbook exercises and homework activities focusing on practicing only the most recently learned material in the context of single-concept problems. Students often approach these end-of-chapter exercises with documented “plug-and-chug” algorithms that do not necessarily scale successfully to situations with multiple interconnected physics concepts [21–23]. As such, the experiments here represent a novel focus on extending instructional methods based on worked examples specifically to synthesis problems and the unique challenges therein, namely, multiple concept recognition.

#### A. Worked examples

Worked examples consist of a problem statement and a corresponding set of solution steps, often with the implicit goal of modeling an expertlike approach to the solution of the problem. Previous research has shown that worked examples can be extremely effective in aiding novice learners as they

\*Corresponding author.  
ding.65@osu.edu

†Corresponding author.  
heckler.6@osu.edu

*Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.*

attempt to master domain-specific knowledge and problem solving skills, especially in highly structured domains such as physics [24–26]. Moreover, seminal work by Sweller *et al.* demonstrated that with careful, principle-based instructional design, studying worked examples can be significantly more effective than individual practice solving problems [24]. This “worked example effect” has traditionally been framed in terms of cognitive load theory—worked examples are effective because they reduce extraneous load associated with inefficient problem solving strategies [27,28]. Rather than devote limited cognitive capacity to plug-and-chug algorithms and equation matching heuristics, a fully worked example allows the novice to instead focus on extracting the relevant solution structure and construct a conceptual schema for subsequent use on other novel problems.

### B. Self-explanation

Effective interventions based on worked examples are often coupled with prompted or spontaneous self-explanations, whereby novices seek to explain the rationale and structure of the worked examples to either themselves or an interested third party. The importance of self-explanation was identified in a study by Chi *et al.*, which asked college students to voluntarily self-explain to an experimenter as they studied examples of introductory mechanics problems [29]. They found that students that generated more high-quality self-explanations performed significantly better on follow-up problem solving tasks than their peers. That result was then confirmed and expanded upon in multiple studies in physics and other domain areas, such as biology, algebra, and computer programming [30–33].

However, as with many of the aforementioned studies on worked examples, the problems and applications used in previous work have predominately focused on mastering isolated concepts and their application to single concept problems, such as Newton’s second law in the context of an equilibrium problem (in the case of the original Chi experiments). As such, our goal is twofold: first, to extend the application of self-explanation specifically to the domain of synthesis problems in physics; and second, to compare the effectiveness of self-explanation within individual worked examples to analogical comparison across a pair of worked examples.

### C. Analogical comparison

Analogical reasoning is a mechanism of applying what has been previously learned from a base situation to a new, analogous target situation. Successful analogical reasoning requires that a person recognize base-target similarity, perform structural mapping, and subsequently apply the base solution to the target [34–39]. In physics, researchers have used analogical reasoning to facilitate student conceptual learning [40–45]. Although the methods and implementation have differed, the primary goal has often been to help novices acquire understanding of a novel concept via analogies to a situation that the student already comprehends

(such as invoking the idea of water flow to understand current in a circuit, or scaffolding a series of analogies to aid conceptual understanding of normal force).

Here, we focus on a specific type of analogical reasoning known as analogical comparison. Analogical comparison invokes student comparison between two worked examples with the intent that students extract the necessary structure to tackle a related target problem. The technique was explored in a study by Gentner, Loewenstein, and Thompson that tested the use of analogical comparison with undergraduates and business-negotiation techniques [38]. In their studies, participants were asked to explicitly compare and contrast two isomorphic base examples before solving a related target problem. It was found that this analogical comparison between base examples facilitated learners to recognize, map, and apply key principles significantly better than did the traditional technique of using only one single base example. Given previously documented student difficulties recognizing component concepts when solving synthesis problems [20,43], we posit that this technique of analogical comparison may be particularly suited to helping students solve physics synthesis problems. Since analogical comparison emphasizes the identification of conceptual structure, it may assist students to overcome the characteristic multiple concept recognition and joint application bottlenecks that were identified in our prior studies on synthesis problems [11,17–20].

This proposal is further supported by previous studies in physics that have tested the effectiveness of isomorphic worked examples and analogical reasoning as a method to improve student problem solving. In particular, Lin and Singh have previously shown that invoking student discussion and comparison of a single isomorphic worked example to a target multiconcept problem can improve student use of the relevant physics concepts [46]. Interestingly, they found that students who were first asked to try to solve the target problem before comparing it to the provided worked example performed significantly better on their subsequent solution of the target problem compared to participants who were provided the worked example, scaffolding prompts, and explicitly told that the target problem and provided worked example shared the same physical concepts (energy conservation and centripetal acceleration).

In comparison to the work of Lin and Singh, where a single isomorphic worked example was provided to the students for study and use on the target problems, the series of studies conducted here specifically employ analogical comparison across pairs of worked examples. By providing pairs of worked examples with similar solution structure, we test the hypothesis that analogical comparison can assist students to extract the overall solution structure of a target synthesis problem while minimizing the impact of surface features from the provided worked examples. In short, analogical comparison—through an appeal to similarities and differences across the worked examples—may serve as an effective way to help students create a generalizable, and thus readily transferable, solution schema.

**D. Research goal and experiment overview**

In light of these previous results, we sought to explore how students utilize worked examples specifically in the context of synthesis problems. In particular, we sought to test whether or not analogical comparison across examples facilitated student recognition of relevant concepts and improved their performance when solving a novel synthesis problem. Along with this overall research goal, we considered the following related questions. First, given the increased complexity of synthesis problems, is it more effective to invoke comparisons between worked examples that break down the target synthesis problem into its single-concept parts, or worked examples that include the concepts in combination? Second, how does the focus of the prompts influence analogical comparison (i.e., prompts involving holistic structure and overall concept recognition vs prompts for fine-grain applications of the individual concepts)? Third, how does analogical comparison across a pair of worked examples compare with self-explanation of each worked example independently? These questions have been addressed by a series of three experiments, illustrated schematically in Fig. 1.

**II. EXPERIMENT 1: ANALOGICAL COMPARISON AND SYNTHESIS PROBLEMS**

**A. Method**

**1. Design**

The goal of the first experiment was to compare several methods of analogical comparison to baseline performance from course instruction alone (control) and to recent practice solving single-concept problems (priming). In order to test the effectiveness of analogical comparison in training students to solve synthesis problems, we designed a target synthesis task that would require application of two physics concepts: energy conservation and circular motion. The target synthesis problem used for this study is shown in Fig. 2(c). In addition to being relevant to the students' course—it represents a canonical situation presented in various problems within introductory physics courses—the problem was chosen based on previous work which documented significant student difficulties with a similar problem [20].

Three different interventions using variations of methods for analogical comparison were designed. These

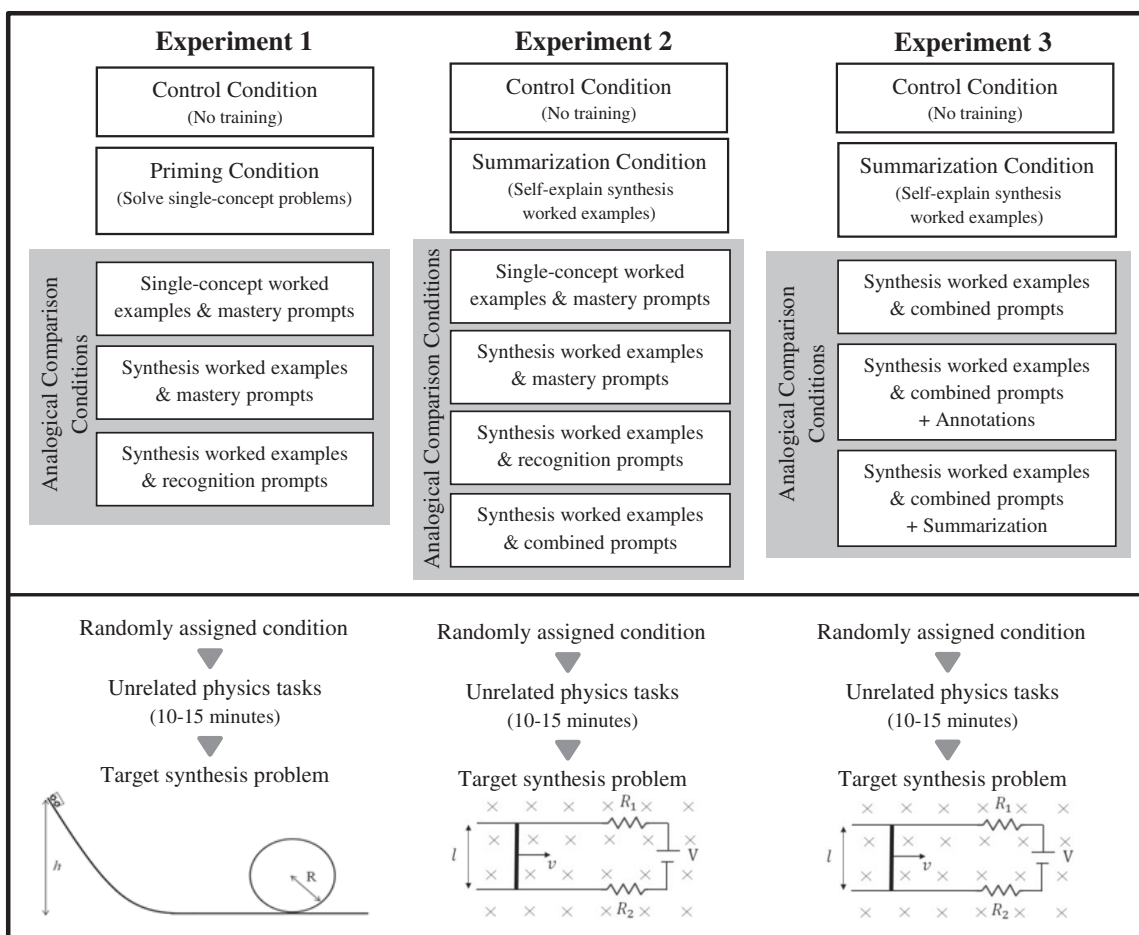


FIG. 1. The experimental design used in each of the three experiments. Experiments 2 and 3 used the same target synthesis problem.

interventions consisted of pairs of worked examples and sets of written prompts that asked students to compare the provided examples. Students were asked to write their responses to these prompts without additional feedback. Although the worked examples were intended to be similar to one another and to the target question (in terms of both physical context and the underlying conceptual structure of the solution), we purposefully varied the work flow of the provided solutions (i.e., the solutions applied the concepts in different orders) to avoid emphasizing a particular solution order for students to trace on the target problem. Examples of problems used as worked examples are included in Figs. 2(a) and 2(b). Full versions of all worked examples are included in Appendix A.

The three analogical comparison interventions varied both the type of worked example to be compared and the questions used for comparison, resulting in the following conditions: (i) single concept examples with mastery prompts (subsequently referred to as single-concept–mastery), (ii) synthesis examples with mastery prompts, and (iii) synthesis examples with recognition prompts.

The two sets of prompts—that is, mastery and recognition—were designed to focus student attention on different elements of the worked examples. The prompts involving single concept mastery explicitly targeted the application of the individual physics concepts within the worked examples, while the recognition prompts focused on concept recognition and their combination within the worked examples. Examples of each type of comparison prompt are included in Table I; the full set of questions are provided in Appendix A.

In addition to investigating different types of comparison questions, we varied the type of worked example provided to the students, namely either single-concept or synthesis problems. To this, the combination of the single-concept–mastery and synthesis-mastery conditions were designed to measure the effect of the type of worked example utilized for analogical comparison. These conditions used the same prompts for comparison with only minor changes to account for different line numbers in the solutions. In addition, the physical contexts of the solutions, diagrams, and problem statements were kept as similar as possible between the

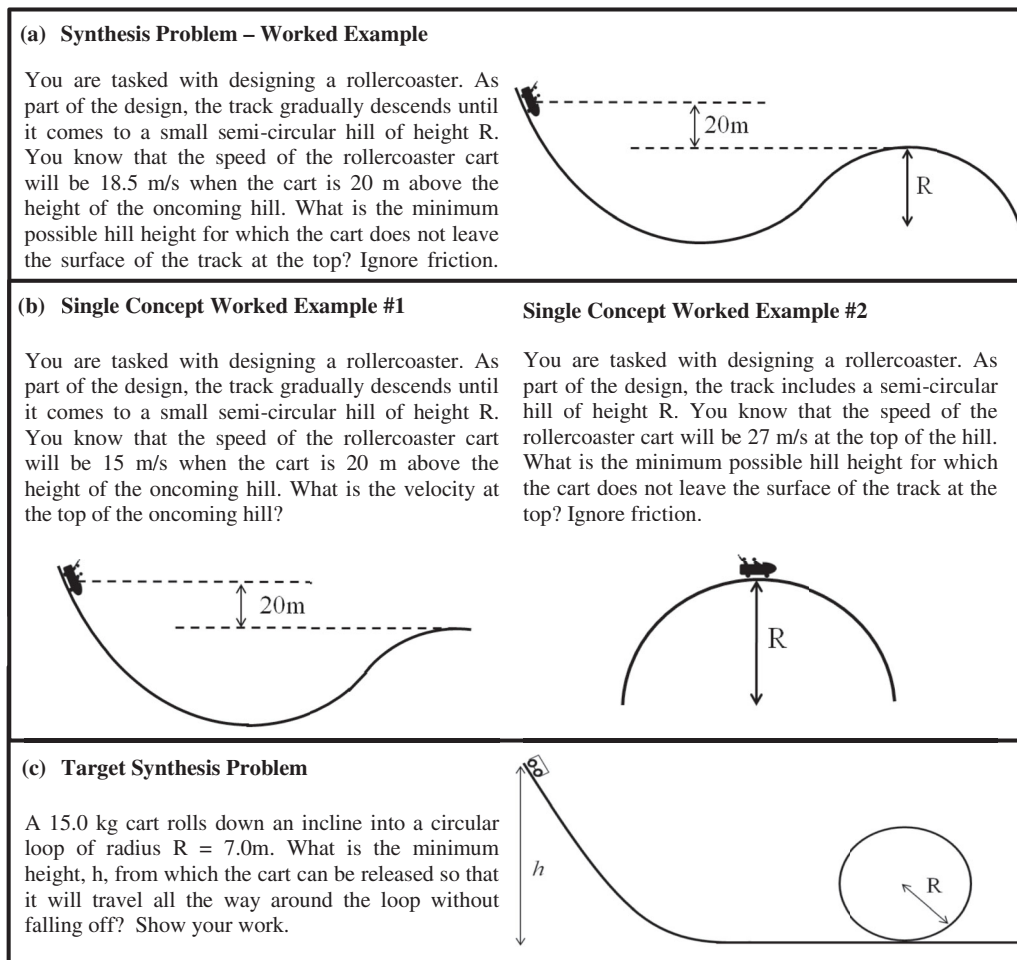


FIG. 2. An example synthesis worked example (a), the corresponding single concept worked examples (b), and the target synthesis problem (c).

TABLE I. Examples of the written prompts provided to students along with the worked examples.

Single-concept mastery prompts	Recognition prompts
Consider the diagram in line 1 of solution 1 and line 6 of solution 2. Explain any similarities and differences between the identified forces in terms of the two physical situations.	What are the main physical concept(s) used in both of the students' solutions?
Consider line 3 of solution 1 and line 8 of solution 2. Is the direction of the $mv^2/R$ term the same or different in the two solutions? Explain your reasoning.	For problem 1, identify the elements of the problem statement or diagram which indicate the need to use each of the physical concept(s) you mentioned in part A. List specific elements and explain your reasoning.
In problem 1 line 3, the student substitutes in zero for $F_N$ . Explain why. Is the situation in problem 2 similar or different to problem 1? Explain your reasoning.	Are the elements you identified above similar between the two problems? Explain your answer highlighting any differences and similarities.

synthesis and single-concept worked examples. In order to keep time on task as similar as possible across interventions, the analogical comparison conditions with synthesis problems included a comparison of only a single pair of worked examples. Students in the single-concept analogical comparison condition compared two pairs of worked examples: a pair of worked examples involving circular motion, and a pair of energy conservation examples.

In principle, there are compelling reasons to expect both methods to be successful. On the one hand, the single-concept problems are the embodiment of a reductionist approach: break the overall problem solution structure into its component parts and minimize cognitive load at each stage of comparison. As a result, the reduced complexity may assist students to recognize how to apply the individual concepts to a following novel problem. On the other hand, the synthesis worked examples are structurally more similar to the target synthesis problem, and include the structural step of joining the two concepts.

Along with the three analogical comparison interventions, there were two additional conditions. A no-training control was included to establish a baseline of student performance solely from course instruction. The final condition was the priming intervention. The priming intervention was included to provide a comparison for effects from recent single-concept practice with the relevant physics concepts, namely, increased concept availability. Rather than having students explicitly compare the worked examples, the priming condition asked students to solve two of the single concept problems—one of each concept—used as worked examples in the analogical comparison conditions.

## 2. Participants

Tasks were administered during Fall 2015. Participants were students in a calculus-based introductory mechanics course at The Ohio State University who participated as part of a 1-hr-long flexible homework assignment for course credit. Students earned full credit for the assignment

based on participation. A total of 196 students were randomly assigned into one of the five study conditions.

## 3. Administration

Students completed the training tasks and target synthesis problem in individual carrels in a quiet room. An equation sheet similar to those used in the course was provided to all students. Tasks were administered and collected by the proctor one at a time, and students were allowed to work at their own pace. Students first completed their selected training, followed by 10–15 min of unrelated physics tasks, and then the target synthesis problem. The intervening, unrelated physics tasks were included to reduce short term memory and priming effects from recent activation of the physical concepts relative to control. Students in the control condition also completed a set of unrelated physics tasks and the target synthesis problem, as shown in Fig. 1.

## 4. Method of analysis

A rubric for assessing student solutions of the target synthesis problem was determined by two of the authors. In an effort to provide an authentic measure of student performance, the rubric was designed to mirror a grading scheme that could be applied in the students' introductory course. The rubric is shown in Table II. After discussing the rubric, the authors coded all of the target student responses independently with an intercoder agreement of 80%. Concept recognition was coded generously (for example, a student earned credit for recognizing energy conservation if they tried to apply a  $\frac{1}{2}mv^2$  term), but required the student to commit to using the concept as part of their solution. Assessment of concept recognition was in complete agreement between the two coders. All disagreements were discussed leading to the agreed upon scores presented here.

## B. Results

Student final course grades in their introductory mechanics class were collected and compared across experimental conditions. To eliminate outliers, two cuts were uniformly

TABLE II. Scoring rubric for the target synthesis problem. 1 point was awarded for each item, for a total of 9 points.

Recognition	
+1	Energy conservation
+1	Centripetal acceleration
Application	
+1	$a_c$ applied correctly with Newton's second law
+1	Identify normal force = 0 (minimum criteria)
+1	Correct initial potential energy
+1	Correct final potential energy
+1	Included final kinetic energy (velocity top of loop $\neq 0$ )
+1	Substitute correct final velocity from centripetal motion constraint
Calculation	
+1	Correct final answer and no mathematical mistakes

conducted across all conditions: students must have completed the course (removing 1 student), and have scored no lower than 2 standard deviations below the course mean (removing a total of 5 students, ranging from 0 to 3 students per condition). Given that the synthesis problem represents the combination of single concepts covered as part of the students' introductory mechanics course, these cuts were conducted to minimize uninformative student difficulties in synthesizing those concepts that may have been due to simple unfamiliarity with the related physics course material. A one-way ANOVA of course grade showed no significant differences across conditions [ $F(4, 185) = 1.228, p = 0.301$ ].

The mean score on the target synthesis problem (out of a maximum of 9) and the number of students per

TABLE III. Mean score on target synthesis problem out of a maximum score of 9 points by condition. Errors shown are standard errors. Analogical comparison conditions are labeled AC.

Condition	Score $\pm$ SE	$N$
Control (No training)	4.15 $\pm$ 0.37	40
Priming	5.12 $\pm$ 0.39	43
AC: Single concept examples with mastery prompts	5.69 $\pm$ 0.40	32
AC: Synthesis examples with mastery prompts	7.23 $\pm$ 0.36	35
AC: Synthesis examples with recognition prompts	6.83 $\pm$ 0.40	40

condition are shown in Table III, with corresponding score distributions included in Fig. 3. The distributions are distinctly non-normal and roughly clustered into two distinct groups: one group centered near a score of 3–4 and the other at a total score of 8–9.

A Kruskal-Wallis  $H$  test was conducted to determine the effectiveness of the 4 interventions versus control (course instruction only). The mean rank of scores on the target synthesis problem was significantly different between conditions [ $\chi^2(4) = 36.2, p < 0.001$ ]. Pairwise comparisons of each intervention to control were conducted using Dunn's procedure with a Bonferroni correction for multiple comparisons ( $n = 4$ ). The adjusted  $p$  values are presented. Results indicated that while the priming condition was not significantly different from control ( $z = 1.97, p = 0.196$ ), all three analogical comparison conditions were significantly higher than control: Single-concept–mastery ( $z = 2.632, p = 0.034$ ), synthesis-mastery ( $z = 5.436, p < 0.001$ ), and synthesis-recognition ( $z = 4.379, p < 0.001$ ).

To test for hypothesized differences ( $H-1$ ) between analogical comparison and priming, ( $H-2$ ) between single-concept and synthesis worked examples, and ( $H-3$ ) between mastery and recognition prompts, we conducted a Kruskal-Wallis  $H$  test across only the intervention conditions. There was a significant difference in mean rank of scores on the target synthesis problem between the interventions [ $\chi^2(3) = 16.72, p = 0.001$ ]. Five pairwise comparisons were conducted using Dunn's procedure with a Bonferroni correction for multiple comparisons. The adjusted  $p$  values are presented. To test  $H-1$ , comparisons showed there were significant differences between priming and both the synthesis-mastery ( $z = 3.718, p < 0.001$ ) and synthesis–recognition conditions ( $z = 2.602, p = 0.045$ ), but not with single-concept–mastery ( $z = 0.721, p = 1.000$ ). To test  $H-2$ , pairwise comparison showed there were significant differences between single-concept–mastery and synthesis-mastery ( $z = 2.772, p = 0.03$ ). To test  $H-3$ , there were no significant differences between synthesis-mastery and synthesis-recognition ( $z = 1.187, p = 1.000$ ). Taken together, these results demonstrate that the effectiveness of analogical comparison extends beyond just single-concept practice and concept activation (as evidenced by

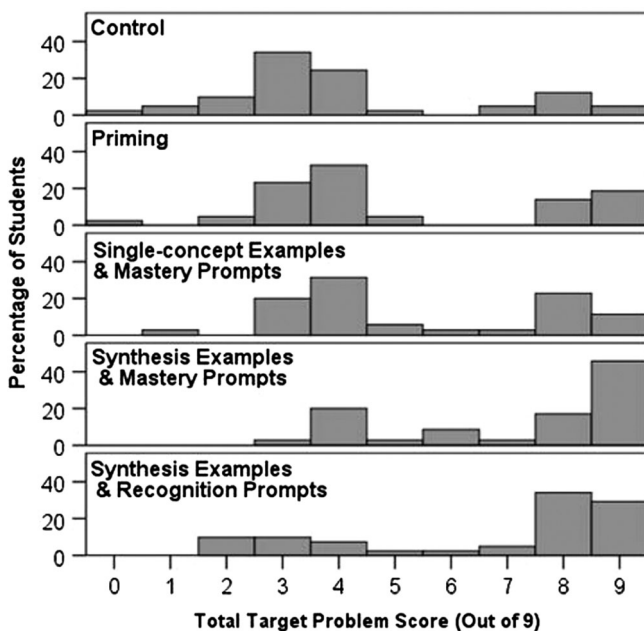


FIG. 3. Distributions of student score on target synthesis problem by treatment condition.

comparison to priming via single-concept problem solving exercises). Moreover, while varying the type of prompts had no significant effect on student performance with the target synthesis problem ( $d = 0.17$ ), students who compared synthesis worked examples performed significantly better on the target problem than those who compared examples highlighting the component concepts in isolation ( $d = 0.70$ ).

In addition to considering total scores on the target synthesis problem, we analyzed the proportion of students recognizing each of the two component concepts. The proportion of students recognizing a concept is shown in Fig. 4. Almost all students ( $\geq 95\%$ ) recognized and utilized energy conservation as part of their solution.

The proportion of students recognizing centripetal acceleration on the target synthesis problem varied considerably. A chi-squared test across treatment conditions showed there was a significant difference in the proportion of students recognizing centripetal acceleration when solving the target synthesis problem [ $\chi^2(3) = 29.899$ ,  $p < 0.001$ ]. *Post hoc* comparisons between treatments were conducted using pairwise chi-squared tests with a Bonferroni correction for multiple comparisons ( $n = 5$ ). The proportion of students recognizing centripetal acceleration in both the synthesis-mastery and synthesis-recognition conditions was significantly different from priming, [ $\chi^2(1) = 18.059$ ,  $p = 0.001$ ] and [ $\chi^2(1) = 19.267$ ,  $p < 0.001$ ], respectively. The proportion of students in the single-concept-mastery condition was not significantly

different than priming [ $\chi^2(1) = 1.099$ ,  $p = 1.000$ ]. Comparison between the single-concept-mastery and synthesis-mastery conditions also showed a significant difference [ $\chi^2(1) = 9.600$ ,  $p = 0.01$ ]. There was no difference between the synthesis-mastery and synthesis-recognition conditions. Taken together these results support the trend suggested by the total score on the target synthesis problem: analogical comparison significantly increased recognition of centripetal acceleration, but only when students compared synthesis examples.

### C. Discussion

There are four important findings from this experiment. First and foremost, training via analogical comparison of worked examples was effective in improving student scores on a target synthesis problem relative to control. Second, training via analogical comparison behaved markedly better than priming (via problem solving exercises). Third, the effectiveness of analogical comparison depended significantly on the type of worked examples to be compared, but not the specific nature of the comparison prompts.

The fourth finding is that student success on the problem was bottlenecked primarily by their ability to recognize the presence of the centripetal motion constraint. Given the analysis of student concept recognition across control and the four treatment conditions, the non-normal distributions of student total scores are telling: without intervention students primarily solved the target synthesis problem as if it were a single-concept problem. Once they were able to conceptually recognize conservation of energy and centripetal acceleration, they almost universally shifted from only applying one concept to correctly applying both. As a result, one of the strongest potential gains from training via analogical comparison—at least, specifically in the context of synthesis problems—may be the improvement of student conceptual recognition.

Moreover, the finding that comparisons of synthesis worked examples was significantly more effective than comparisons of single-concept examples (using the same prompts) supports the importance of this full structural transfer [38]. Combined with the fact that priming students via explicit problem-solving practice was not significantly better than course instruction alone, these results are also prescriptive. Namely, these results weaken the often-held assumption that students can repeatedly practice physics concepts in isolation, and simultaneously expect success on problems that combine them. Instead, these results suggest that integration does not happen spontaneously, at least not for solving complex physics problems. As such, in contrast to the vast majority of introductory homework problems and end-of-chapter exercises, success with synthesis problems may best be facilitated by explicit practice with synthesis problems.

We consider two potential explanations for the finding of no significant difference between the comparison prompts

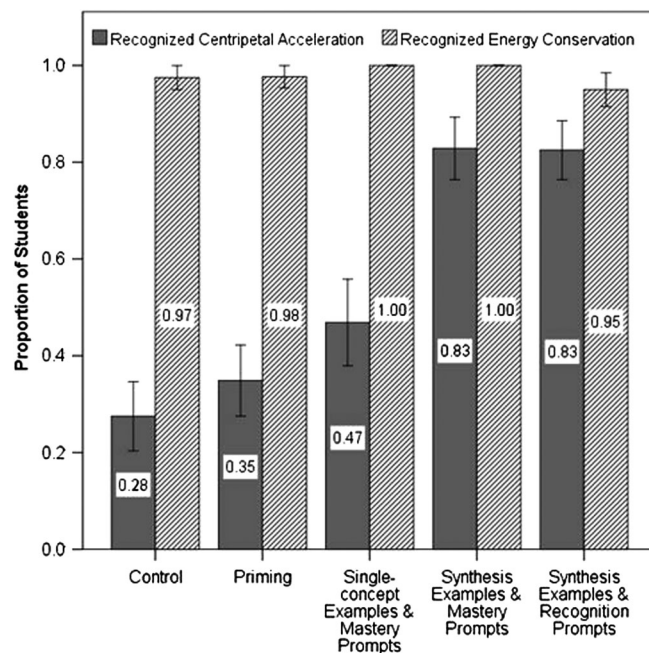


FIG. 4. Proportion of students in each condition demonstrating recognition of centripetal acceleration and energy conservation. Error bars are standard errors.

focused on individual concept application and those focused on overall concept recognition and structure. First, noting that both conditions were quite successful relative to additional practice solving single-concept problems, it is possible that ceiling effects are limiting the possibility for any difference in overall effectiveness. Second, given that students tend to switch from applying a one-concept only solution on the target synthesis problem in the control and priming conditions to providing fully correct solutions after completing the analogical comparison tasks, it is possible that the greatest benefit of the analogical comparison prompts is that they force the student to sufficiently encode the two synthesis examples; because the students are so successful with the recognition of the individual concepts once they have encoded the combined concept structure, the specific comparisons themselves are less important.

Overall, the results of this experiment suggest that analogical comparison can be an effective technique for training students to solve synthesis problems—at least for problems that demonstrate the level of conceptual and mathematical complexity represented by the target synthesis problem. However, even considering the vast array of potential physics concepts and combinations, the success here is promising. After all, this target synthesis problem and the corresponding base worked examples are already more involved than other previous, successful examples of analogical comparison [38,39]. Still, the results here suggest natural follow-up questions: Is analogical comparison effective for a synthesis problem with increased complexity? Is analogical comparison as effective as another known effective problem solving intervention method, namely, self-explanation?

### III. EXPERIMENT 2: ANALOGICAL COMPARISON VS EXPLANATION

#### A. Method

##### 1. Design

The goals of experiment 2 were to explore the effectiveness of analogical comparison in the context of more complicated introductory-level synthesis problems [see the target problem in Fig. 5(c)] and to compare the effectiveness of analogical comparison to the method of self-explanation. A full solution of the target problem requires three main conceptual components: simple circuit analysis (Ohm’s law), induced EMF (Faraday’s law), and magnetic force. In particular, the concepts of magnetic force and Faraday’s law represent distinct and documented challenges for students [47,48]. In the case of magnetic force, students must successfully interpret cross products. Faraday’s law requires an implicit understanding of magnetic flux and the consideration of direction via Lenz’s law. Correspondingly, a compact solution of the target synthesis problem utilizes not only three basic physics equations, but considerably more algebraic manipulations than that required to solve

experiment 1. Along with the target synthesis problem, we designed a corresponding set of single-concept and synthesis worked examples [see Figs. 5(a) and 5(b)].

Experiment 2 had six conditions (see Fig. 1). Four of the conditions (control and three analogical comparison conditions) were similar in structure to experiment 1. The fifth condition was aimed at avoiding a potential shortcoming of the other analogical comparison conditions. This condition represented a “best-effort” attempt to scaffold comparisons through a mix of recognition and single-concept focused prompts. In particular, the prompts explicitly addressed the concept of induced emf and its application within the two synthesis worked examples. For example, one prompt in this condition states “One of the important concepts is that of an induced emf due to a changing magnetic flux. Compare the physical reason for a changing magnetic flux in each of the problems. Explain your answer highlighting any differences between the two problems.” In contrast, the other conditions did not explicitly invoke the concept of Faraday’s law. Further, the combined “best-attempt” condition tried to scaffold comparison that followed the worked examples: starting with recognition of the relevant concepts, through consideration of the physical context for Faraday’s law, and then subsequent application of the component concepts in the worked examples. The other analogical comparison conditions were restricted to focus on either recognition of the concepts or their application as in experiment 1. Full versions of all worked examples and corresponding comparison prompts are included in Appendix B.

Finally, the sixth condition was the self-explanation condition. This condition simply prompted students to explain (write) both of the synthesis worked examples as if to a friend, but did not invoke explicit comparison between the two. Ample space was provided for the explanation.

##### 2. Participants

Tasks were administered during Spring 2016. Participants were students in the second semester of a calculus-based introductory electromagnetism course at The Ohio State University. Students participated as part of a 1-hr-long flexible homework assignment for course credit. Students completed the flexible assignment over a three week window, after course instruction on Faraday’s law, and in close proximity to a course midterm covering the relevant material. A total of 254 students were randomly assigned into one of the six study conditions.

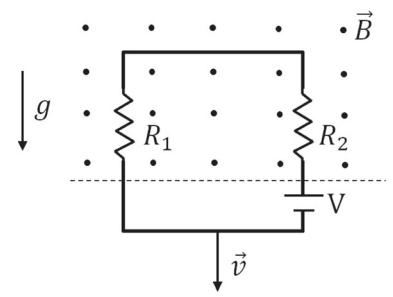
##### 3. Administration

Students completed the training tasks and target synthesis problem in individual carrels in a quiet room. An equation sheet similar to those used in the course was provided to all students. Tasks were administered and collected by the proctor one at a time, and students were allowed to work at their own pace. Whereas students in the analogical comparison conditions were given all relevant



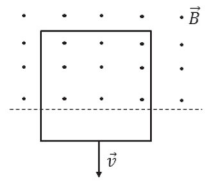
**(a) Synthesis Problem – Worked Example**

A square circuit of length  $L = 40\text{cm}$  and unknown mass  $m$  is falling (gravity directed as shown) through a region with a uniform and perpendicular magnetic field  $B = 2.3\text{ T}$  directed out of the page. The resistors have resistances of  $R_1 = 10\Omega$  and  $R_2 = 25\Omega$ , and the voltage of the battery is  $9\text{V}$ . At the instant shown, when part of the circuit is still in the region with a magnetic field, the velocity of the object is constant and equal to  $10.0\text{ m/s}$  downward. Find the mass of the circuit.



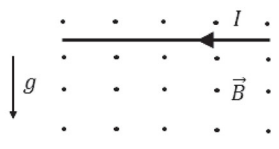
**(b) Single Concept Example #1**

A square loop of conducting wire, length  $40\text{ cm}$  is being pulled at a constant speed  $10.0\text{ m/s}$  through a region with a uniform magnetic field  $B = 2.3\text{ T}$  directed out of the page as shown. At the instant shown, when part of the circuit is still in the region with a magnetic field, find the direction and magnitude of the induced *emf* in the loop.



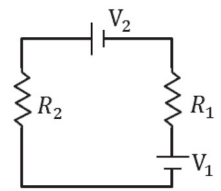
**Single Concept Example #2**

A current carrying wire of length  $L = 40\text{cm}$  is falling (gravity directed as shown) at a constant velocity through a region with a uniform and perpendicular magnetic field  $B = 2.0\text{ T}$  directed out of the page. If the current is  $0.5\text{A}$ , find the mass of the wire.



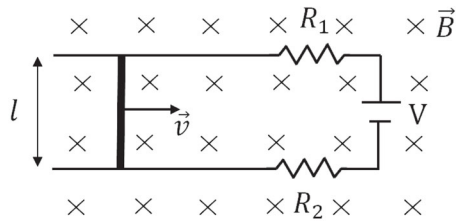
**Single Concept Example #3**

Given  $V_1=9\text{V}$ ,  $V_2=5\text{V}$ ,  $R_1=10\Omega$  and  $R_2=15\Omega$ , find the magnitude of the current in the circuit shown.



**(c) Target Synthesis Problem**

A conducting bar of negligible resistance and length  $30\text{ cm}$  is sliding along a pair of frictionless rails at a speed of  $v = 4.0\text{ m/s}$  as shown in the figure. A uniform magnetic field of  $B = 11.0\text{ T}$  is directed into the page. The battery voltage is  $24\text{V}$  and the resistances are  $R_1 = 10\Omega$  and  $R_2 = 15\Omega$ .



In order to keep the rod moving at a constant speed, an additional force needs to be applied on it. Find the magnitude and direction of this applied force.

FIG. 5. An example synthesis problem provided as a worked example (a), corresponding single concept problems (b), and the target synthesis problem (c).

worked examples and prompts together, students in the self-explanation condition were given one synthesis worked example to summarize at a time. Students first completed their selected training, followed by 10–15 min of unrelated physics tasks, and then the target synthesis problem. Students in the control condition completed a set of unrelated physics tasks and the target synthesis problem.

#### 4. Method of analysis

A rubric for assessing student solutions of the target synthesis problem was determined by two of the authors. The rubric is shown in Table IV. As in experiment 1, recognition of component physics concepts was coded generously, but required the student to commit to using the

concept as part of their solution. A random sample of 25 student solutions was coded by two researchers with an intercoder agreement of 84%. Disagreements were discussed and resolved.

#### B. Results

Student final course grades in their introductory electromagnetism class were collected and compared across conditions. The same cuts conducted in experiment 1 were applied to eliminate outliers. Students must have completed the course (removing no students), and have scored no lower than 2 standard deviations below the mean (removing a total of 9 students, ranging from 1 to 4 students per condition). A one-way ANOVA of course

TABLE IV. Scoring rubric for the target synthesis problem. 1 point was awarded for each item, for a total of 10 points.

Recognition		
+1	Faraday's law or induced emf	
+1	Ohm's law	
+1	Magnetic force	
Application		
+1	Correct induced emf	
+1	Correct induced emf direction or Lenz's law	
+1	Ohm's law: series circuit or combined resistors in series	
+1	Ohm's law: combined voltage sources	
+1	Correct magnetic force equation	
+1	Correct applied force direction	
Calculation		
+1	Correct calculations and no mathematical mistakes	

grade showed no significant differences across conditions [ $F(5, 239) = 0.554$ ,  $p = 0.735$ ].

Mean scores on the target synthesis problem are shown in Table V. We first compared the interventions to control. A one-way ANOVA showed significant differences between conditions [ $F(5, 239) = 5.961$ ,  $p < 0.001$ ]. A Tukey HSD *post hoc* showed significant differences between self-explanation and control ( $d = 1.14$ ,  $p < 0.001$ ) and synthesis-combined prompts and control ( $d = 0.67$ ,  $p = 0.018$ ). The other analogical comparison conditions were not significantly different from control: single-concept-mastery ( $d = 0.18$ ,  $p = 0.959$ ), synthesis-mastery ( $d = 0.44$ ,  $p = 0.363$ ), and synthesis-recognition ( $d = 0.44$ ,  $p = 0.296$ ).

A one-way ANOVA was conducted between intervention conditions to test for hypothesized differences between single-concept worked examples and synthesis worked examples, between mastery and recognition prompts, and between self-explanation and the combined, best-effort analogical comparison condition. The one-way ANOVA showed significant differences between interventions

TABLE V. Mean score on target synthesis problem out of a maximum score of 10 points by condition. Errors shown are standard errors. Analogical comparison conditions are labeled AC.

Condition	Score $\pm$ SE	$N$
Control	5.19 $\pm$ 0.35	42
AC: Single concept examples with mastery prompts	5.60 $\pm$ 0.35	42
AC: Synthesis examples with mastery prompts	6.16 $\pm$ 0.35	38
AC: Synthesis examples with recognition prompts	6.19 $\pm$ 0.35	42
AC: Synthesis examples with combined prompts	6.75 $\pm$ 0.37	40
Self-explanation	7.54 $\pm$ 0.28	41

[ $F(4, 198) = 4.697$ ,  $p = 0.001$ ]. A Tukey HSD *post hoc* showed no significant difference in total score on the target synthesis problem between single-concept-mastery and synthesis-mastery ( $d = 0.25$ ,  $p = 0.776$ ). There was no significant difference between synthesis-mastery and synthesis-recognition ( $p = 1.00$ ). Finally, although there were no significant differences between self-explanation and synthesis-combined ( $d = 0.38$ ,  $p = 0.48$ ), the self-explanation treatment significantly outperformed all of the other analogical comparison conditions: Single-concept-mastery ( $d = 0.95$ ,  $p = 0.001$ ), synthesis-mastery ( $d = 0.70$ ,  $p = 0.042$ ), and synthesis-recognition ( $d = 0.66$ ,  $p = 0.042$ ).

To sum up, only the best-effort attempt at analogical comparison (synthesis worked examples using a combination of scaffolded comparison prompts) and the self-explanation intervention were significantly better than course instruction alone (control) in terms of overall student performance on the target synthesis problem. In addition, there were no statistical differences in the effectiveness of analogical comparison based on either the type of worked example or the type of comparison prompts. Finally, summarization via self-explanation was the most effective intervention, with significant differences in student performance versus all analogical comparison conditions except for the highly scaffolded best-attempt condition.

In addition to overall performance on the target synthesis problem, we compared student conceptual recognition across conditions. The proportion of students recognizing each component physics concept is shown in Fig. 6. The vast majority of students recognized and utilized Ohm's law ( $\geq 93\%$ ) and magnetic force due to a current carrying wire ( $\geq 83\%$ ) across all conditions.

A chi-squared test across treatment conditions showed there was a significant difference in the proportion of students recognizing and utilizing Faraday's law on the target synthesis problem [ $\chi^2(4) = 25.543$ ,  $p < 0.001$ ]. To test for hypothesized differences between the interventions, *post hoc* comparisons between treatments were conducted using pairwise chi-squared tests with a Bonferroni correction for multiple comparisons ( $n = 3$ ). The adjusted  $p$  values are reported. To test for the hypothesized difference due to type of worked example, a comparison of single-concept-mastery and synthesis-mastery showed no significant difference in the proportion of students recognizing Faraday's law [ $\chi^2(1) = 0.029$ ,  $p = 1.00$ ]. To test for the hypothesized difference due to type of comparison prompt, a comparison of synthesis-mastery and synthesis-recognition also showed no significant difference [ $\chi^2(1) = 1.749$ ,  $p = 0.558$ ]. Finally, a comparison between synthesis-combined and self-explanation showed a significant difference in the proportion of students recognizing Faraday's law on the target synthesis problem [ $\chi^2(1) = 6.316$ ,  $p = 0.036$ ].

Overall, there was a statistically significant difference in the proportion of students identifying Faraday's law between the self-explanation condition and the best

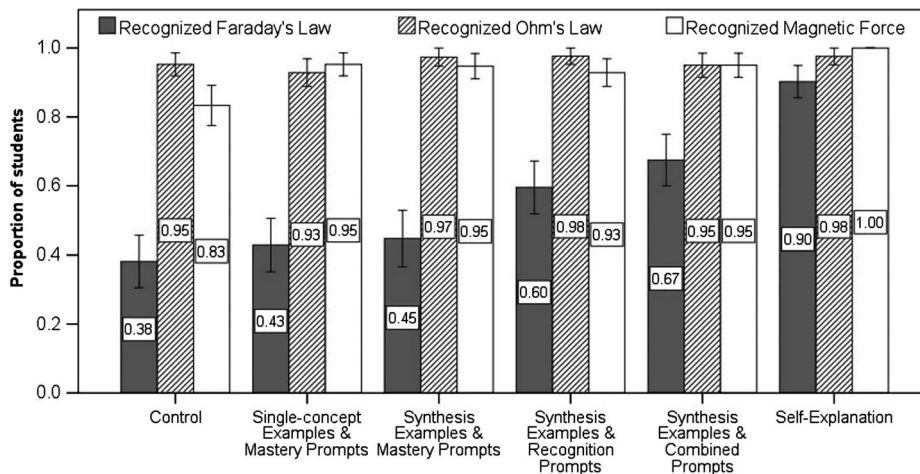


FIG. 6. Proportion of students in each condition demonstrating recognition of Faraday's law, Ohm's law, and magnetic force on the target synthesis problem. Error bars are standard errors.

attempt, combined analogical comparison treatment with the self-explanation group outperforming the combined analogical comparison group. However, there was no significant effect from either the type of worked example or the type of comparison prompts on student recognition of Faraday's law for the target synthesis problem.

In order to further examine the effect of training on how students approached the target synthesis problem, we compared the proportion of students across conditions who recognized and applied Faraday's law, Ohm's law, and the magnetic force on a current carrying wire, and also explicitly calculated a total current in the circuit that combined the voltage due to the battery with the induced emf (but not necessarily with correct directions or magnitudes). This combination was used to represent the minimum structure necessary for a correct approach to the target synthesis problem. The proportion of students successfully meeting this threshold is shown in Fig. 7.

A chi-squared test across treatment conditions showed there was a significant difference in the proportion of students meeting this structural threshold [ $\chi^2(4) = 32.482$ ,  $p < 0.001$ ]. To test for hypothesized differences between the interventions, *post hoc* comparisons between treatments were conducted using pairwise chi-squared tests with a Bonferroni correction for multiple comparisons ( $n = 3$ ). Tests for hypothesized differences due to type of worked example and type of prompt showed no significant differences (i.e., comparison between Single-concept-mastery and synthesis-mastery and between synthesis-mastery and synthesis-recognition, respectively). However, comparison between synthesis-combined and self-explanation did show a significant difference in the proportion of students who applied all component concepts and calculated the combined total current [ $\chi^2(1) = 8.320$ ,  $p = 0.012$ ]. Simply put, the difference does not come from the first four treatment groups in Fig. 7; instead, it comes from the last group.

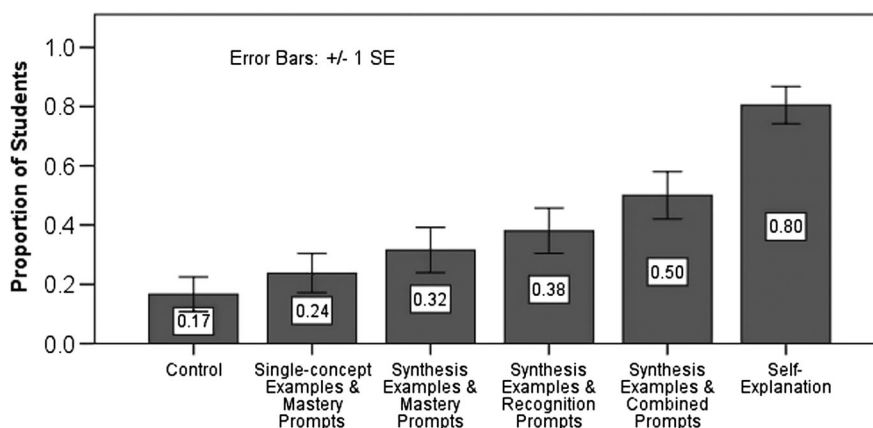


FIG. 7. Proportion of students in each condition who employed the correct solution structure on the target synthesis problem. Error bars are standard errors.

### C. Discussion

There are two main takeaways from experiment 2. First, of the analogical comparison interventions, only the scaffolded best-attempt analogical comparison condition performed significantly better than control in terms of overall performance on the target synthesis problem. Moreover, with other factors controlled, single-concept and synthesis worked examples produced similar results. The case was the same comparing prompts focused on mastery or recognition. Second, summarization via self-explanation of synthesis worked examples alone was not only the most effective intervention in improving students' overall performance on the target problem, but it also significantly improved the recognition and use of Faraday's law compared to every other intervention, including the best-attempt analogical comparison condition.

The increased complexity of the target problem in experiment 2 (compared to experiment 1) manifested itself in the details of student performance on the problem. In experiment 1, student conceptual recognition was the dominant bottleneck to correctly solving the target problem; once that difficulty was successfully overcome by the analogical comparison interventions, students correctly applied the component physics concepts. In experiment 2, recognition alone was not enough to guarantee a completely correct solution. The summarization condition resulted in 90% of students recognizing and applying Faraday's law (and even more recognizing the other two component concepts), but the mean score was still only 7.54/10, in part due to remaining difficulties applying Lenz's law and cross-product-based reasoning.

The synthesis worked examples were also more complex. We suggest that this manipulation of the synthesis worked example complexity is a likely reason for the subsequent lack of statistically significant gains for those analogical comparison conditions. The importance of the increased complexity seems even more likely considering that the only statistically different analogical comparison condition versus control was the combined analogical comparison treatment. The combined prompts provided more information to the student (explicitly identifying Faraday's law) and scaffolded the comparison of the two worked examples more extensively than any of the other analogical comparison interventions. In order for the analogical comparison to be effective with more complicated base examples, additional scaffolding was necessary.

The necessity for additional scaffolding may not be too surprising given the cognitive load and increased demands from the more complicated worked examples. However, it is important to note that once again comparisons of considerably simpler single-concept examples did not significantly improve student performance on the target problem. It appears that the structure of a synthesis problem cannot simply be broken into parts, even when students are asked to compare the set of parts one after another.

Unfortunately, this also suggests that analogical comparison with single-concept examples alone is unlikely to be an effective way to help students reduce the cognitive load inherent to complicated synthesis problems.

On the other hand, despite the lack of instructional scaffolding or information beyond the worked example, students in the self-explanation condition performed significantly better than the control group on the target synthesis problem. Further, students in the self-explanation condition performed significantly better in recognizing the need for Faraday's law compared to the best analogical comparison condition—despite the fact that the best-effort analogical comparison condition explicitly pointed out the concept as one of its comparison prompts. This result was further supported by the proportion of students who generated the correct solution structure on the target problem—using all three identified physics concepts and explicitly combining both the induced emf and the battery voltage (cf. Fig. 7).

Given these results, there are several possible explanations for the relative success of the self-explanation condition. First, self-explanation might have been more successful because the task of explaining each problem, one at a time, allowed students to better encode the entire structure of each independent worked example. In contrast, students in the analogical comparison condition may have made sense of smaller-grain component concepts across the base examples, but without encoding the full structure of either example. This lack of encoding the full structure in the analogical comparison conditions could be due to a simple failure of students to satisfactorily read through the base worked examples—that is, beyond what was necessary to answer each specific invoked comparison—or more nuanced differences in how the students extracted the structure of the provided solutions. Experiment 3 was designed to help exclude the first possibility. Two mechanisms proposed by Chi [49] to explain the success of self-explanation, namely, inference generation and mental-model revision, may account for other underlying differences in student encoding.

The mechanism of inference generation suggests that summarization via self-explanation may have led to higher performance by prompting students to fill in necessary information and reasoning steps missing in the worked examples. There is some evidence for such an effect: most students included not only the relevant physics concepts in their summaries, but also additional justifications. For example, Faraday's law and induced emf were addressed in 97% of student summaries and 57% of students described the physical reason for the two cases of changing magnetic flux. Moreover, student summaries seemed to follow the reasoning of the provided solutions. The majority of student summaries discussed the concepts in the order that they were presented in the worked solutions while also identifying important intermediate quantities. In particular,

students often included explicit identification of the electric current as a crucial unknown, e.g., “The next step is to solve for the current since  $L$ ,  $B$ ,  $g$  are all known...however, to properly solve this we need to solve for the  $\mathcal{E}$  created by the  $B$  field.” Although that recognition was probably driven in part by students relying on a given-unknown problem solving heuristic, it may have allowed students to identify the electric current as a structural connection between the physics concepts in the problem. In contrast, the connections between the analogical comparisons prompts may not have been as strongly internalized by students, even though the combined analogical comparison condition invoked comparisons that explicitly targeted those exact elements.

The second mechanism, mental-model revision, could also account for the relative success of the summarization condition. The target synthesis problem—and its potentially novel combination of an induced emf and a battery—represented a significant challenge for students. It is possible that students had an incomplete or disjointed prior understanding of electromotive force. If that was the case, summarization via self-explanation may have helped students to reconcile their mental model with the worked examples. Future work is needed to differentiate between these possible mechanisms.

#### IV. EXPERIMENT 3: ANALOGICAL COMPARISON WITH EXPLANATION

##### A. Method

###### 1. Design

The third experiment was built upon the previous study to both replicate the relative success of the self-explanation intervention and test whether analogical comparison and self-explanation of the individual worked examples can be combined for further benefit. As such, this study used the same target problem employed in experiment 2. In addition to a no-training control, 4 treatments were included in the experimental design. To replicate the results of experiment 2, we once again included a best-effort analogical comparison condition and self-explanation condition, using the same worked examples as before. The prompts for the analogical comparison condition were similar to those used previously and explicitly invoked the concept of Faraday’s law. Given student’s prior difficulty with applying the individual concepts—in particular, difficulties with direction-based considerations due to Lenz’s law and cross products—we made adjustments to the prompts to encourage further comparison of the important directions identified in the worked examples. The full set of prompts is included in Appendix C.

The other two conditions were designed to test whether an increased emphasis on encoding the individual worked examples before analogical comparison would increase student performance on the target synthesis problem. First, we included an “annotation” condition that asked students

to very briefly label the two individual worked examples. The prompts were intended to be brief checks to verify that students had successfully read through the example solutions. As such, they simply asked students to identify both the concepts used and the basic goal of sections of the worked examples (i.e., “Ohm’s law” and “find the total current”). In contrast to the self-explanation prompt, students were only asked to label the individual solutions rather than provide additional justifications or elaborations. The presentation of these reading annotations is shown in Appendix C.

The last condition sought to test the hypothesized idea that self-explanation and analogical comparison could be combined for additional benefit. Given the increased complexity of the base worked examples, we posited that inviting students to first summarize the individual worked examples independently would facilitate subsequent analogical comparison. Consequently, students may have a better holistic understanding of the base examples and how individual comparisons fit within the two overall solution structures, rather than viewing them as a set of unconnected, piecewise comparisons. As such, the combined self-explanation and analogical comparison condition asked students to first briefly summarize each worked example before prompting them to compare across the two worked examples.

###### 2. Participants

Tasks were administered during Fall 2016. Participants were students in an off-sequence calculus-based introductory electromagnetism course at The Ohio State University who participated as part of a 1-hr-long flexible homework assignment for course credit. The flexible homework assignment was administered over a three week period near the end of the semester, approximately one month on average after course instruction on Faraday’s law. A total of 232 students were randomly assigned into one of the five study conditions.

###### 3. Administration

Students completed the training tasks and target synthesis problem in individual carrels in a quiet room. Tasks were administered and collected by the proctor one at a time and an equation sheet was provided to all students. Students first completed their selected training, followed by 10–15 minutes of unrelated physics tasks, and then the target synthesis problem, as shown in Fig. 1. Whereas students in the analogical comparison and annotation conditions were given all relevant worked examples and prompts together, students in both the self-explanation and combined analogical comparison and self-explanation conditions were given only a single synthesis worked example to summarize at a time. Students in the control condition once again completed a set of unrelated physics tasks and the target synthesis problem.

Student time on task was digitally recorded by the proctor. Because of time constraints from the testing format, students in the combined self-explanation and analogical comparison condition were given a time limit of approximately 7 min per individual summary. Time on task during the training interventions was collected for the vast majority of students (95%).

#### 4. Method of analysis

Student solutions to the target synthesis problem were graded with the same rubric used in experiment 2 (shown in Table IV). A random sample of 25 student solutions was coded independently by two researchers. Any differences in coding were discussed and resolved leading to an intercoding agreement of 88%.

### B. Results

Student final course grades in their introductory electromagnetism class were collected and compared across conditions. The same cuts conducted previously were applied to eliminate outliers. Students must have completed the course (removing no students), and have scored no lower than 2 standard deviations below the mean (removing a total of 6 students, ranging from 0 to 2 students per condition). Almost all students satisfactorily completed the training tasks. One student who did not complete the training task was removed from the combined analogical comparison and self-explanation condition. A one-way ANOVA of course grade showed no significant differences across conditions [ $F(4, 220) = 0.522$ ,  $p = 0.719$ ].

Mean scores on the target synthesis problem are shown in Table VI. First, interventions were compared to control. A one-way ANOVA showed significant differences in total score on the target synthesis problem between conditions [ $F(4, 220) = 11.351$ ,  $p < 0.001$ ]. A Tukey HSD *post hoc* showed significant differences for all treatment conditions compared to control: Comparison-only ( $d = 0.77$ ,  $p = 0.003$ ), comparison and annotations ( $d = 0.92$ ,  $p = 0.001$ ), self-explanation ( $d = 1.05$ ,  $p < 0.001$ ), and comparison and self-explanation ( $d = 1.44$ ,  $p < 0.001$ ).

A one-way ANOVA was conducted between intervention conditions to test for hypothesized differences in student

performance on the target synthesis problem. The one-way ANOVA showed significant differences between the treatments [ $F(3, 176) = 3.002$ ,  $p = 0.032$ ]. In order to test for hypothesized differences, a Tukey HSD *post hoc* was conducted. The *post hoc* tests showed no significant difference in total score on the target synthesis problem between comparison-only and comparison and annotations ( $d = 0.08$ ,  $p = 0.977$ ), nor between comparison-only and self-explanation ( $d = 0.23$ ,  $p = 0.653$ ), but there was a significant difference between comparison-only and comparison and self-explanation ( $d = 0.58$ ,  $p = 0.030$ ).

In summary, the combination of self-explanation and analogical comparison was significantly better than analogical comparison alone. However, there was no significant difference between students who completed only the analogical comparison prompts and the students who were explicitly asked to read through each example and provide brief annotations before making comparisons. There was also no significant difference between self-explanation and analogical comparison ( $d = 0.23$ ), though the trend was in the same direction as in experiment 2 ( $d = 0.38$ ), with self-explanation performing nominally better than analogical comparison. Taken together, this replication may suggest a small, but potentially significant effect ( $d \approx 0.3$ ).

The proportion of students recognizing each component physics concept and employing it as part of their solution on the target synthesis problem is shown in Fig. 8. Across all conditions, the majority of students successfully recognized and utilized Ohm's law ( $\geq 80\%$ ) and magnetic force due to a current carrying wire ( $\geq 93\%$ ). A chi-squared test across all conditions showed there was a significant difference in the proportion of students recognizing and utilizing Faraday's law on the target synthesis problem [ $\chi^2(4) = 39.140$ ,  $p < 0.001$ ]. However, such a difference went away if only the treatment groups were compared [ $\chi^2(3) = 5.658$ ,  $p = 0.129$ ]. This suggests that while the four treatment conditions all significantly improved concept recognition versus control, they did not differ significantly from one another.

In addition, we compared the proportion of students who met a minimum threshold for a correct approach on the target problem, namely, recognize and apply all three concepts and calculate a total current using both the induced emf and battery voltage. The results are shown in Fig. 9. Along with clear differences between the interventions and control, a chi-squared test was used to compare students who did not self-explain the individual worked examples with those who did as part of their training. There was a significant difference in the proportion of students who met the proposed threshold on the target synthesis problem [ $\chi^2(1) = 24.360$ ,  $p < 0.001$ ]. Students who self-explained the individual worked examples as a part of their training were significantly more likely to recognize all three concepts and combine the two sources of emf.

TABLE VI. Mean score on target synthesis problem out of a maximum score of 10 points by condition. Errors shown are standard errors.

Condition	Score $\pm$ SE	<i>N</i>
Control	4.04 $\pm$ 0.30	45
Analogical comparison	5.74 $\pm$ 0.35	46
Analogical comparison and reading annotations	5.93 $\pm$ 0.32	45
Self-explanation	6.29 $\pm$ 0.34	45
Analogical comparison and self-explanation	7.07 $\pm$ 0.34	44

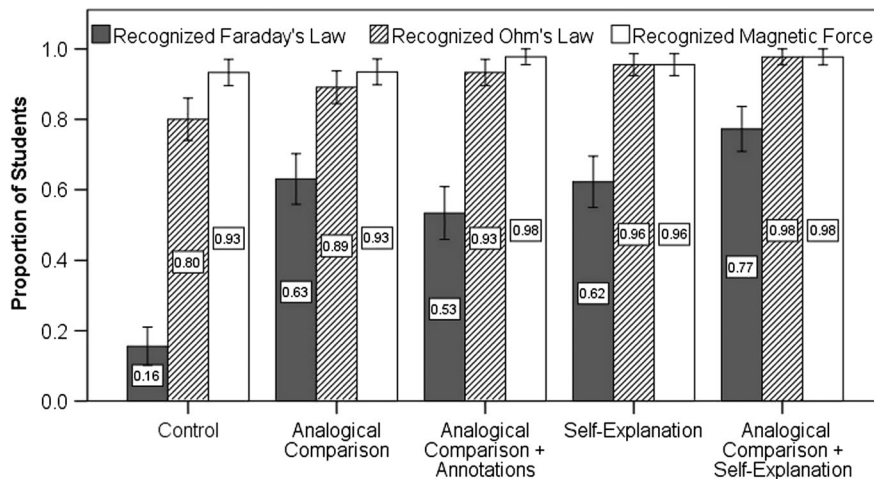


FIG. 8. Proportion of students in each condition demonstrating recognition of Faraday’s law, Ohm’s law, and magnetic force on the target synthesis problem. Error bars are standard errors.

Time on task during training was recorded for the vast majority of students (95%) and compared across the four intervention conditions. Students without timing data were removed from the subsequent analysis, resulting in the corresponding boxplots presented in Fig. 10. A median test showed significant differences in time spent during training across the conditions. In particular, comparison and annotation and comparison and self-explanation represent an increase of approximately 20% and 35% (5 and 8 min, respectively) in training time over the comparison-only condition.

To assess the effect of time on task and student aptitude in determining student performance on the target synthesis problem, we compared total score on the target synthesis problem between the comparison and self-explanation and comparison-only conditions using a general linear model, accounting for a main effect (condition) and two covariates (course grade and time on task). Outliers in time on task were removed based on inspection of the boxplot presented in Fig. 10. Both course grade (partial  $\eta^2 = 0.14, p = 0.001$ )

and time on task (partial  $\eta^2 = 0.05, p = 0.048$ ) were found to significantly predict student performance on the target synthesis problem. Condition was marginally significant (partial  $\eta^2 = 0.04, p = 0.093$ ). These results suggest that student aptitude, as measured by course grade, is the strongest predictor of subsequent performance on the target synthesis problem. Moreover, though the model lacked the statistical power to observe the effect at the 0.05 level, it suggests some evidence that combining analogical comparison and summarization may provide a small, positive effect beyond that accounted only by the additional time on task.

### C. Discussion

Taken together, the results from experiment 3 support three broad conclusions. First, both analogical comparison

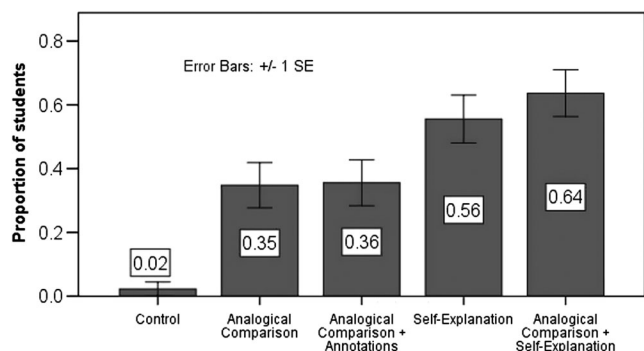


FIG. 9. Proportion of students in each condition demonstrating the correct general solution structure on the target problem. Error bars are standard errors.

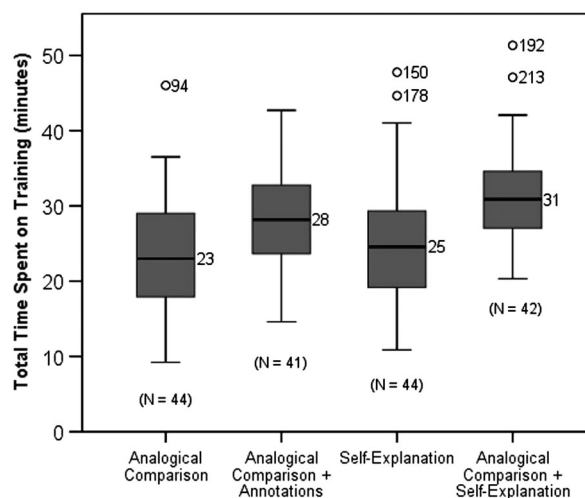


FIG. 10. Box plot of student time spent on training by condition.

and self-explanation were effective in improving student performance on the target synthesis problem versus control, though there were differences between analogical comparison and self-explanation in the proportion of students employing the correct solution structure on the target problem. Second, the combination of self-explanation and analogical comparison was significantly more effective than analogical comparison alone. Third, student annotations of the worked examples were accurate, but did not significantly improve student performance on the subsequent target problem relative to analogical comparison alone.

While the mean score on the target synthesis problem for the control condition in this study was 4.04/10, with only 16% of students recognizing and applying Faraday's law, the mean score on the same target problem in experiment 2 was 5.19/10, with 38% of students recognizing and applying Faraday's law. In order to provide context for these results in light of previous findings, we note that the population sample used in this study differed in two potentially important ways from the sample in experiment 2. First, though the two samples were drawn from different semesters of the same introductory electromagnetics course, students in this study completed the course off sequence. Second, students in this study completed the training task and target problem over a three week period near the end of the semester, approximately one month on average after course instruction on Faraday's law. In contrast, students in experiment 2 completed the training over a similar period that began closely following in-course instruction, and in proximity to an in-course exam on the relevant topics. Although we cannot exclude differences in the on- and off-course sequence, previous research tracking student understanding over the duration of an introductory course suggests timing differences between in-course instruction and administration of the training and target synthesis problem are a potential explanation for the observed differences in baseline student performance between the two studies [50,51].

This difference in baseline performance suggests an important distinction when discussing the relative effectiveness of the analogical comparison and self-explanation conditions. This study found no significant differences in either total score on the target synthesis problem or the proportion of students recognizing and applying Faraday's law between the self-explanation and analogical comparison conditions. In contrast, experiment 2 found that the self-explanation condition resulted in significantly more students recognizing and applying Faraday's law than in the analogical comparison condition. Moreover, the overall proportion of students in the self-explanation condition who were able to construct the correct solution structure for the target synthesis problem was considerably lower in experiment 3 than in experiment 2.

At the same time, these findings support the hypothesis that there is a meaningful difference in how successfully students encoded the base worked examples between the

self-explanation and analogical comparison conditions: in particular, the replicated finding that significantly more students in the self-explanation condition constructed the correct solution structure for the target synthesis problem than in the analogical comparison condition. Moreover, the finding of no significant difference between analogical comparison alone and analogical comparison with annotations suggests that these differences are likely not due to students simply failing to sufficiently read through the worked examples in the comparison conditions. In other words, students in the annotation condition satisfactorily labeled physical concepts and key steps within both worked examples with no difference in student performance on the target synthesis problem compared with just analogical comparison alone.

As a result, it is more likely that the success of analogical comparison of the two worked examples was limited by cognitive load and not a failure of students to appropriately attend to the task—here, and potentially in experiment 2. There are several additional pieces of evidence. First, student performance on the target problem once again indicated significant and persistent student difficulty with applications of the single concepts—in particular, determining physical directions associated with Lenz's law and cross products. Unlike experiment 1 where students demonstrated a high degree of mastery of the two component concepts after recognizing the need for their simultaneous application, students continued to struggle with these single-concept difficulties regardless of intervention. Second, there was a significant difference between analogical comparison alone and analogical comparison after self-explanation, in terms of both total score and the proportion of students demonstrating the correct solution structure on the target synthesis problem, suggesting that the initial self-explanation did aid comparison.

One limitation of this study is that it was unable to make a definitive distinction between the value added by the combination of analogical comparison and self-explanation and the corresponding additional time on task. This limitation was a consequence of constraints involving the administration of the task, available contact time, and number of participants. However, the overall significant difference between analogical comparison alone and the combination of self-explanation and analogical comparison is still of particular value, as it suggests at least one pedagogically relevant way to help students analyze similar, complicated synthesis problems. In other words, the additional approximate 8 min to time on task from the combined intervention was well spent, engaging, and resulted in significant gains on the target synthesis problem; in contrast, the additional time spent required by the reading annotations was not inherently productive.

## V. CONCLUSION

Taken as a whole, the three experiments demonstrate that the instructional methods of analogical comparison and



self-explanation of worked examples can successfully be extended to improve student performance on target synthesis problems. As such, this work represents a novel contribution to the study of these techniques beyond previous work predominantly studying their application with single-concept examples. Moreover, the results of these experiments suggest several principles regarding the conditions under which the two instructional methods are likely to be effective.

First, analogical comparison only resulted in significant increases in student performance over baseline when the comparisons were invoked between synthesis examples. There were no significant improvements when students were asked to compare corresponding sets of single-concept problems, despite the same prompts and surface features in the worked examples. This is particularly important given the increase in the target synthesis problem complexity from experiment 1 to 2—in terms of both the conceptual difficulty represented by the involved concepts and the requisite number of algebraic manipulations necessary for a complete solution. The finding that breaking the base worked examples into component parts was not significantly different than either unguided problem solving practice (experiment 1) or course instruction alone (experiments 1 and 2) emphasizes the importance of the combined structure and joint application of concepts within synthesis problems. Even with explicit and sequential comparisons of the component parts, students cannot be expected to successfully transfer those parts to a novel synthesis problem without extensive efforts to explicitly scaffold the missing structure.

Second, these results show that while analogical comparison and self-explanation of worked examples can be effective in improving student conceptual recognition and the use of the correct solution structure on a target synthesis problem, pervasive difficulties associated with single-concept mastery—such as how to apply Lenz's law—may not be as successfully remedied. In part, this suggests that synthesis problems and the instructional methods used here can best be employed to help students explicitly practice concept recognition, as opposed to the plug-and-chug heuristics often associated with single-concept problems. Meanwhile, issues of single-concept mastery may benefit from complementary and targeted practice in single-concept settings. As such, synthesis problems represent another type of tool for physics instructors, similar to other classes of physics problems like context-rich problems and jeopardy problems.

Third, there is evidence that self-explanation and analogical comparison can be combined for additional gains. Although it is not completely clear whether those gains are solely due to increased time on task or represent an additional inherent difference between the treatments, this finding is of pedagogical value. The addition of  $< 10$  min of total time on task necessary for students to briefly

self-explain the individual worked examples before comparison resulted in significant improvements in total score on the target synthesis problem.

There are several important limitations to the studies presented here. First, the three experiments studied synthesis problems involving only two sets of physics concepts. Further study with an increased variety of concepts and combinations is necessary. As a corollary, the problems studied here share a subtle but important commonality: they are both structured so that students can arrive at an answer, albeit incorrect, without considering at least one of the component concepts. In experiment 1, students can forgo (and frequently did) the circular motion constraint and subsequent application of centripetal acceleration if they naively assumed that the velocity of the cart was zero at the top of the loop. In experiments 2 and 3, students could arrive at an answer by only considering the battery voltage. As such, students were not blocked from the final answer only because of a missing unknown, but rather conceptual consideration of the physical situation. It is possible that the large gains in concept recognition reflect this inability of students to successfully rely only on an equation-hunting heuristic. Mathematically sequential synthesis problems that require students to use one concept to first solve for an unknown necessary for the application of the second concept may not represent as significant a challenge for students (cf. Ref. [19]).

A second limitation of this series of experiments is that they do not account for potential interactions with feedback during training. In particular, the effectiveness of analogical comparison might increase when students are provided immediate feedback on their comparisons, either through peer-mediated feedback in a group work setting or via individualized tutoring in computer-based instruction. In a similar vein, these experiments did not explicitly manipulate the presentation of the worked example prompts and solutions. It is possible that the inclusion of expertlike explanations or highlighted presentation of the worked examples may help reduce cognitive load, and subsequently improve the effectiveness of analogical comparison with more demanding worked examples. However, even without such additional support, the fact that these interventions proved successful in supporting student performance suggests that such methods may be productive as part of individual assignments to scaffold student problem solving, particularly in large classes with limitations for one-on-one interaction with an instructor.

A final key limitation is that our current study focused on interventions to support student problem solving on a target problem that shared the same physical concepts and general solution structure as the provided worked examples. Previous research on student categorization tasks with introductory-level single-concept physics problems has shown that students are much more likely to rely on such surface features in their classification of problem solving

approaches compared to the principle-based methods often employed by experts [52–55]. As such, one potential critique is that this study is merely documenting trivial, surface-feature transfer. That is, successful students may not have created a deep underlying problem representation of the target problem, but instead superficially mapped the worked examples to the target synthesis problem based only on the similarity of surface features (i.e., in experiment 1, all problems contained hills and circular loops).

While this study was not directly designed to investigate this possibility, there are a number of reasons to suspect that student performance improvements were driven by more than surface feature matching and direct replication of the work flow in the provided solutions. First, while we wanted to limit the confound of far transfer difficulties between the worked example and the target task, the experimental design included efforts to ensure that the target task was not too similar to the worked examples. For example, there are significant superficial differences between the diagrams and physical context for the worked examples and the target tasks (see Figs 2 and 5), and the order of the work flow within the worked-example solutions is varied and different from the typical solutions for the target tasks. Second, if success on the target problem was primarily modulated by exposure to the same collective bag of physics concepts, basic physics equations, and other, problem-specific surface features (for example, circular tracks and changes in height), one might have expected the single-concept analogical comparisons to also result in significant improvements, since the solutions to the single concept problems included the same concepts and basic physics equations, just in different combinations. However, experiments 1 and 2 both found that comparisons of single-concept worked examples did not significantly improve student performance relative to control. Third, for experiment 2, only the self-explanation and most scaffolded analogical comparison condition significantly improved student concept recognition and execution on the target synthesis problem—providing synthesis worked examples and solutions was not enough to increase performance without additional scaffolding or an opportunity for students to elaborate on the structure of the worked examples. In other words, on this more complicated problem, only the two conditions that most emphasized the underlying structure of the worked examples were successful. To further support this third point, we discussed in Sec. III. C how many students in the self-explanation condition produced summaries that were richer than simple reproductions of solutions, and that there are two proposed theoretical explanations for this, including mental-model revision.

At its core, the key instructional issue is that transfer with single-concept problems is qualitatively different from transfer with synthesis problems. In this study, we find that synthesis brings a dimension of complexity to a transfer task that must be addressed. In particular, unlike single-concept problems where the goal is often to get students to use

problem cues to identify a particular principle for use (rather than match an equation to the corresponding variable set of the problem), synthesis problems have multiple, potentially competing conceptual cues within a single-problem statement. As such, the interaction between these conceptual cues and the identification and joint application of the concepts, can be relatively more complicated. This difficulty is likely exacerbated by the fact that students are likely biased to consider only a single concept given their previous exposure to more traditional physics problems.

With these issues in mind, it remains an open question to what extent the improvements in student performance observed here for these relatively isomorphic training and target examples will transfer to problems with less similar surface features or altogether different combinations of physics concepts. However, there are reasons to be optimistic on both accounts. First, we note that the target problem in the second and third experiments—in part as a result of its complexity—arguably differed more from its worked example counterparts than the target problem used in the first experiment. Whereas the first experiment consisted of the same physical entities (hills and circular loops) repeated in the worked examples and target problem, the second and third experiments included more varied physical situations representing different contexts for Faraday’s law (a circuit falling out of a field, a stationary circuit in a changing field, and a moving rod). Although more explicit scaffolding was necessary, students were still able to successfully compare and transfer the training to the target problem. This may suggest that transfer to yet even more dissimilar problems and examples may be possible with appropriate scaffolding. As such, future work to explore the effectiveness of these techniques as modulated by transfer distance would be interesting and useful.

The second reason for optimism is that there is evidence that a primary benefit of these interventions for synthesis problems is that they did in fact shift a significant number of students from applying only a single concept (or concept subset) to correctly applying the full set of required physics concepts. Although training on a single problem involving a particular pair of physics concepts may be unlikely to transfer directly to another involving a completely different combination of concepts, it is possible that repeated exposure to synthesis problems may represent a useful way to help students to recognize deep problem structure, consider conceptual cues, and de-emphasize the use of equation-hunting heuristics typically associated with single-concept end-of-chapter problems. As such, synthesis worked examples—combined with either analogical comparisons or self-explanations—may represent a useful tool to help students develop complex problem solving skills.

#### ACKNOWLEDGMENTS

This work is supported by the National Science Foundation (Grant No. DRL—1252399).

## APPENDIX A: WORKED EXAMPLES AND PROMPTS FOR EXPERIMENT 1

### 1. Synthesis worked examples

**Problem 1.** A block with mass  $M = 2.0$  kg slides from a horizontal surface into a vertical circular track with a radius of  $R = 3.0$  m. Assume that friction between the block and the track is negligible. What is the minimum speed the block must have at the bottom of the loop that will permit it to slide all the way around the circular track without leaving the track at the top?



#### Student solution 1

At the top of the loop:



$$F_{\text{Net}} = F_g + F_N = ma_c \quad (\text{A1})$$

$$F_{\text{Net}} = mg + 0 = \frac{mv^2}{R} \quad (\text{A2})$$

$$g = \frac{v^2}{R} \quad (\text{A3})$$

$$v = \sqrt{gR} \quad (\text{A4})$$

$$\frac{1}{2}mv_b^2 = \frac{1}{2}mv^2 + mgh \quad (\text{A5})$$

$$\frac{1}{2}mv_b^2 = \frac{1}{2}m(gR) + mg(2R) \quad (\text{A6})$$

$$v_b^2 = gR + 4gR \quad (\text{A7})$$

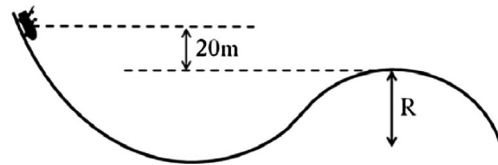
$$v_b = \sqrt{5gR} = \sqrt{5\left(9.81 \frac{\text{m}}{\text{s}^2}\right)(3.0 \text{ m})} \quad (\text{A8})$$

$$\boxed{v_b = 12.1 \text{ m/s}} \quad (\text{A9})$$

**Problem 2.** You are tasked with designing a rollercoaster. As part of the design, the track gradually descends until it comes to a small semicircular hill of height  $R$ .

You know that the speed of the rollercoaster cart will be  $18.5$  m/s when the cart is  $20$  m above the height of the

oncoming hill. What is the minimum possible hill height for which the cart does not leave the surface of the track at the top? Ignore friction.



#### Student solution 2

$$\frac{1}{2}mv_i^2 + mgh = \frac{1}{2}mv_f^2 \quad (\text{A10})$$

$$v_f^2 = v_i^2 + 2gh \quad (\text{A11})$$

$$v_f = \sqrt{(18.5 \text{ m/s})^2 + 2(9.81 \text{ m/s}^2)(20 \text{ m})} \quad (\text{A12})$$

$$v_f = 27.1 \text{ m/s} \quad (\text{A13})$$

$$\text{At the top of the hill:} \quad (\text{A14})$$



$$F_{\text{Net}} = F_N - F_g = -ma_c \quad (\text{A15})$$

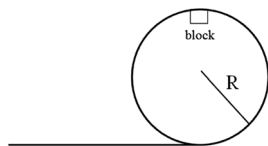
$$F_{\text{Net}} = 0 - mg = -\frac{mv_f^2}{R} \quad (\text{A16})$$

$$R = \frac{v_f^2}{g} = \frac{(27.1 \text{ m/s})^2}{9.81 \text{ m/s}^2} \quad (\text{A17})$$

$$\boxed{R = 75 \text{ m}} \quad (\text{A18})$$

### 2. Single-concept worked examples

**Problem 1.** A block with mass  $M = 2.0$  kg slides around a vertical circular track with a radius of  $R = 3.0$  m. Assume that friction between the block and the track is negligible. What is the minimum speed the block must have at the top of the track in order to ensure that it does not leave the track at the top?



$$R = \frac{v_f^2}{g} \tag{A31}$$

$$R = \frac{(27 \frac{\text{m}}{\text{s}})^2}{9.81 \frac{\text{m}}{\text{s}^2}} \tag{A32}$$

$$\boxed{R = 74 \text{ m}} \tag{A33}$$

**Student solution 1**

At the top of the loop:



$$F_{\text{Net}} = F_g + F_N = ma_c \tag{A22}$$

$$F_{\text{Net}} = mg + 0 = \frac{mv^2}{R} \tag{A23}$$

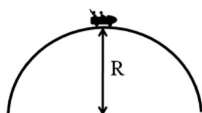
$$g = \frac{v^2}{R} \tag{A24}$$

$$v = \sqrt{gR} \tag{A25}$$

$$v = \sqrt{(9.8 \text{ m/s}^2)(3.0 \text{ m})} \tag{A26}$$

$$\boxed{v_b = 5.4 \text{ m/s}} \tag{A27}$$

**Problem 2.** You are tasked with designing a rollercoaster. As part of the design, the track includes a semi-circular hill of height  $R$ .



You know that the speed of the rollercoaster cart will be 27 m/s at the top of the hill. What is the minimum possible hill height for which the cart does not leave the surface of the track at the top? Ignore friction.

**Student solution 2**

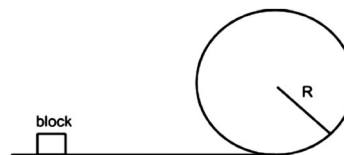
At the top of the hill:



$$F_{\text{Net}} = F_N - F_g = -ma_c \tag{A29}$$

$$F_{\text{Net}} = 0 - mg = -\frac{mv_f^2}{R} \tag{A30}$$

**Problem 1.** A block with mass  $M = 1.5 \text{ kg}$  slides from a horizontal surface into a circular track with a radius of  $R = 2.0 \text{ m}$ . Assume that friction between the block and the track is negligible. If the speed of the block at the top of the track is 3.0 m/s, what was the speed of the block at the bottom before entering the loop?



**Student solution 1**

$$\frac{1}{2}mv_b^2 = \frac{1}{2}mv_i^2 + mgh \tag{A34}$$

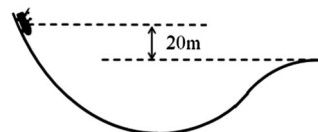
$$\frac{1}{2}mv_b^2 = \frac{1}{2}mv_i^2 + mg(2R) \tag{A35}$$

$$v_b^2 = v_i^2 + 4gR \tag{A36}$$

$$v_b = \sqrt{\left(3.0 \frac{\text{m}}{\text{s}}\right)^2 + 4(9.81 \text{ m/s}^2)(2.0 \text{ m})} \tag{A37}$$

$$\boxed{v_b = 9.4 \text{ m/s}} \tag{A38}$$

**Problem 2.** You are tasked with designing a rollercoaster. As part of the design, the track gradually descends until it comes to a small semicircular hill of height  $R$ .



You know that the speed of the rollercoaster cart will be 15 m/s when the cart is 20 m above the height of the oncoming hill. What is the velocity at the top of the oncoming hill?

**Student solution 2**

$$\frac{1}{2}mv_i^2 + mgh = \frac{1}{2}mv_f^2 \tag{A39}$$

$$v_f^2 = v_i^2 + 2gh \tag{A40}$$

$$v_f = \sqrt{(15 \text{ m/s})^2 + 2(9.81 \text{ m/s}^2)(20 \text{ m})} \quad (\text{A41})$$

$$\boxed{v_f = 24.8 \text{ m/s}} \quad (\text{A42})$$

### 3. Mastery prompts (synthesis version)

Each of the following questions refers to the two different problem-solution pairs on page 1.

- (A) Consider the diagram in line 1 of solution 1 and line 6 of solution 2. Explain any similarities and differences between the identified forces in terms of the two physical situations.
- (B) Consider line 3 of solution 1 and line 8 of solution 2. Is the direction of the  $\frac{mv^2}{R}$  term the same or different in the two solutions? Explain your reasoning.
- (C) In Problem 1 line 3, the student substitutes in zero for  $F_N$ . Explain why.

Is the situation in problem 2 similar or different to problem 1? Explain your reasoning.

(D) Consider line 6 in solution 1 and line 1 in solution 2. Identify the sources of energy considered in the two solutions.

Why is the term  $mgh$  on the left side of solution 2, but on the right side in solution 1?

Where did the student in solution 1 assign  $y = 0$ ? Explain your reasoning.

- (E) A friend in an introductory physics class at another institution asks for help in understanding and tackling similar problems. What will you tell your friend so that s/he can fully understand the problem?

### 4. Recognition prompts (synthesis version)

Each of the following questions refers to the two different problem-solution pairs on page 1.

- (A) What are the main physical concept(s) used in both of the students' solutions?
- (B) For problem 1, identify the elements of the problem statement or diagram which indicate the need to use each of the physical concept(s) you mentioned in Part A. List specific elements and explain your reasoning.
- (C) For problem 2, identify the elements of the problem statement or diagram which indicate the need to use each of the physical concept(s) you mentioned in Part A. List specific elements and explain your reasoning.
- (D) Are the elements you identified above similar between the two problems? Explain your answer highlighting any differences and similarities.
- (E) Another student compared these two problems and solutions, and remarked:

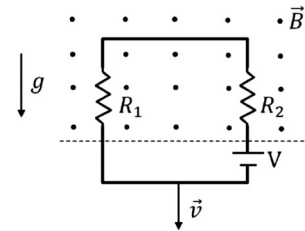
"I get why the initial and final velocity in problem 2 are both nonzero, but I don't understand why the final velocity in problem 1 must be nonzero."

- Using the physical concept(s) you identified in part A, what would you say to help this student?
  - What elements of problem 1 indicate that the final velocity should be nonzero?
- (F) A friend in an introductory physics class at another institution asks for help in understanding and tackling similar problems. What will you tell your friend so that s/he can fully understand the problem?

## APPENDIX B: WORKED EXAMPLES AND PROMPTS FOR EXPERIMENT 2

### 1. Synthesis worked examples

**Problem I.** A square circuit of length  $L = 40 \text{ cm}$  and unknown mass  $m$  is falling (gravity directed as shown) through a region with a uniform and perpendicular magnetic field  $B = 2.3 \text{ T}$  directed out of the page. The resistors have resistances of  $R_1 = 10 \Omega$  and  $R_2 = 25 \Omega$ , and the voltage of the battery is  $9 \text{ V}$ . At the instant shown, when part of the circuit is still in the region with a magnetic field, the velocity of the object is constant and equal to  $10.0 \text{ m/s}$  downward. Find the mass of the circuit.



### Student solution I

$$\begin{array}{c} \uparrow F_B \\ \square \\ \downarrow F_g \end{array} \quad (\text{B1})$$

$$F_{\text{Net}} = F_B - F_g = ma \quad (\text{B2})$$

$$ILB - mg = 0 \quad (\text{B3})$$

$$m = \frac{ILB}{g} \quad (\text{B4})$$

$$|\varepsilon| = \frac{d\Phi_B}{dt} \quad (\text{B5})$$

$$|\varepsilon| = \frac{d(BA)}{dt} = \frac{d(BLy)}{dt} = BL \frac{dy}{dt} = BLv \quad (\text{B6})$$

$$\varepsilon = BLv \text{ counterclockwise} \quad (\text{B7})$$

$$\sum V = I_{\text{total}} R_{\text{total}} \tag{B8}$$

$$V + \varepsilon = I(R_1 + R_2) \tag{B9}$$

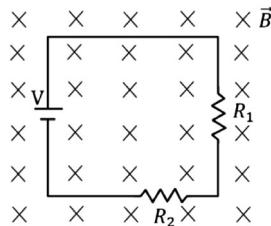
$$I = \frac{V + \varepsilon}{R_1 + R_2} \tag{B10}$$

$$m = \frac{LB(V + BLv)}{g(R_1 + R_2)} \tag{B11}$$

$$m = \frac{(0.4 \text{ m})(2.3 \text{ T})[9 \text{ V} + (2.3 \text{ T})(0.4 \text{ m})(10.0 \frac{\text{m}}{\text{s}})]}{9.81 \frac{\text{m}}{\text{s}^2}(35 \Omega)} \tag{B12}$$

$$m = 0.049 \text{ kg} \tag{B13}$$

**Problem II.** A square circuit of length  $L = 3.0 \text{ m}$  is in a region with a changing magnetic field as shown. The magnetic field is directed into the page, and varies with time as  $B(t) = B_o t$ , where  $B_o = 0.125 \text{ T/s}$ . The battery voltage is  $5 \text{ V}$ ,  $R_1 = 25 \Omega$  and  $R_2 = 75 \Omega$ . What is the magnitude and direction of the force on the top wire at time  $t = 30.0 \text{ s}$ ?



**Student solution II**

$$|\varepsilon| = \frac{d\Phi_B}{dt} \tag{B14}$$

$$|\varepsilon| = \frac{d(BA)}{dt} = L^2 \frac{dB}{dt} = L^2 B_o \tag{B15}$$

$$\varepsilon = L^2 B_o \text{ counterclockwise} \tag{B16}$$

$$\sum V = I_{\text{total}} R_{\text{total}} \tag{B17}$$

$$V - \varepsilon = I(R_1 + R_2) \tag{B18}$$

$$I = \frac{V - \varepsilon}{R_1 + R_2} \tag{B19}$$

$$\text{At } t = 30.0: B(t) = B_o t \rightarrow B = 3.75 \text{ T} \tag{B20}$$

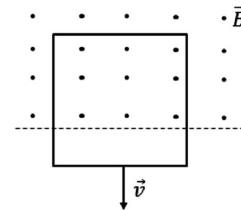
$$|F| = ILB = \frac{(V - L^2 B_o)LB}{R_1 + R_2} \tag{B21}$$

$$|F| = \frac{[5 \text{ V} - (3.0 \text{ m})^2(0.125 \frac{\text{T}}{\text{s}})](3.0 \text{ m})(3.75 \text{ T})}{100 \Omega} \tag{B22}$$

$$F = 0.44 \text{ N} \text{ towards the top of the page} \tag{B23}$$

**2. Single-concept worked examples**

**Problem I.** A square loop of conducting wire, length  $40 \text{ cm}$  is being pulled at a constant speed  $10.0 \text{ m/s}$  through a region with a uniform magnetic field  $B = 2.3 \text{ T}$  directed out of the page as shown. At the instant shown, when part of the circuit is still in the region with a magnetic field, find the direction and magnitude of the induced emf in the loop.



**Student solution I**

$$|\varepsilon| = \frac{d\Phi_B}{dt} \tag{B24}$$

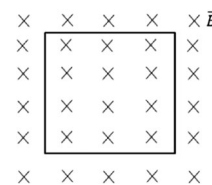
$$|\varepsilon| = \frac{d(BA)}{dt} = \frac{d(BLy)}{dt} = BL \frac{dy}{dt} \tag{B25}$$

$$|\varepsilon| = BLv \tag{B26}$$

$$|\varepsilon| = (2.3 \text{ T})(0.4 \text{ m}) \left(10.0 \frac{\text{m}}{\text{s}}\right) \tag{B27}$$

$$\varepsilon = 9.2 \text{ V counterclockwise} \tag{B28}$$

**Problem II.** A square loop of conducting wire, length  $3.00 \text{ m}$  is in a region with a changing magnetic field as shown. The magnetic field is into the page and varies with time as  $B(t) = B_o t$ , where  $B_o = 0.125 \text{ T/s}$ . What is the magnitude and direction of the induced emf in the loop?



**Student solution II**

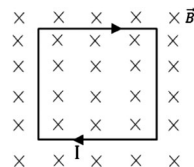
$$|\varepsilon| = \frac{d\Phi_B}{dt} \tag{B29}$$

$$|\varepsilon| = \frac{d(BA)}{dt} = L^2 \frac{dB}{dt} \tag{B30}$$

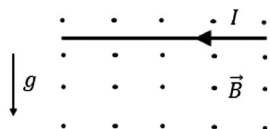
$$|\varepsilon| = L^2 B_0 \tag{B31}$$

$$|\varepsilon| = (3.00 \text{ m})^2 \left(0.125 \frac{\text{T}}{\text{s}}\right) \tag{B32}$$

$$\varepsilon = 1.13 \text{ V counterclockwise} \tag{B33}$$



**Problem I.** A current carrying wire of length  $L = 40 \text{ cm}$  is falling (gravity directed as shown) at a constant velocity through a region with a uniform and perpendicular magnetic field  $B = 2.0 \text{ T}$  directed out of the page. If the current is  $0.5 \text{ A}$ , find the mass of the wire.



**Student solution I**

$$\tag{B34}$$

$$F_{\text{Net}} = F_B - F_g = ma \tag{B35}$$

$$ILB - mg = 0 \tag{B36}$$

$$m = \frac{ILB}{g} \tag{B37}$$

$$m = \frac{(0.5 \text{ A})(0.4 \text{ m})(2.0 \text{ T})}{9.81 \text{ m/s}^2} = \boxed{0.041 \text{ kg}} \tag{B38}$$

**Problem II.** A square loop of conducting wire, length  $L = 3.0 \text{ m}$  is in a region with a uniform magnetic field  $B = 1.2 \text{ T}$  as shown. If there is a current  $I = 0.25 \text{ A}$  in the loop, what is the magnitude and direction of the force on the top wire?

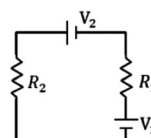
**Student solution II**

$$F = ILB \tag{B39}$$

$$F = (0.25 \text{ A})(3.0 \text{ m})(1.2 \text{ T}) \tag{B40}$$

$$F = \boxed{0.9 \text{ N}} \text{ towards the top of the page} \tag{B41}$$

**Problem I.** Given  $V_1 = 9 \text{ V}$ ,  $V_2 = 5 \text{ V}$ ,  $R_1 = 10 \Omega$  and  $R_2 = 15 \Omega$ , find the magnitude of the current in the circuit shown.



**Student solution I**

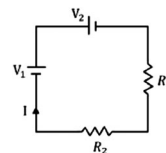
$$\sum V = I_{\text{total}} R_{\text{total}} \tag{B42}$$

$$V_1 + V_2 = I(R_1 + R_2) \tag{B43}$$

$$I = \frac{V_1 + V_2}{R_1 + R_2} \tag{B44}$$

$$I = \frac{14}{35} \text{ A} = \boxed{0.4 \text{ A}} \tag{B45}$$

**Problem II.** There is a clockwise current  $I = 0.04 \text{ A}$  in the circuit shown. Given  $V_1 = 5.0 \text{ V}$ ,  $R_1 = 25 \Omega$  and  $R_2 = 75 \Omega$ , find the voltage of the unknown battery  $V_2$ .



**Student solution II**

$$\sum V = I_{\text{total}} R_{\text{total}} \tag{B46}$$

$$V_1 - V_2 = I(R_1 + R_2) \tag{B47}$$

$$V_2 = V_1 - I(R_1 + R_2) \tag{B48}$$

$$V_2 = 5 \text{ V} - (0.04 \text{ A})(100 \Omega) \tag{B49}$$

$$\boxed{V_2 = 1.0 \text{ V}} \tag{B50}$$

### 3. Mastery prompts (synthesis version)

Each of the following questions refers to the two different problem-solution pairs on page 1.

- (A) What physical quantity is calculated in line 6 in solution I and line 2 in solution II?
- (B) Explain any similarities and differences between line 6 in solution I and line 2 in solution II. *Hint: As part of your answer, explain the  $\frac{dy}{dt}$  term in problem I and the  $\frac{dB}{dt}$  term in problem II in light of the two physical situations.*
- (C) Consider line 7 in solution I and line 3 in solution II. Explain why both have a counterclockwise direction, even though the direction of the magnetic field is different in problem I and problem II.
- (D) Consider line 9 in solution I and line 5 in solution II. Explain why solution I substitutes in  $V + \epsilon$  while solution II substitutes in  $V - \epsilon$ .
- (E) Why is the force due to the magnetic field on the loop in problem I directed towards the top of the page (solution I line 1)? Explain your reasoning.
- (F) The magnetic field in Problem II is in the opposite direction of problem I. Why is the magnetic force on the top wire in problem II also directed towards the top of the page (solution II line 9)? Explain your reasoning.
- (G) A friend in an introductory physics class at another institution asks for help in understanding and tackling similar problems. What are the most important ideas or information that you can tell your friend so that s/he can solve a similar problem?

### 4. Recognition prompts

Each of the following questions refers to the two different problem-solution pairs on page 1.

- (A) What are the main physical concept(s) used in both of the students' solutions?
- (B) For **problem I**, identify the elements of the problem statement and/or diagram which indicate the need to use each of the physical concept(s) you mentioned in Part A. List the elements and explain your reasoning for each of them.
- (C) For **problem II**, identify the elements of the problem statement and/or diagram which indicate the need to use each of the physical concept(s) you mentioned in Part A. List the elements and explain your reasoning for each of them.
- (D) Are the elements you identified in part B and C similar or different between the two problems? Explain your answer highlighting any similarities and/or differences.
- (E) In problem I line 9, the student substitutes in for the total  $emf$  for both circuits. Explain why the student

includes more than just the battery voltage, using the physical concepts you identified in Part A. In what ways is the situation in problem II line 5 similar or different from problem I line 9? Explain your reasoning.

- (F) A friend in an introductory physics class at another institution asks for help in understanding and tackling similar problems. What are the most important ideas or information that you can tell your friend so that s/he can solve a similar problem?

### 5. Combined prompts

Each of the following questions refers to the two different problem-solution pairs on page 1.

- (A) What are the main physical concept(s) used in both of the students' solutions?
- (B) For **problem I**, identify the elements of the problem statement and/or diagram which indicate the need to use each of the physical concept(s) you mentioned in Part A. List the elements and explain your reasoning for each of them.
- (C) For **problem II**, identify the elements of the problem statement and/or diagram which indicate the need to use each of the physical concept(s) you mentioned in Part A. List the elements and explain your reasoning for each of them.
- (D) One of the important concepts in both problems is that of an induced  $emf$  due to a changing magnetic flux. Compare the physical reason for a changing magnetic flux in each problem. Explain your answer highlighting any differences between the two problems.
- (E) A changing magnetic flux induces an  $emf$ . Explain any similarities and differences between line 6 in solution I and line 2 in solution II. *Hint: As part of your answer, explain the  $\frac{dy}{dt}$  term in problem I and the  $\frac{dB}{dt}$  term in problem II in light of the two physical situations and your answers to part D.*
- (F) Consider line 7 in solution I and line 3 in solution II. Explain why both induced  $emfs$  have a counterclockwise direction, even though the direction of the magnetic field is different in problem I and problem II.
- (G) The battery is not the only source driving the current in both problems. Consider line 9 in solution I and line 5 in solution II. Explain why solution I substitutes in  $V + \epsilon$  while solution II substitutes in  $V - \epsilon$ .
- (H) A friend in an introductory physics class at another institution asks for help in understanding and tackling similar problems. What will you tell your friend so that s/he can fully understand the problem?



## APPENDIX C: WORKED EXAMPLES AND PROMPTS FOR EXPERIMENT 3

## 1. Synthesis worked examples and annotation

<b>C.1: Synthesis Worked Examples + Annotation</b>	
<p><b>Student Solution II</b></p> $ \varepsilon  = \frac{d\Phi_B}{dt}$ $ \varepsilon  = \frac{d(BA)}{dt} = L^2 \frac{dB}{dt} = L^2 B_0$ $\varepsilon = L^2 B_0 \text{ counter-clockwise}$ $\Sigma V = I_{total} R_{total}$ $V - \varepsilon = I(R_1 + R_2)$ $I = \frac{V - \varepsilon}{R_1 + R_2}$ <p>At <math>t = 30.0\text{s}</math>:</p> $B = B_0 t = (0.125 \text{ T/s})(30.0\text{s}) = 3.75 \text{ T}$ $ F  = ILB = \frac{(V - L^2 B_0)LB}{R_1 + R_2}$ $ F  = \frac{\left(5V - (3.0\text{m})^2 \left(0.125 \frac{\text{T}}{\text{s}}\right)\right) (3.0 \text{ m})(3.75 \text{ T})}{100 \Omega}$ <div style="border: 1px solid black; display: inline-block; padding: 2px;"><math>F = 0.44 \text{ N}</math></div> <i>towards the top of the page</i>	<div style="display: flex; flex-direction: column; align-items: center;"> <div style="margin-bottom: 10px;">(1) {</div> <div style="margin-bottom: 10px;">(2) {</div> <div style="margin-bottom: 10px;">(3) {</div> <div style="margin-bottom: 10px;">(4) {</div> <div style="margin-bottom: 10px;">(5) {</div> <div style="margin-bottom: 10px;">(6) {</div> <div style="margin-bottom: 10px;">(7) {</div> <div style="margin-bottom: 10px;">(8) {</div> <div style="margin-bottom: 10px;">(9) {</div> <div style="margin-bottom: 10px;">(10) {</div> </div> <div style="border: 1px solid black; padding: 5px; margin-bottom: 10px;"> <p><u>The concept(s) used in this section are:</u></p>   <p><u>The goal of this section is:</u></p> </div> <div style="border: 1px solid black; padding: 5px; margin-bottom: 10px;"> <p><u>The concept(s) used in this section are:</u></p>   <p><u>The goal of this section is:</u></p> </div> <div style="border: 1px solid black; padding: 5px;"> <p><u>The concept(s) used in this section are:</u></p>   <p><u>The goal of this section is:</u></p> </div>

## 2. Combined prompts

- (A) What are the main physical concept(s) used in both of the students' solutions?
- (B) Compare the main solution steps used to solve both problems. Explain in what ways they are similar and/or different.
- (C) One of the important concepts in both problems is that of an induced *emf* due to a changing magnetic flux. Compare the physical reason for a changing magnetic flux in each problem. Explain your answer highlighting any differences between the two problems.

- (D) A changing magnetic flux induces an *emf*. Explain any similarities and differences between line 6 in Solution I and line 2 in Solution II. *Hint: As part of your answer, explain the  $\frac{dy}{dt}$  term in problem I and the  $\frac{dB}{dt}$  term in problem II in light of the two physical situations and your answers to part C.*
- (E) Consider line 7 in solution I and line 3 in solution II. Explain why both induced *emfs* have a counter-clockwise direction, even though the direction of the magnetic field is different in problem I and problem II. *Note: It is not enough to only state the name of a rule or physical law; you must explain and compare the application of that rule or law in the two cases.*
- (F) The battery is not the only source driving the current in both problems. Consider line 9 in solution I and line 5 in solution II. Explain why solution I substitutes in  $V + \varepsilon$  while solution II substitutes in  $V - \varepsilon$ .
- (G) Consider line 1 in solution I and line 10 in solution II. Explain why the force of the magnetic field on both current carrying wires is directed towards the top of the page in both problems, even though the direction of the magnetic field is different in problem I and problem II. *Note: It is not enough to only state the name of a rule or physical law; you must explain and compare the application of that rule or law in the two cases.*

- 
- [1] J. L. Docktor and J. P. Mestre, Synthesis of discipline-based education research in physics, *Phys. Rev. ST Phys. Educ. Res.* **10**, 020119 (2014).
- [2] L. Hsu, E. Brewster, T. M. Foster, and K. A. Harper, Resource letter RPS-1: Research in problem solving, *Am. J. Phys.* **72**, 1147 (2004).
- [3] D. P. Maloney, An overview of physics education research on problem solving, in *Getting Started in PER*, edited by C. Henderson and K. A. Harper (American Association of Physics Teachers, College Park, MD, 2011), Reviews in PER Vol. 2.
- [4] J. H. Larkin, Processing information for effective problem solving, *Eng. Educ.* **70**, 285 (1979).
- [5] J. H. Larkin and F. Reif, Understanding and teaching problem-solving in physics, *Int. J. Sci. Educ.* **1**, 191 (1979).
- [6] F. Reif and J. Heller, Knowledge structure and problem solving in physics, *Educ. Psychol.* **17**, 102 (1982).
- [7] A. Van Heuvelen and D. P. Maloney, Playing physics jeopardy, *Am. J. Phys.* **67**, 252 (1999).
- [8] P. T. Hardiman, R. Dufresne, and J. P. Mestre, The relation between problem categorization and problem solving among experts and novices, *Memory Cogn.* **17**, 627 (1989).
- [9] P. B. Kohl and N. D. Finkelstein, Patterns of multiple representation use by experts and novices during physics problem solving, *Phys. Rev. ST Phys. Educ. Res.* **4**, 010111 (2008).
- [10] D. E. Meltzer, Relation between students' problem-solving performance and representational format, *Am. J. Phys.* **73**, 463 (2005).
- [11] L. Ding, N. W. Reay, A. Lee, and L. Bao, Exploring the role of conceptual scaffolding in solving synthesis problems, *Phys. Rev. ST Phys. Educ. Res.* **7**, 020109 (2011).
- [12] P. Heller, R. Keith, and S. Anderson, Teaching problem solving through cooperative grouping. Part 1: Group versus individual problem solving, *Am. J. Phys.* **60**, 627 (1992).
- [13] P. Heller and M. Hollabaugh, Teaching problem solving through cooperative grouping. Part 2: Designing problems and structuring groups, *Am. J. Phys.* **60**, 637 (1992).
- [14] C. A. Ogilvie, Changes in students' problem-solving strategies in a course that includes context-rich, multifaceted problems, *Phys. Rev. ST Phys. Educ. Res.* **5**, 020102 (2009).
- [15] P. D. Antonenko, C. A. Ogilvie, D. S. Niederhauser, J. Jackman, P. Kumsaikaew, R. R. Marathe, and S. M. Ryan, Understanding student pathways in context-rich problems, *Educ. Inf. Technol.* **16**, 323 (2011).
- [16] Q. X. Ryan, E. Frodermann, K. Heller, L. Hsu, and A. Mason, Computer problem-solving coaches for introductory physics: Design, and usability studies, *Phys. Rev. Phys. Educ. Res.* **12**, 010105 (2016).
- [17] L. Ding, N. W. Reay, A. Heckler, and L. Bao, Sustained effects of solving conceptually scaffolded synthesis problems, *AIP Conf. Proc.* **1289**, 133 (2010).
- [18] L. Ding, N. W. Reay, A. Lee, and L. Bao, Using conceptual scaffolding to foster effective problem solving, *AIP Conf. Proc.* **1179**, 129 (2009).
- [19] B. Ibrahim, L. Ding, A. F. Heckler, D. R. White, and R. Badeau, Students' conceptual performance on synthesis physics problems with varying mathematical complexity, *Phys. Rev. Phys. Educ. Res.* **13**, 010133 (2017).
- [20] D. R. White, R. Badeau, A. F. Heckler, and L. Ding, Bottlenecks in solving synthesis problems, in *Proceedings of the Physics Education Research Conference, Minneapolis, MN, 2014*, edited by P. V. Engelhardt, A. D. Churukian, and D. L. Jones (AIP, New York, 2015).
- [21] D. E. Allen, B. J. Duch, and S. E. Groh, The power of problem-based learning in teaching introductory science courses, *New Directions for Teaching and Learning* **1996**, 43 (1996).
- [22] B. J. Duch, Writing problems for deeper understanding, in *The Power of Problem-Based Learning: A Practical "How To" for Teaching Undergraduate Courses in any Discipline*, edited by B. J. Duch, S. E. Groh, and D. E. Allen (Stylus Publishing Sterling, VA, 2001), p. 47.
- [23] M. S. Sabella and E. F. Redish, Knowledge organization and activation in physics problem solving, *Am. J. Phys.* **75**, 1017 (2007).

- [24] J. Sweller and G. A. Cooper, The use of worked examples as a substitute for problem solving in learning algebra, *Cognit. Instr.* **2**, 59 (1985).
- [25] X. Zhu and H. A. Simon, Learning mathematics from examples and by doing, *Cognit. Instr.* **4**, 137 (1987).
- [26] R. K. Atkinson, S. J. Derry, A. Renkl, and D. Wortham, Learning from examples: Instructional principles from the worked examples research, *Rev. Educ. Res.* **70**, 181 (2000).
- [27] J. Sweller, Cognitive load during problem-solving: Effects on learning, *Cogn. Sci.* **12**, 257 (1988).
- [28] A. Renkl and R. K. Atkinson, An example order for cognitive skill acquisition, in *In Order to Learn: How the Sequence of Topics Influences Learning*, edited by F. E. Ritter, J. Nerb, E. Lehtinen, and T. M. O'Shea (Oxford University Press New York, 2007), pp. 95–105.
- [29] M. T. H. Chi, M. Bassok, and M. Lewis, Self-explanations: How students study and use examples in learning to solve problems, *Cogn. Sci.* **13**, 145 (1989).
- [30] M. T. H. Chi, N. De Leeuw, M.-H. Chiu, and C. Lavancher, Eliciting Self-Explanations Improves Understanding, *Cogn. Sci.* **18**, 439 (1994).
- [31] M. T. H. Chi and K. VanLehn, The content of physics self-explanations, *J. Learn. Sci.* **1**, 69 (1991).
- [32] V. Alevan and K. Koedinger, An effective metacognitive strategy: Learning by doing and explaining with a computer-based Cognitive Tutor, *Cogn. Sci.* **26**, 147 (2002).
- [33] K. VanLehn, R. Jones, and M. T. H. Chi, A model of the self-explanation effect, *J. Learn. Sci.* **2**, 1 (1992).
- [34] R. Catrambone and K. J. Holyoak, Overcoming contextual limitations on problem-solving transfer, *J. Exp. Psychol. Learn. Mem. Cognit.* **15**, 1147 (1989).
- [35] D. Gentner, Structure mapping: A theoretical framework for analogy, *Cogn. Sci.* **7**, 155 (1983).
- [36] D. Gentner, The mechanisms of analogical learning, *Similarity and Analogical Reasoning*, edited by S. Vosniadou and A. Ortony (Cambridge University Press London, 1989), pp. 199–241.
- [37] D. Gentner and A. B. Markman, Structure mapping in analogy and similarity, *Am. Psychol.* **52**, 45 (1997).
- [38] D. Gentner, J. Loewenstein, and L. Thompson, Learning and transfer: A general role for analogical encoding, *J. Educ. Psychol.* **95**, 393 (2003).
- [39] R. W. Ferguson and K. D. Forbus, Telling juxtapositions: Using repetition and align-able difference in diagram understanding, in *Advances in Analogy Research*, edited by K. Holyoak, D. Gentner, and B. Kokinov (New Bulgarian University Sofia, Bulgaria, 1998), pp. 109–117.
- [40] D. Gentner and D. R. Gentner, Flowing waters or teeming crowds: Mental models of electricity, in *Mental Models*, edited by D. Gentner and A. L. Stevens (Lawrence Erlbaum Hillsdale, NJ, 1983), pp. 99–129.
- [41] D. E. Brown and J. Clement, Overcoming misconceptions via analogical reasoning: abstract transfer versus explanatory model construction, *Instr. Sci.* **18**, 237 (1989).
- [42] J. Clement, Using bridging analogies and anchoring intuitions to deal with students' preconceptions in physics, *J. Res. Sci. Teach.* **30**, 1241 (1993).
- [43] L. Ding, N. W. Reay, A. Lee, and L. Bao, Exploring the role of conceptual scaffolding in solving synthesis problems, *Phys. Rev. ST Phys. Educ. Res.* **7**, 020109 (2011).
- [44] N. S. Podolefsky and N. D. Finkelstein, Use of analogy in learning physics: The role of representations, *Phys. Rev. ST Phys. Educ. Res.* **2**, 020101 (2006).
- [45] N. S. Podolefsky and N. D. Finkelstein, Analogical scaffolding and the learning of abstract ideas in physics: An example from electromagnetic waves, *Phys. Rev. ST Phys. Educ. Res.* **3**, 010109 (2007).
- [46] S.-Y. Lin and C. Singh, Using isomorphic problems to learn introductory physics, *Phys. Rev. ST Phys. Educ. Res.* **7**, 020104 (2011).
- [47] D. P. Maloney, O'T. L. Kuma, C. J. Hieggelke, and A. Van Heuvelen, Surveying students' conceptual knowledge of electricity and magnetism, *Am. J. Phys.* **69**, S12 (2001).
- [48] K. Zuza, J.-M. Almuđí, A. Leniz, and J. Guisasaola, Addressing students' difficulties with Faraday's law: A guided problem solving approach, *Phys. Rev. ST Phys. Educ. Res.* **10**, 010122 (2014).
- [49] M. Chi, Self-explaining expository texts: The dual processes of generating inferences and repairing mental models, *Adv. Instr. Psychol.* **5**, 161 (2000).
- [50] A. F. Heckler and E. C. Sayre, What happens between pre- and post-tests: Multiple measurements of student understanding during an introductory physics course, *Am. J. Phys.* **78**, 768 (2010).
- [51] L. Ding, N. W. Reay, A. Lee, and L. Bao, Effects of testing conditions on conceptual survey results, *Phys. Rev. ST Phys. Educ. Res.* **4**, 010112 (2008).
- [52] M. T. Chi, P. J. Feltovich, and R. Glaser, Categorization and representation of physics problems by experts and novices, *Cogn. Sci.* **5**, 121 (1981).
- [53] J. L. Docktor, J. P. Mestre, and B. H. Ross, Impact of a short intervention on novices' categorization criteria, *Phys. Rev. ST Phys. Educ. Res.* **8**, 020102 (2012).
- [54] A. Mason and C. Singh, Assessing expertise in introductory physics using categorization task, *Phys. Rev. ST Phys. Educ. Res.* **7**, 020110 (2011).
- [55] E. R. Savelsbergh, T. de Jong, and M. G. Ferguson-Hessler, Choosing the right solution approach: The crucial role of situational knowledge in electricity and magnetism, *Phys. Rev. ST Phys. Educ. Res.* **7**, 010103 (2011).