International Journal of Science Education

ISSN: 0950-0693 (Print) 1464-5289 (Online) Journal homepage: http://www.tandfonline.com/loi/tsed20

When procedures discourage insight: epistemological consequences of prompting novice physics students to construct force diagrams

Eric Kuo, Nicole R. Hallinen & Luke D. Conlin

INTERNATIONAL JOURNAL OF SCIENCE EDUCATION, 2017
http://dx.doi.org/10.1080/09500693.2017.1308037

Routledge
Taylor & Francis Group

Check for updates

When procedures discourage insight: epistemological consequences of prompting novice physics students to construct force diagrams

Eric Kuo [a,b], Nicole R. Hallinen [b] and Luke D. Conlin [b]

[a]Department of Physics, Stanford University, Stanford, CA, USA; [b]Graduate School of Education, Stanford University, Stanford, CA, USA

ABSTRACT
One aim of school science instruction is to help students become adaptive problem solvers. Though successful at structuring novice problem solving, step-by-step problem-solving frameworks may also constrain students' thinking. This study utilises a paradigm established by Heckler [(2010). Some consequences of prompting novice physics students to construct force diagrams. *International Journal of Science Education*, *32*(14), 1829–1851] to test how cuing the first step in a standard framework affects undergraduate students' approaches and evaluation of solutions in physics problem solving. Specifically, prompting the construction of a standard diagram before problem solving increases the use of standard procedures, decreasing the use of a conceptual shortcut. Providing a diagram prompt also lowers students' ratings of informal approaches to similar problems. These results suggest that reminding students to follow typical problem-solving frameworks limits their views of what counts as good problem solving.

ARTICLE HISTORY
Received 10 October 2016
Accepted 14 March 2017

KEYWORDS
Problem solving; epistemology; physics education

One goal of school science is to help students develop adaptive problem-solving expertise so they can successfully solve a wide range of problems. Curricula provide students with practice problem sets with the aim of preparing students to solve not only these familiar problems but also new ones encountered in the future. With this goal in mind, science curricula are typically designed around general principles rather than idiosyncratic examples. For example, physics students learn Newton's second law, $\Sigma\vec{F} = m\vec{a}$: a general principle relating any object's acceleration to the net force acting on it.

Many science curricula teach problem solving through step-by-step frameworks to scaffold students' developing problem-solving expertise. For Newton's second law, typical problem-solving steps include drawing 'free-body diagrams' – standard representational tools for recording all the forces acting on an object – and applying the relation $\Sigma\vec{F} = m\vec{a}$ to solve for the desired quantities. The scaffolding provided by such frameworks is designed to be general enough to reliably lead students to correct solutions for a range of problems.

CONTACT Eric Kuo erickuo@pitt.edu 718 Learning Research & Development Center, 3939 O'Hara Street, Pittsburgh, PA 15260, USA

However, following these procedures without question could be detrimental, because adherence to scripted routines may discourage efficient alternatives. As an analogy, hikers who only ever follow an established trail through a forest may never get lost, but they may also never find a time-saving shortcut to the destination. The signposts marking the trail may discourage the use of shortcuts as strongly as they mark the familiar path. Similarly, the scaffolding of problem-solving frameworks may serve as signposts, both pointing toward familiar procedures and away from alternative approaches. Limiting problem solving to standard procedures may adversely affect students' epistemologies – their views about knowledge and learning – by restricting the kinds of approaches they view as appropriate for problem solving.

In the current study, we investigate the ways in which a prompt to follow a standard procedural step limits students' use of conceptual shortcuts in solving physics problems. This investigation replicates previous work by Heckler (2010), who showed that prompting diagram construction before solving a force problem increased the use of standard procedures. Extending this work, we propose that procedural prompts affect students' views of what kinds of solutions are appropriate. To study this question, we asked students to evaluate the appropriateness of an informal problem-solving approach.

We begin by reviewing the literature on problem-solving frameworks, discussing the benefits and risks of following procedures and considering the role that students' epistemological views play in problem solving.

Problem-solving procedures, shortcuts, and epistemologies

Instructional frameworks that structure problem solving

Common instructional supports for physics problem solving provide scaffolding through steps for novices to follow. Early research on problem solving in physics revealed that experts attend to the underlying conceptual structure of problems (Chi, Feltovich, & Glaser, 1981; Chi, Glaser, & Rees, 1982; Larkin, McDermott, Simon, & Simon, 1980). Subsequently designed step-by-step problem-solving frameworks pushed novice students into analysing physical concepts and selecting relevant physics principles before diving into mathematical manipulations (Heller & Heller, 2001; Reif & Heller, 1982; Ryan, Frodermann, Heller, Hsu, & Mason, 2016; Van Heuvelen, 1991). Such frameworks provide scaffolding to 'reduce complexity and choice by providing additional structure to the task' (Reiser, 2004, p. 283).

Although the details vary, most frameworks follow these general steps:

- Construct a diagram representing the problem.
- Define known and unknown quantities.
- Select relevant physics principles/equations.
- Map values in the problem to the equations.
- Compute the solution.
- Check the solution.

These frameworks have been shown to increase problem-solving success (Docktor, Strand, Mestre, & Ross, 2015; Heller, Keith, & Anderson, 1992; Reif & Heller, 1982; VanLehn et al., 2005), though primarily for problems similar to the training tasks.

The dangers of making thinking routine

Although procedures may structure problem solving and increase solution accuracy, following procedures may not be optimal for developing new insights. Schwartz, Bransford, and Sears (2005) distinguish routine expertise, skill at solving familiar problems efficiently, from adaptive expertise, skill at dealing with new problems in new contexts. They warn that learning and applying known procedures in familiar contexts may overshadow students' capacity for innovation on new problems and leave students unprepared for future adaptation.

One potential danger of procedures is that they can become so routine that more efficient alternatives are not considered. In a classic example, Luchins and Luchins (1959) showed how reuse of old routines can prevent discovery of new approaches. Their task presented three water jugs of different sizes with the goal of producing a certain volume of water. In the first block of problems, the solution method was always the same ('Fill the middle jug and then pour water out to fill the left jug once and the right jug twice'). Participants found the solution and practised it on several problems. In a subsequent block, the nature of the problems changed. The old solution still worked, but the problems could now also be solved with new, simpler solutions (e.g. 'Fill the left and right jugs and add them together'). Even though there were now simpler solutions, participants trained on the first block of problems tended to continue following their original approach.

Although both solution methods were effective on later problems, they differed in their efficiency. The old procedure, while effective, was unnecessarily complicated for the new problems. These students were apparently not searching for ways to streamline their solutions, instead opting to continue using established methods. Schwartz, Chase, and Bransford (2012) call this 'overzealous transfer,' where replication of old methods deters the development of new, more efficient ones.

The value of breaking from routines in physics problem solving

Even on standard textbook physics problems, overzealous adherence to standard problem-solving routines may overshadow valued alternative approaches. In one example, Hammer (1989) contrasted two students' approaches to this problem:

> One ball is thrown horizontally with a velocity v_0 from a height h, and another is thrown down with the same initial speed. Which ball will land first?

One student, Liza, wrote the relevant kinematic equation relating position, speed, and acceleration, $x = x_0 + v_0 t + 1/2\ at^2$, for the horizontal and vertical components of each ball. After some calculations, she concluded that the ball thrown downward would travel farther downward in a certain time t, so it would hit first. In contrast, Ellen explained that the answer was obvious: the ball thrown down will hit first because it has greater downward speed to cover the same vertical distance.

In another example, Kuo, Hull, Gupta, and Elby (2013) detailed two approaches to comparing the speeds of two falling objects. Both approaches correctly determined the difference in the speeds of the two objects, but in very different ways. One student's solution mirrored the steps of standard problem-solving frameworks, drawing a picture of the situation, selecting the relevant equation, $v = v_0 + at$, and performing calculations for the final speed of each object. The other student quickly determined that both objects experience the same change in speeds (represented in the equation as *at*), so the final difference in speeds is equal to the initial difference.

In these examples, procedural computations and qualitative approaches both led to the correct answer. However, approaches that broke from the standard procedures (which we label ‘shortcuts’) proved valuable in two respects. First, they were more efficient, leading to the correct answer more quickly and with less computational work. Second, the shortcuts demonstrated an understanding of the principles underlying these problems. Rather than relying on mathematical manipulations, the shortcuts drew on students’ sense of what result the calculations must yield.

Another consideration is how repeated problem-solving practice can crystallise students’ views of how to approach problem solving. Beyond the previous example, Hammer noted consistent differences between Liza’s and Ellen’s problem-solving approaches. For Liza, doing physics meant executing learned procedures. By contrast, Ellen sought to connect the procedures to common sense. Given these differing attitudes, it is perhaps not surprising that Liza gave a formal calculation, whereas Ellen took a qualitative shortcut. This is one example of how students’ epistemologies, views on knowledge and learning, can affect their physics problem solving.

Epistemological consequences of procedural instruction

As in Liza’s and Ellen’s cases, problem-solving approaches can be local instantiations of students’ broader epistemological views. As such, it is important to know how instruction affects students’ epistemologies toward problem solving. Surveying physics students’ attitudes toward learning shows a consistent result: typical classroom instruction pushes students to favour formal mathematics and procedures over developing conceptual understanding (Madsen, McKagan, & Sayre, 2015).

Outside of surveys, students may indicate their epistemological preference for formal procedures in classroom activities by rejecting correct, informal methods. Schoenfeld (1988) showed that students who had quickly created valid mathematical proofs spent extended periods of time rewriting their proofs in a standard two-column format. His conclusion was that the instructional emphasis on this conventional form caused students to value the formal structure of proofs over their underlying mathematical substance. In another example, Moore and Schwartz asked college students to evaluate the correctness of non-standard approaches to calculating variance. Students rejected these novel approaches, not because they were incorrect, but because they were not the sanctioned method (as reported in Schwartz & Martin, 2004). Students’ epistemologies may also influence what they consider appropriate explanations, as in the case of a student who incorrectly inserted formal vocabulary into an informal explanation, consistent with that student’s repeated rejections of informal, ‘common-sense’ ideas in science class (Lising & Elby, 2005).

These examples illustrate the role that epistemologies play in students’ valuation of different solutions. The key question is how these attitudes develop. What components of classroom instruction can push students to overvalue formalisms? In this study, we predict that prompts to use standard problem-solving frameworks strengthen views that informal approaches are inappropriate. Although such prompts successfully introduce structure into students’ problem-solving approaches, we predict that they also push students to devalue novel, efficient solutions.

Prompting diagram construction before problem solving

A procedure and a shortcut for solving a force problem

Figure 1 shows the two problems used in this study. How might someone solve Q1? As described earlier, we contrast two possible approaches: a typical procedure and a shortcut.

The standard procedural approach would be to analyse the net force on each box, then mathematically combine the results for the individual boxes to reach the answer. Figure 2 shows the steps of applying the procedure to solve this problem. First, individual free-body diagrams for the three boxes are drawn. These diagrams indicate all the forces acting on each separate box in the problem. The horizontal forces acting on each box are the pulls from the ropes and the force of friction opposing the motion. Newton's second law, $\Sigma F = ma$, can be applied to *each box separately*. As the boxes are moving with constant speed, the acceleration is zero, so the sum of the horizontal forces is zero. To solve for T_1, the tension in the rope between Liz's and Dan's boxes, Newton's second law should be applied to Dan's and Rex's boxes. These two implementations of Newton's second law combined with an expression for the frictional force yields the final answer: 120 N. This approach (drawing a free-body diagram, applying Newton's second law to each object, and combining the equations to solve for the unknown quantity) is a paradigmatic example of problem-solving procedures taught in many introductory physics courses.

This particular problem also allows for a shortcut solution, shown in Figure 3. The shortcut relies on a conceptual insight: the desired tension in the rope is due to pulling Dan and Rex's boxes together. In the shortcut approach, Dan's and Rex's boxes are treated as one combined unit, and Newton's second law is applied to *the combination of 'Dan+Rex.'* The resulting mathematical analysis of the tension in the rope is comparatively simpler and again leads to the correct answer.

Q1

Three siblings, Margaret, Dan, and Liz, are playing in the basement. With some rope, they attached three boxes together in a line like a train. Liz sits in the first box, Dan in the second, and they put the dog Rex in the third box. Margaret grabs on to the first box and pulls the "train" around the basement. When the kids (and the dog) are sitting in their box, each box has a total mass of 30 kg, and the coefficient of friction for the boxes on the basement floor is $\mu_k = 0.2$. At one point, Margaret is pulling horizontally and the "train" is moving with constant velocity $v = 2.0$ m/s on the level basement floor.

With how much force is the rope from Liz's box pulling on Dan's box? Show your work.

Q2

A red wagon and a blue wagon, both full of toys, are connected to each other by a rope. Nick connects another rope to the red wagon and uses his rope to pull the wagons up a 25° hill at constant speed. Nick pulls parallel to the angle of the hill. Each wagon has a mass of 10 kg. You can use the approximations sin(25°) ~ 0.4 and cos(25°) ~ 0.9 to simplify the calculations.

What is the difference in tension between the two ropes? Show your work.

Figure 1. The two problem-solving tasks used in this study: Q1 (the train problem) and Q2 (the ramp problem).

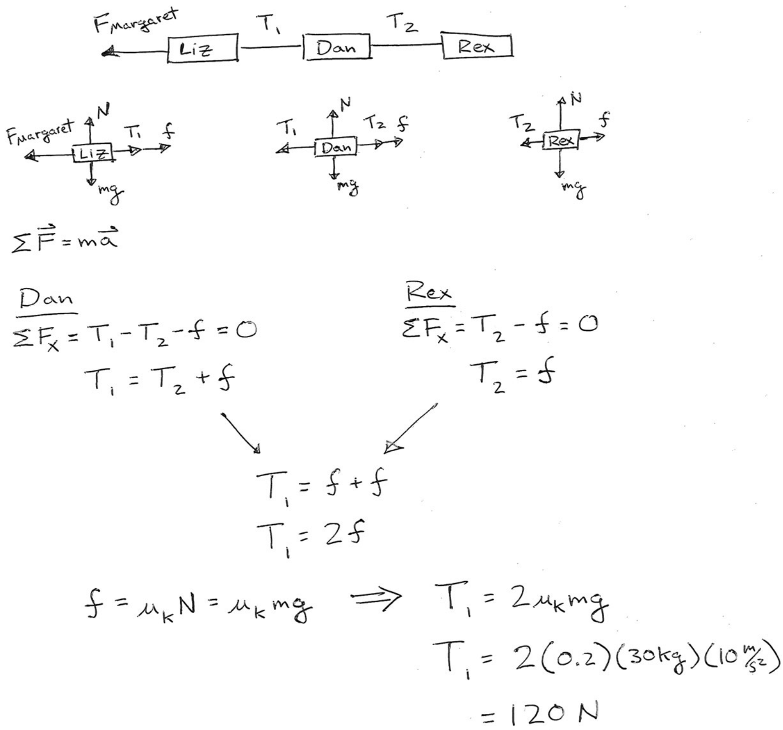

Figure 2. The prototypical procedure to solve the train problem.

Both the standard procedure and shortcut are valid and yield correct answers. They both rely on correctly identifying the forces and using Newton's second law. The key difference between the procedure and the shortcut is *when* the boxes are combined. In the standard procedure, the combination happens algebraically, after Newton's second law is applied to each box. In the shortcut, the boxes are combined conceptually, before

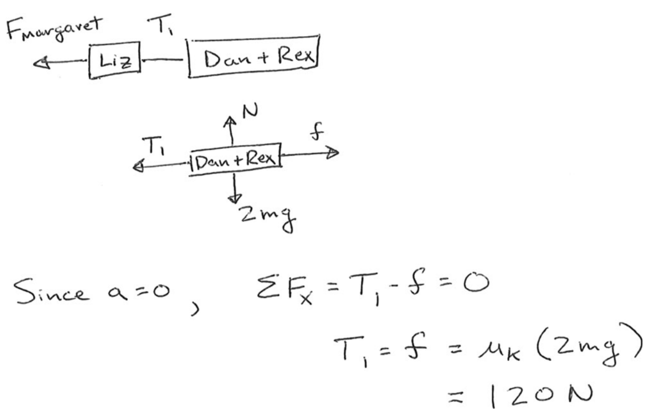

Figure 3. A shortcut to solve the train problem. Dan's and Rex's boxes are combined conceptually before use of Newton's second law.

equations are introduced. Compared to the standard procedure, the shortcut demonstrates two insights:

- Conceptual understanding – two objects travelling with the same velocity can be analysed as one combined unit.
- Efficiency – combining the boxes will simplify the calculations required.

While shortcuts are sometimes acknowledged in typical instruction, the standard procedure receives more emphasis. One reason is that the shortcut is not a *generally applicable* method. If asked to solve for another variable, such as T_2, then combining the boxes would be counterproductive. This highlights a strength of the standard procedure: it is general and can be used in *any* standard force problem to produce the correct answer. A potential weakness of the standard procedure is that it may discourage alternative approaches. Students using the shortcut approach demonstrate adaptivity by using a conceptual insight to break from standard procedures to find a more efficient solution.

The effect of prompting procedures

The potential for standard procedures to discourage adaptive problem solving highlights a way that typical problem-solving instruction can backfire. Instructors often scaffold students' problem solving by providing step-by-step frameworks, along with prompts for when to use them. One standard instructional scaffold for force problems is to tell students to 'draw a diagram representing all the forces' acting on the relevant objects. This hint is designed to help students structure their problem solving by generating free-body diagrams, leading to the next step in the standard procedure: applying Newton's second law to each object.

Heckler (2010) found that prompting students to draw free-body diagrams increased their use of formal problem-solving approaches and decreased their accuracy in solving problems. In that study, half of the students in an introductory physics course received a prompt telling them to draw a free-body diagram before solving force problems. The other half were simply asked to solve the same problems as a control group. The prompted group was more likely to draw formal free-body diagrams and to follow the standard procedure. In the current study, we primarily seek to replicate this result, comparing formal, procedural approaches to shortcut use.

A remaining question from the previous study is whether diagram prompts passively encouraged students to initiate standard procedures or actively discouraged pursuit of informal solutions. In extending this line, we test the prediction that a standard procedural prompt increases dissatisfaction with informal solutions. If our prediction is correct, it suggests that cues to execute standard problem-solving frameworks may trigger epistemological views that devalue informal approaches, including novel shortcuts.

As mentioned, Heckler's previous study also showed that the diagram prompts made students' problem solving less accurate. Because diagram prompts increased formal procedures and decreased correctness, our hypothesis is that formal procedures are less likely to be correct than more informal alternatives. However, the percentage of correct answers produced by each approach was not directly reported in Heckler (2010). In the current study, we seek to test this hypothesis directly.

Method

Materials

The primary tasks were split into the *problem-solving phase* and the *evaluation phase*. The problem-solving phase consisted of two force problems for students to solve themselves (Figure 1). Q1 was the train problem, described previously (taken from Heckler [2010]). Q2 was similarly designed to allow both procedural and shortcut approaches. Half of the participants were prompted to draw a particular free-body diagram before solving the problem. The problem-solving phase is a replication of Heckler's experimental design.

The evaluation phase presented an informal solution to another problem, shown in Figure 4. The solution was an example of an informal approach used by a student in Heckler's study. In this problem, a box pushed unequally in opposing directions is held in place by the force of friction. The informal solution violated the formal procedure by partitioning the analysis into two phases. First, the net force of the two pushes is found. Then, this net pushing force is set equal to friction to solve for the minimum mass required. In contrast, the procedural approach would use Newton's second law to set the sum of *all the forces* to zero in one step. This is the critical difference between the informal approach and the procedural approach. Participants were asked to rate the solution using a

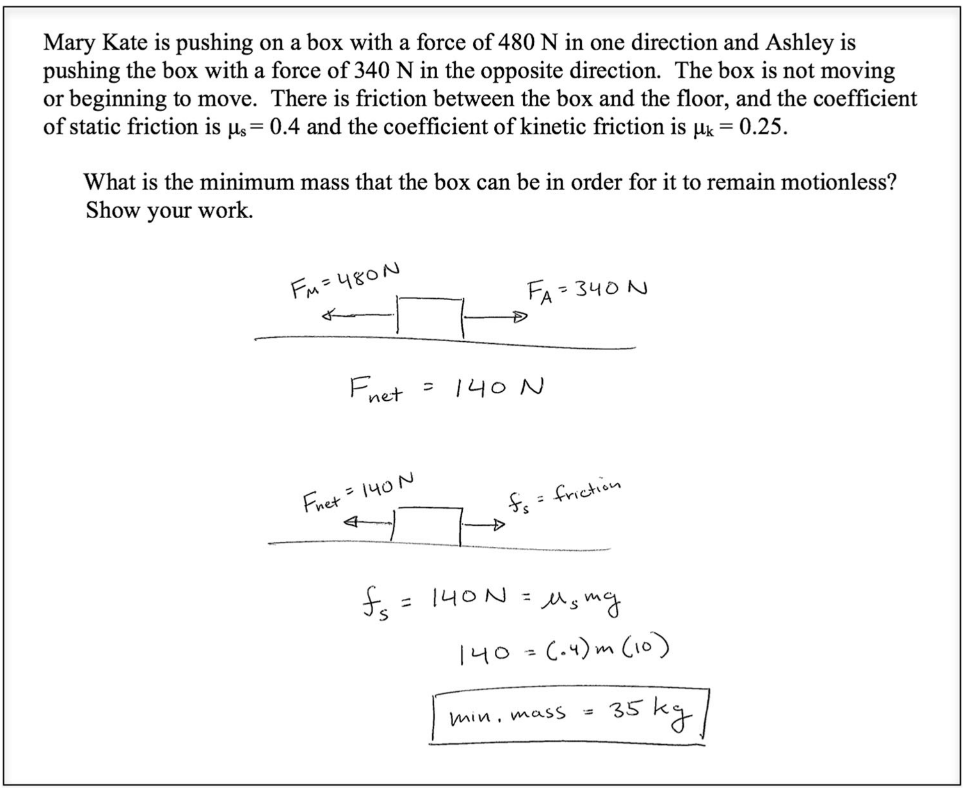

Figure 4. The informal solution students saw in the evaluation phase.

5-point scale, ranging from 'very bad' to 'very good,' and to explain their evaluation. Once again, half of the participants saw a prompt to draw a free-body diagram in the problem text.

The purpose of the evaluation phase was to see whether the formalising effects of prompting free-body diagrams in students' own problem solving would also apply to solution evaluation. We predicted that the diagram prompt would increase desire for procedural approaches and dissatisfaction with the informal solution. This would provide evidence that prompting diagrams not only affects students' own problem solving, but also causes them to devalue correct, non-standard methods. An alternative hypothesis is that the effect of prompting might fade when students are presented with a correct, informal solution. The rationale for this alternative is that a conceptually valid solution, no matter how informal, should be rated highly.

Notably, the problem-solving phase provides us with a baseline of how much students adhere to procedural routines. This will provide a comparison point for understanding the evaluation phase results. All things being equal, we predict that students who engage in procedural problem solving should find more fault with informal solutions. However, we also predict that the diagram prompt will disrupt this and increase the demand for formal procedures.

Participants

One hundred thirty-six undergraduate students from a large, private university participated in this study. The students, typically life sciences majors, were enrolled in an algebra-based introductory physics course. Most students were juniors and seniors.

The 10-week physics course covered mechanics, fluid dynamics, and thermodynamics. Though the course incorporated common large-lecture educational reforms (an electronic classroom voting system, pre-lecture reading assignments, small-group problem-solving sessions in recitation), the instructional problem-solving approach for force problems emphasised the standard procedure. Specifically, the instruction in this course emphasised these steps: draw a picture, select axes, draw a free-body diagram, identify constraints that allow conclusions about acceleration, and then write Newton's second law by horizontal and vertical components to solve for the desired unknown quantity. Instruction on forces and this problem-solving approach took place during two 50-minute lectures in the first two weeks of the course, with additional opportunities to practice on a homework assignment, during problem-solving practice in recitation, and on a midterm exam.

Design

All students first completed the problem-solving phase, then six learning attitude survey items, and finally the evaluation phase. Students were randomly assigned to two conditions, control ($n = 70$) and prompt ($n = 66$), to test the effects of diagram prompts. Students in the prompt condition saw prompts to draw force diagrams in the problem-solving phase and evaluation phase, shown in Table 1, whereas students in the control condition did not.

The six learning attitude survey items, provided in Appendix A, were drawn from the Colorado Learning and Attitudes Student Survey (Adams et al., 2006). The selected survey

Table 1. The diagram prompts included in the prompt condition.

Question	Prompt
Problem-solving: Q1	Draw a free-body diagram clearly indicating the forces on Dan's box.
Problem-solving: Q2	Draw a free-body diagram for each wagon, clearly indicating the forces.
Evaluation Task	Draw a free-body diagram clearly indicating the forces.

items related to problem solving and the importance of recall and memorisation in learning physics. This was our measure of students' problem-solving epistemologies in this physics class. Students responded using a 5-point Likert scale labelled from 'strongly disagree' to 'strongly agree.' Regardless of condition, all students saw the same items.

Procedure

This study was conducted in recitation sections during the last week of the 10-week academic quarter. A researcher attended each of the 12 recitation sections to distribute materials. Students in each recitation section were randomly assigned to condition. We obtained final course GPA for 127 students ($n_{\text{control}} = 64$, $n_{\text{prompt}} = 63$). There was no difference in course GPA by condition (Control: $M = 3.22$, SD = 0.60; Prompt: $M = 3.31$, SD = 0.54), $t(125) = 0.87$, $p = .39$, confirming random assignment.

The materials were distributed as one stapled packet and participants completed them using pencil and paper. Students worked individually and were encouraged to complete the problem-solving phase, attitudinal survey items, and evaluation phase in the order presented. Students had 20 minutes to complete all the tasks. They were permitted to use calculators, but did not have access to other resources, such as notes or textbooks.

Coding

Problem-solving codes

Three attributes of students' problem-solving solutions on Q1 and Q2 were coded: diagram type, problem-solving approach, and correctness. Diagram type was coded according to two main categories: (i) only standard free-body diagrams of *separate boxes* drawn or (ii) an insightful diagram of *combined boxes drawn*. A third category (*ambiguous/no diagram)* was used if the diagrams were incomplete or none were drawn. Approach was coded according to whether the solution employed the standard *procedure* or a *shortcut* that conceptually analysed a combined system. A third category captured *other* approaches. Correctness was coded according to whether the solution approach was valid and produced the correct answer (barring errors unrelated to the solution process, as in Heckler [2010]). Appendix B shows the detailed coding scheme and examples of student work.

Evaluation codes

We coded students' responses for whether justifications of their ratings cited violations of the formal procedure. Because we did not want 'formal complaints' to be directly attributable to the diagram prompt (i.e. students who see an explicit prompt to construct a free-body diagram complain more about incomplete or informal diagrams), we coded specifically for complaints about the mathematical approach rather than the diagram. Students who incorrectly stated that the sample solution was wrong were excluded from this coding.

A complaint was coded *formal approach* if it noted (i) that the net force should be zero and should include all of the forces or (ii) the absence of a general statement of Newton's second law ($\Sigma F = ma$). Example responses coded as *formal approach* complaints include:

> She used the right numbers, but $F_{net} = 0$ which tells you F_s must = 140 to offset the 480 N force.
>
> $F_{net} = 0$ since box doesn't move.
>
> Would have been nicer to relate to 2^{nd} law, shows complete thought process.
>
> I'd write an equation $\Sigma F_{net} = F_M - (F_A + F_{fr}) = 0$.

Interrater reliability

After an initial round of coding and discussion, the first and third authors independently applied the problem-solving codes (diagram type, approach, and correctness) to 20 responses of both Q1 and Q2, producing 92% agreement (average $\kappa = .85$, lowest code $\kappa = .76$). After discussing these codes and refining the coding scheme,[1] the first author coded all remaining responses.

The first and second authors independently applied the evaluation codes to 33 responses, grouping together all formal complaints, whether they related to either the diagram or approach. There was 91% agreement on this code ($\kappa = .81$).[2] In a second stage of coding, the two coders specified which formal complaints pertained specifically to the approach, agreeing 100%. The first author then coded all remaining responses.

Results

In this section, we begin by describing the main problem-solving phase result: prompting free-body diagrams increased procedural approaches and decreased the use of shortcut approaches. Then, we present the evaluation phase result: the diagram prompt increased complaints about the informal approach by students who did not follow the procedures in their own problem solving. Examining the problem-solving results in depth, we show the relations between prompting, diagram type drawn, approach taken, and correctness. Finally, we explore the ways in which students' problem solving aligned with their course GPA and problem-solving attitudes.

Main problem-solving result: the effect of prompting diagrams on approach

From the initial sample of 136 participants, we excluded blank responses or barely-started responses in which no approach was evident, because we were primarily interested in the effects of diagram prompts on solution attempts. For Q1, there were no blanks in the control condition and three blanks (5%) in the prompt condition. For Q2, there were two blanks (3%) in control and seven blanks (11%) in prompt. The difference in percentage of blank responses by condition was not significant for Q1 (Fisher's exact: $p = .11$) or Q2 (Fisher's exact: $p = .09$).

With the remaining sample, we explored the influence of diagram prompts on problem-solving approaches, predicting that prompting would increase the use of procedures relative to shortcuts. Figure 5 shows the distribution of problem-solving approaches by condition. The main comparison between the procedure and shortcut approaches confirmed our prediction on both problems, Q1: $\chi^2(1, N = 103) = 5.18$, $p = .02$; Q2: $\chi^2(1, N = 96) = 9.62$, $p = .002$. This finding matches Heckler's: when prompted to construct free-body diagrams, students are more likely to execute procedural routines. There was no condition difference for how many students took other approaches, as compared to those who used either a procedure or shortcut, Q1: $\chi^2(1, N = 133) = 0.84$, $p = .35$; Q2: $\chi^2(1, N = 127) = 0.44$, $p = .51$.

Evaluation results

After problem solving, 102 students completed the evaluation task ($n_{\text{control}} = 50$, $n_{\text{prompt}} = 52$). On a 5-point scale, the prompted group rated the informal solution lower (Control: $M = 3.94$, SD = 0.93; Prompt: $M = 3.54$, SD = 0.83), $t(100) = 2.30$, $p = .02$, $d = 0.45$. We take this difference as validation of our prediction: the diagram prompt increased dissatisfaction with the sample informal approach.

There are two plausible alternative interpretations. First, this result may be more indicative of students' prior problem solving. The prompted group's tendency to solve problems using procedures may carry over into their evaluation task. These students' dissatisfaction with the informal solution may reflect their preference for procedures in their own previous problem solving. Another alternative interpretation is that prompted students could specifically object to the quality of the diagrams drawn, since the prompt requests a canonical free-body diagram.

To investigate these alternative interpretations, we looked specifically at formal approach complaints, independent from those concerning the free-body diagram. We hypothesised that two factors would predict formal approach complaints. First, we expected students' evaluation to match their own problem solving: greater use of formal procedures in problem solving should predict more frequent formal approach complaints. To investigate this, we used students' *procedural approach score*, the

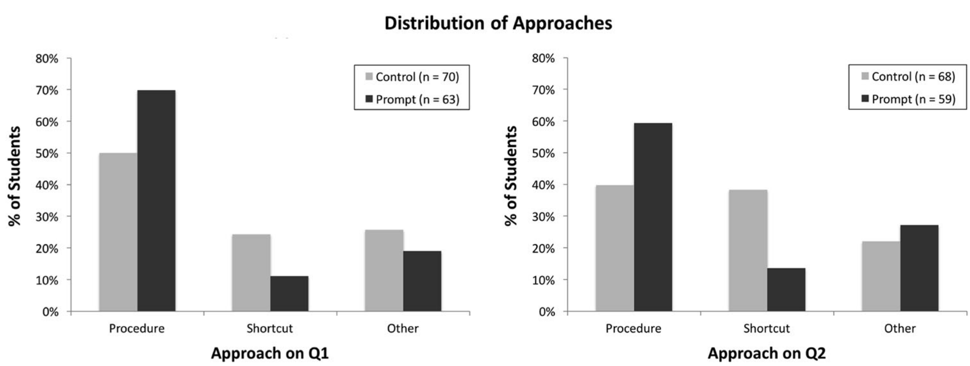

Figure 5. The percentage of students that used each approach in solving Q1 and Q2 in the problem-solving phase, broken out by condition.

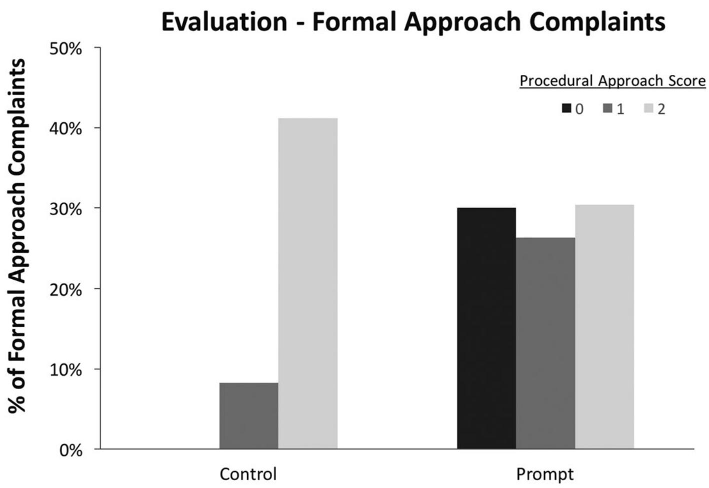

Figure 6. The percentage of students who made formal approach complaints in the evaluation phase, broken out by prompting and procedural approach score.

number of procedural approaches they used on Q1 and Q2, ranging from 0 to 2. Second, we expected the condition difference of the diagram prompt to trigger demands for formal approaches.

Figure 6 shows the percentage of formal approach complaints, split by condition and procedural approach score. The control group showed the expected effect of prior procedure use on formal approach complaints: students with a higher procedural approach score were more likely to demand formal approaches. That is, control students' prior problem-solving approaches matched their evaluation criteria. However, for the prompted group, the percentage of formal complaints did not increase with procedural approach score. The diagram prompt signalled to all students the need for formal approaches, independently of how they solved problems themselves. The condition differences were largest for students who did not use formal approaches in their own problem solving (procedural approach score = 0). For this group, 30% of prompted students demanded formal procedures when it came time to evaluate a new solution, even though they did not use such procedures in their own problem solving. In comparison, no control students with a procedural approach score of zero demanded formal procedures.

We performed a two-way logistic regression on formal approach complaints, using condition and procedural approach score as predictors. The full table of results is presented in Appendix C. The interaction between condition and procedural approach score was significant, $\beta = 2.35$, $p = .03$, highlighting the different relations between procedural approach score and formal approach complaints for the two conditions. The condition term was also significant, $\beta = -4.15$, $p = .03$, indicating the difference between the two conditions when procedural approach score is zero. Overall, prompting formal diagrams increased dissatisfaction with informal problem-solving approaches, especially for students who did not use procedures in their own problem solving.

How the prompts influence problem-solving approach

Given our main problem-solving finding that prompting affects students' solution approaches, we now delve deeper into other problem-solving features to investigate how the effects of prompting take hold.

Diagram types drawn

Figure 7 shows the distribution of diagram types by condition. Prompting increased the construction of separate boxes diagrams as compared to all other types on both problems, Q1: $\chi^2(2, N = 133) = 26.1$, $p < .001$; Q2: $\chi^2(2, N = 127) = 28.3$, $p < .001$. This serves as a manipulation check: averaging the two problems, students in the prompt condition followed the prompt to draw diagrams of separate boxes 93% of the time. In the control condition, students drew diagrams showing separate boxes only 52% of the time. While some control students drew a diagram of the combined boxes, 40% drew ambiguous diagrams or no diagram at all. When not explicitly prompted to draw a diagram, many students opted for incomplete or ill-specified diagrams.

Effect of diagram type on approach

Next, we investigated whether the effect of prompting on approach was mediated by the diagram type drawn. We performed a three-way log-linear analysis (using condition, diagram type, and approach as factors) to test for associations between the three factors simultaneously. In log-linear model selection analysis, the full model, including all one-way, two-way, and three-way associations, is pared down until removal of any association would lead to a significant change in model fit. The final model leaves only the significant associations between condition, diagram type, and approach. See Appendix D for full results of the log-linear analysis.

Figure 8 shows the significant associations between factors for Q1 and Q2. For both problems, prompting affected the type of diagram drawn, and diagram type was associated with approach. When diagram type is taken into account, condition did not directly impact approach.

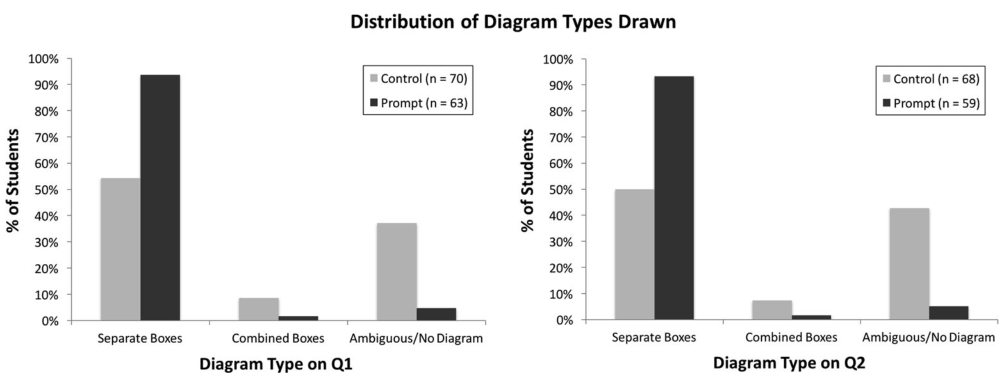

Figure 7. The percentage of students who drew each diagram type on Q1 and Q2 in the problem-solving phase, broken out by condition.

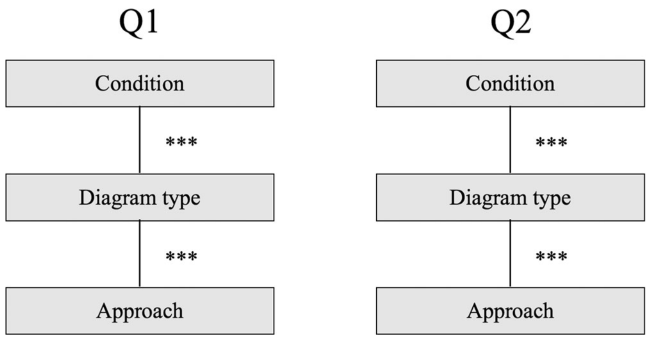

Figure 8. Associations between factors involved in the problem-solving phase. ***$p < .001$.

Table 2 shows the distribution of approaches used for each diagram type, averaged for Q1 and Q2. As expected, separate boxes diagrams supported a procedural approach, and combined boxes diagrams tended toward the shortcut approach. Prompting the construction of a standard free-body diagram increased separate box diagrams and more separate box diagrams led to increased use of procedural approaches. Perhaps unexpectedly, the suboptimal ambiguous/no diagram type did not correspond to suboptimal approaches. Rather, approaches that followed these diagrams were split between the procedure, shortcut, and other approaches. Many of the shortcut approaches were generated by students who drew ambiguous diagrams or no diagrams at all in the control condition.

No effect of prompting on correctness

Unlike Heckler's previous result, we found no condition differences in correctness on either question (Q1: Control = 37%, Prompt = 40%, $\chi^2(1, N = 133) = 0.09$, $p = .76$; Q2: Control = 57%, Prompt = 53%, $\chi^2(1, N = 127) = 0.30$, $p = .59$).

Diagram and solution approach affect correctness

To further investigate the effects of the different problem-solving factors (condition, diagram type, approach) on correctness, we extended the previous log-linear analysis to include correctness. Approaches coded as *other* were much worse than either the procedure or shortcut at reaching a correct final answer (0% on Q1, 16% on Q2). Many of these approaches started with incorrect principles (e.g. energy) or grossly misapplied Newton's laws. For this reason, we excluded solutions utilising other approaches from this analysis, focusing on differences between the two appropriate ways to solve this problem.

The overall associations with correctness were different for Q1 and Q2. For Q1, the correctness of the solution was associated with both diagram type ($p = .005$) and approach

Table 2. The average distribution of approaches following each diagram type drawn for Q1/Q2.

	Approach used (%)		
Diagram type	Procedure	Shortcut	Other
Separate boxes	68	11	21
Combined boxes	8	85	7
Ambiguous/no diagram	23	42	35

Table 3. Correctness broken out by diagram type and approach for Q1 and Q2 (excluding 'other' approaches). Fraction of solutions correct is shown in parentheses.

	% Correct solutions	
Problem-solving factor	Q1	Q2
Diagram type		
Separate boxes	54% (43/80)	69% (46/67)
Combined boxes	67% (4/6)	83% (5/6)
Ambiguous/no diagram	24% (4/17)	61% (14/23)
Approach		
Procedure	46% (36/79)	65% (40/62)
Shortcut	63% (15/24)	74% (25/34)

($p = .01$). For Q2, no factors were associated with correctness. Table 3 shows the correctness by diagram type and approach for Q1 and Q2. For Q1, diagram type was associated with correctness. This seems to be primarily due to the lower accuracy for students who drew ambiguous diagrams or did not draw diagrams. By contrast, on Q2, ambiguous/no diagrams were associated with about as many correct answers as the separate boxes diagram.

In line with our hypothesis, shortcut approaches were more correct than procedural approaches on both questions, though the difference (17%) was significant for Q1, but not for Q2 (9%). Looking at the errors made on the procedure and the shortcut for Q1, we found that a key difference was in the number of incomplete solutions (i.e. applying Newton's second law without reaching a final answer). Eighty-one percent of incorrect procedural approaches were incomplete. In these cases, the force analysis was not performed on enough of the individual boxes. By contrast, none of the incorrect shortcut approaches were incomplete in this way.[3] This result highlights a weakness specific to the procedural approach. When starting a procedural approach, students may recognise that they need to apply Newton's second law to objects in the problem, but not know which objects to analyse or how to combine the equations to find the desired quantity. For the shortcut, the choice to combine the relevant boxes reveals a goal-oriented insight, and this insight makes clear that Newton's second law should be applied to the combined unit.

The influences of diagram type and approach on correctness show the competing effects of prompting diagram construction. Prompted students almost all drew separate boxes diagrams, avoiding the risks of ambiguous/no diagrams. Yet, prompting also decreased the use of shortcut approaches, which were more likely to produce correct answers. In terms of producing correct answers, the benefits of prompting reliable methods were offset by the hazards of stifling insightful approaches.

Problem solving, course achievement, and attitudes

Next, we investigated whether problem-solving approaches related to a standard performance measure (course GPA) and to surveyed problem-solving attitudes. As with students' procedural approaches in the problem-solving phase, we generated scores for shortcut approach, other approach, and correct solutions on the problem-solving questions (ranging from 0 to 2). Table 4 shows the correlations of these scores with course grades and problem-solving attitudes. There were no significant condition differences with respect to these correlations.

Table 4. Correlations of problem-solving score with course GPA and problem-solving attitudes.

Problem-solving score	Correlation with course GPA ($n = 127$)	Sig.	Correlation with problem-solving attitudes ($n = 133$)	Sig.
Procedural approach score	**.44**	<.001	**.26**	.003
Shortcut approach score	−.05	.60	−.13	.12
Other approach score	**−.45**	<.001	−.13	.15
Correct score	**.55**	<.001	**.20**	.03

Course grade

As one would expect, correct solutions were positively correlated with course performance. Students who performed well on our problems performed well in the course overall. Although shortcut approaches led to more correct answers than procedural approaches, procedural approaches were significantly positively correlated with course grade while shortcut approaches were not. We interpret this as showing that course grades are closely aligned to the standard procedures typically emphasised in introductory physics instruction. Approaches coded as 'other' were negatively correlated with course grade, consistent with these approaches being mostly incorrect.

Problem-solving attitude survey questions

Attitudinal survey data was collected from 133 students ($n_{\text{control}} = 68$, $n_{\text{prompt}} = 65$). We coded these responses in accordance with standard scoring practices; responses judged most expert-like by physics instructors received a score of 5. These attitude survey questions served as our measure of students' epistemologies. There was no significant difference in the aggregate scores between the control ($M = 3.27$, SD = 0.49) and prompt conditions ($M = 3.29$, SD = 0.51), $t(131) = 0.27$, $p = .79$.

Problem-solving attitude score significantly correlated with procedural approach score and correct solutions. The correlation between attitudes and correctness mirrors findings that favourable scores on attitudinal surveys correlate with course performance in physics class (e.g. Finkelstein & Pollock, 2005). However, the positive correlation between procedural approach use and favourable attitudes was surprising.

Discussion

This study replicated a main finding from Heckler's original study on the effects of prompting standard diagrams before problem solving. Namely, diagram prompts increased the use of standard procedural approaches and decreased more informal methods. In extending Heckler's study, we found that one cause of this finding may be epistemological: these prompts affected students' stances toward what counts as valid problem solving. Specifically, prompting force diagrams pushed students to demand formal methods, even if they did not use such formalisms in their own problem solving. This new finding suggests that even a light-touch instructional scaffold can push students to overvalue formal methods and discourage them from generating shortcuts.

We caution against interpreting our results as indicating that diagrams and sketches are not a productive part of scientific thinking. The act of constructing diagrams supports conceptual understanding in science (e.g. Gobert & Clement, 1999), and diagrams can organise information spatially to facilitate problem solving (Larkin & Simon, 1987). Moreover, students in our study who drew any type of clear diagram were either as accurate or more accurate than students who did not draw clear diagrams. Simply constructing a diagram did not lead to negative results, but constructing a diagram *as part of a standard procedure* did. The question, then, is not whether diagrams are helpful or harmful. The question is how to leverage the benefits of drawing diagrams as tools for making sense of science concepts, without pressuring students to overvalue and follow procedural approaches. Future work can explore whether modified diagram prompts can maintain the benefits of scaffolds without being overly prescriptive.

The broader instructional hypothesis that follows from our findings is that procedural problem-solving scaffolds can strengthen epistemological views that value formalism over flexibility. By pointing students toward reliable procedures, we may be unintentionally suppressing their search for insights. The course sampled in our study did emphasise a procedure for force problems, and, consistent with the positive correlation between procedural problem solving and course grade, it is likely that course assessments valued use of those procedures. This course's portrayal of problem-solving expertise as procedural may also explain the positive correlation between procedural problem solving and problem-solving attitudes. We interpret this correlation as reflective of this particular course's values rather than that of expert problem-solving values more generally. To strengthen this interpretation, more work triangulating survey responses with problem-solving behaviour and instructional experiences is needed.

On prompting affecting correctness

We found no correctness difference by condition. These results differ from Heckler (2010), where prompting decreased the correctness of final answers. On the train problem common to both studies, the average performance of both conditions in our study was at the level of a prompted group in Heckler's study (40% correct). In comparison, Heckler's associated control group was 60% correct. This discrepancy may have resulted from the unusually fast pace of our physics course, which spends only two 50-minute lectures on force problems. It is possible that the unprompted students in our study did not have enough practice with force problems to answer more accurately than prompted students.

However, we did find that standard procedures were less likely to lead to correct answers than an informal approach that leveraged a conceptual shortcut. Most errors in the procedural approach were due to incomplete attempts; knowing to start the standard procedure does not guarantee knowing how to finish it. This suggests a potential mechanism for Heckler's results: prompting diagrams increases procedural approaches, which decreases correctness. While this did not hold for our sample, again, it is possible that the brief time and procedural focus of instruction prevented condition differences from carrying into solution correctness.

How can we teach routines without making thinking routine?

Our findings have uncovered an apparent disconnect between adaptive, expert-like insights (shortcuts) and behaviours valued in introductory course instruction (procedures). To address this disconnect, we return to our discussion of the original goals of problem-solving frameworks. Problem-solving frameworks are intended to help beginning students structure their problem solving as an initial step in the development of problem-solving expertise. Like most forms of direct instruction, the hope is for this explicit scaffolding to help students learn established methods and approaches (Kirschner, Sweller, & Clark, 2006). Our findings indicate that students in these instructional environments can even demonstrate small problem-solving insights, using efficient shortcuts that are distinct from standard frameworks.

However, our findings also indicate that standard instructional scaffolds to follow problem-solving frameworks also discourage shortcuts. Even at the end of a physics course, these scaffolds fix student thinking on scripted routines, rather than serving as stepping stones to new insights. Still, problem-solving frameworks could be the first step in a progression toward adaptive problem-solving expertise. As students gain more practice using standard problem-solving approaches, they could eventually develop shortcuts and adaptive strategies. However, in this instructional paradigm, a challenge for subsequent instruction is how to address students' views of problem solving that have been shaped by forcing early adherence to these frameworks.

Alternatively, instructional approaches could aim to mitigate the constraining effects of these frameworks from the beginning. Novel physics courses that successfully foster expert-like attitudes toward learning physics (e.g. Brewe, Traxler, de la Garza, & Kramer, 2013; Redish & Hammer, 2009) could provide a useful counterpoint to traditional instruction. Instead of implementing prescribed procedures, such courses typically aim to help students construct connections between various physics ideas and tools by engaging in rich modelling or reasoning tasks. Initial evidence suggests that such courses can help students break from standard procedures and implement conceptual shortcuts (Elby, Kuo, Gupta, & Hull, 2015). Future work could investigate ways that diagram prompts may be taken up by students in reform courses who have less experience with following procedural frameworks. We hypothesise that curricula that avoid emphasising procedures might mitigate the procedural influence of prompting diagrams. Conversely, the question for these curricula is whether the move away from procedures negatively impacts the reliability of students' solutions. Without procedural safety nets, can students reason through force problems correctly? It is crucial to consider the influence of the curriculum on both solution accuracy and problem-solving insight.

To that end, our findings indicate a need for changes in assessment. Commonly, problem-solving success in science class is measured according to whether the final solution is correct. In these cases, the approach taken does not matter, except for where errors might be made. Barring a few recent exceptions (e.g. Docktor et al., 2016), grading rubrics that do assess approaches explicitly value the execution of standard procedures, typically assigning points for completing each step correctly. This assessment trend was reflected in our results: course grade was correlated with procedural approaches on our problem-solving tasks. Moreover, these standard problem-solving assessments generally do not include any incentives for thinking outside of these step-by-step frameworks (Hull, Kuo, Gupta, & Elby, 2013). This was also reflected in our results by the lack of correlation between course grade and problem-solving shortcut use. By shifting from the typical focus of problem-solving assessments, we have shown that even standard physics problems can serve as litmus tests for problem-solving flexibility, revealing whether students are searching for and have the skills to find novel efficiencies. When evaluating the success of instructional methods, both immediate and long-term teaching goals deserve consideration. In the current study, attention to problem-solving shortcuts highlighted a potential long-term downside of prompting students to follow reliable procedures.

Conclusion

Science education addresses two primary goals: to teach basic skills and to prepare students to adapt those skills for future problems. In one proposed learning trajectory, students would initially focus on learning basic skills, eventually learning to use these skills more flexibly in the future. Flexible problem solving is pushed off as a long-term goal rather than one to be addressed initially. The risk of this approach is that it does not value the seeds of adaptive problem solving, as assessments value procedural methods over insight. Although the move to conceptually combine boxes on a force problem is a relatively small deviation from the standard procedure, fostering insights like these may help develop a student's eye for opportunities to act adaptively.

Introductory science courses are often viewed as gatekeepers to further STEM coursework and careers. But are the criteria for passing through this gate aligned with what we want students to learn? It is not a guarantee that students who demonstrate the greatest procedural fluency early on are most on-track for developing adaptive problem-solving expertise. By ignoring initial adaptive insights, we may be missing an opportunity to identify and support early forms of problem-solving expertise. If our goal is to develop students' adaptive problem-solving expertise in conjunction with procedural fluency, then as instructors and researchers we must find ways to recognise and cultivate both. Doing so will require us to look beyond our own standard instructional procedures, to discover new educational insights.

Notes

1. Five out of six diagram type disagreements had only one rater labelling the diagram ambiguous. In all of these instances, the authors resolved that these should be ambiguous, because the diagrams were not well labelled and could be interpreted in multiple ways. Stronger evidence of labelled diagrams would be required to not be coded as ambiguous. Seven out of nine approach disagreements were about whether an approach was ambiguous. The substance of the disagreement was between how articulate or complete the procedure or shortcut needed to be. The resolution was that approaches could be coded as procedure (or shortcut) if it was clear that students were expressing relationships between forces on one of the boxes (or on a system of two or more boxes). Six out of seven of these disagreements were resolved to be either procedure or shortcut. From the five correctness disagreements, the correctness was refined as follows: for Q2, correctly solving for the two tensions without taking the difference is sufficient, and the answer needs to be numerical or in terms of variables given in the problem text.
2. On the evaluation code, there were three disagreements that produced a clarification: simply requesting more work or clearer explanations should not be coded as a formal complaint.
3. The most common error (five out of nine errors) for the shortcut approach was to solve for the incorrect tension.

Acknowledgements

We thank the instructional staff of the physics course for working with us to conduct this research. The AAA Lab and Wieman Research Group provided thoughtful suggestions on the study design and analysis. Finally, thanks to Andy Elby, Cristina Zepeda, and two anonymous reviewers, who provided useful comments on a previous draft. Eric Kuo is currently at the Learning Research & Development Center, University of Pittsburgh. Nicole Hallinen is currently in the Departments of Psychology and Psychological Studies in Education, Temple University. Luke Conlin is currently in the Department of Chemistry and Physics, Salem State University.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Institute of Education Sciences [grant number R305B150014].

ORCID

Eric Kuo http://orcid.org/0000-0001-5292-6188

References

Adams, W. K., Perkins, K. K., Podolefsky, N. S., Dubson, M., Finkelstein, N. D., & Wieman, C. E. (2006). New instrument for measuring student beliefs about physics and learning physics: The Colorado learning attitudes about science survey. *Physical Review Special Topics – Physics Education Research*, *2*(1), 010101.

Brewe, E., Traxler, A., de la Garza, J., & Kramer, L. H. (2013). Extending positive CLASS results across multiple instructors and multiple classes of Modeling Instruction. *Physical Review Special Topics – Physics Education Research*, *9*(2), 020116.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5*(2), 121–152.

Chi, M. T. H., Glaser, R., & Rees, E. (1982). Expertise in problem solving. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 1, pp. 7–75). Hillsdale, NJ: Erlbaum.

Docktor, J. L., Dornfeld, J., Frodermann, E., Heller, K., Hsu, L., Jackson, K. A., … Yang, J. (2016). Assessing student written problem solutions: A problem-solving rubric with application to introductory physics. *Physical Review Physics Education Research*, *12*(1), 010130.

Docktor, J. L., Strand, N. E., Mestre, J. P., & Ross, B. H. (2015). Conceptual problem solving in high school physics. *Physical Review Special Topics – Physics Education Research*, *11*(2), 020106.

Elby, A., Kuo, E., Gupta, A., & Hull, M. M. (2015). *Tensions and trade-offs in instructional goals for physics courses aimed at engineers*. In 2015 American Society for Engineering Education Annual Conference & Exposition Proceedings, Seattle, WA.

Finkelstein, N. D., & Pollock, S. J. (2005). Replicating and understanding successful innovations: Implementing tutorials in introductory physics. *Physical Review Special Topics – Physics Education Research*, *1*(1), 010101.

Gobert, J. D., & Clement, J. J. (1999). Effects of student-generated diagrams versus student-generated summaries on conceptual understanding of causal and dynamic knowledge in plate tectonics. *Journal of Research in Science Teaching*, *36*(1), 39–53.

Hammer, D. (1989). Two approaches to learning physics. *The Physics Teacher*, *27*, 664–670.

Heckler, A. F. (2010). Some consequences of prompting novice physics students to construct force diagrams. *International Journal of Science Education*, *32*(14), 1829–1851.
Heller, P., & Heller, K. (2001). *Cooperative group problem solving in physics*. Brooks/Cole. Retrieved from http://www.physics.umn.edu/groups/physed/Research/CGPS/Green%20Book/cover_tc.pdf
Heller, P., Keith, R., & Anderson, S. (1992). Teaching problem solving through cooperative grouping. Part 1: Group versus individual problem solving. *American Journal of Physics*, *60*(7), 627–636.
Hull, M. M., Kuo, E., Gupta, A., & Elby, A. (2013). Problem-solving rubrics revisited: Attending to the blending of informal conceptual and formal mathematical reasoning. *Physical Review Special Topics – Physics Education Research*, *9*(1), 010105.
Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, *41*(2), 75–86.
Kuo, E., Hull, M. M., Gupta, A., & Elby, A. (2013). How students blend conceptual and formal mathematical reasoning in solving physics problems. *Science Education*, *97*(1), 32–57.
Larkin, J. H., McDermott, J., Simon, D. P., & Simon, H. A. (1980). Expert and novice performance in solving physics problems. *Science*, *208*(4450), 1335–1342.
Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, *11*(1), 65–100.
Lising, L., & Elby, A. (2005). The impact of epistemology on learning: A case study from introductory physics. *American Journal of Physics*, *73*(4), 372–382.
Luchins, A. S., & Luchins, E. H. (1959). *Rigidity of behavior: A variational approach to the effect of Einstellung*. Oxford: University of Oregon Press.
Madsen, A., McKagan, S. B., & Sayre, E. C. (2015). How physics instruction impacts students' beliefs about learning physics: A meta-analysis of 24 studies. *Physical Review Special Topics – Physics Education Research*, *11*(1), 010115.
Redish, E. F., & Hammer, D. (2009). Reinventing college physics for biologists: Explicating an epistemological curriculum. *American Journal of Physics*, *77*(7), 629–642.
Reif, F., & Heller, J. I. (1982). Knowledge structure and problem solving in physics. *Educational Psychologist*, *17*(2), 102–127.
Reiser, B. J. (2004). Scaffolding complex learning: The mechanisms of structuring and problematizing student work. *Journal of the Learning Sciences*, *13*(3), 273–304.
Ryan, Q. X., Frodermann, E., Heller, K., Hsu, L., & Mason, A. (2016). Computer problem-solving coaches for introductory physics: Design and usability studies. *Physical Review Physics Education Research*, *12*(1), 010105.
Schoenfeld, A. H. (1988). When good teaching leads to Bad results: The disasters of 'Well-taught' mathematics courses. *Educational Psychologist*, *23*(2), 145–166.
Schwartz, D. L., Bransford, J. D., & Sears, D. (2005). Efficiency and innovation in transfer. In J. P. Mestre (Ed.), *Transfer of learning from a modern multidisciplinary perspective* (pp. 1–51). Greenwich, CT: Information Age.
Schwartz, D. L., Chase, C. C., & Bransford, J. D. (2012). Resisting overzealous transfer: Coordinating previously successful routines with needs for new learning. *Educational Psychologist*, *47*(3), 204–214.
Schwartz, D. L., & Martin, T. (2004). Inventing to prepare for future learning: The hidden efficiency of encouraging original student production in statistics instruction. *Cognition and Instruction*, *22*(2), 129–184.
Van Heuvelen, A. (1991). Overview, Case Study Physics. *American Journal of Physics*, *59*(10), 898–907.
VanLehn, K., Lynch, C., Schulze, K., Shapiro, J. A., Shelby, R., Taylor, L., … Wintersgill, M. (2005). The Andes physics tutoring system: Lessons learned. *International Journal of Artificial Intelligence in Education*, *15*(3), 147–204.

Appendix A: Survey items on problem-solving attitudes

1. When I solve a physics problem, I locate an equation that uses the variables given in the problem and plug in the values.
2. There is usually only one correct approach to solving a physics problem.
3. If I get stuck on a physics problem on my first try, I usually try to figure out a different way that works.
4. Understanding physics basically means being able to recall something you have read or been shown.
5. If I do not remember a particular equation needed to solve a problem on an exam, there's nothing much I can do (legally!) to come up with it.
6. When studying physics, I relate the important information to what I already know rather than just memorising it the way it is presented.

Appendix B: Detailed problem-solving coding scheme

Here, we describe the problem-solving coding scheme, providing examples from student work.

Diagram type

Though many students drew sketches of the problem situation, only diagrams that included at least one arrow were considered free-body diagrams. Diagrams were coded as a *Separate boxes* diagram, *Combined boxes* diagram, or *Ambiguous/No Diagram*.

Separate boxes

Drawn diagrams indicate the forces for individual boxes or wagons. In Figure B1, sample student work shows how this type of diagram involves creating force diagrams for each box or wagon individually.

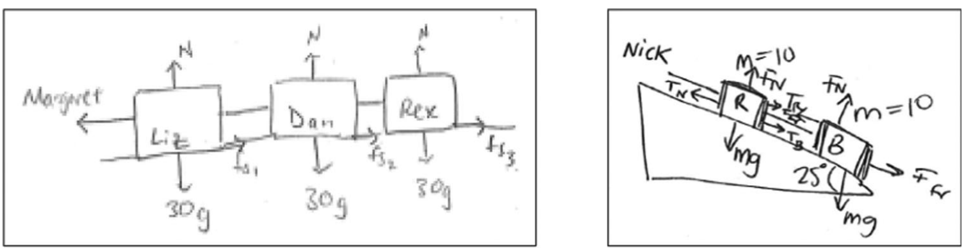

Figure B1. Diagrams coded as *Separate boxes* for Q1 (left) and Q2 (right). In both diagrams, the student indicates the forces on each individual unit in the problem.

Combined boxes

At least one diagram indicates the forces acting on some combination of multiple boxes/wagons. Sometimes, the system being combined was explicitly labelled in a diagram. Other times, this was implicit, signified by which forces were included in the diagram. On both problems, the *Combined boxes* code means the student drew at least one diagram involving combined boxes, even if they also drew more standard diagrams representing each box separately. Figure B2 shows *Combined boxes* diagrams.

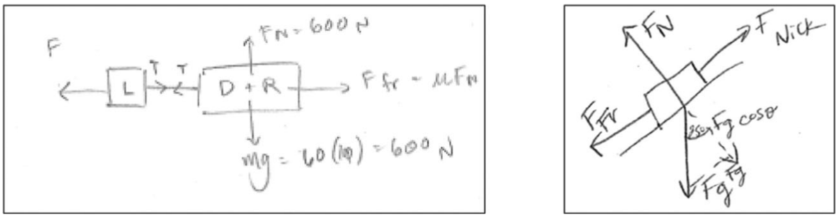

Figure B2. Sample student-generated diagrams that show multiple boxes combined as a single conceptual unit. The Q1 diagram (left) shows an example of an explicit label indicating a combined unit (D + R). The Q2 diagram (right) shows an example of the lack of a second box or an intermediate tension force implicitly indicating a combined unit.

Ambiguous/no diagram

Diagrams that could not be categorised into either of the previous two categories were coded as ambiguous. This includes diagrams that were not detailed enough to distinguish whether boxes were being combined or not. Figure B3 shows the examples for this code. In a few cases, there was no diagram of the boxes with forces indicated, receiving a *No diagram* code.

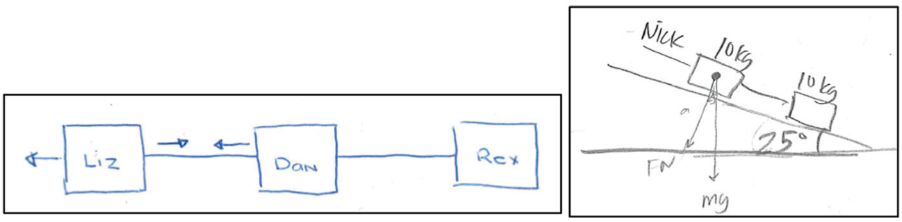

Figure B3. Sample student-generated diagrams coded as Ambiguous for Q1 (left) and Q2 (right). Here, although arrows indicating forces were drawn, not enough detail was provided to code these diagrams as clearly representing separate or combined boxes.

Problem-solving approach

Much like the diagrams, we coded students' solutions into three main approach categories – *procedural* approaches, evidence of a conceptual *shortcut*, and *other* approaches.

Procedure

This standard approach involves analysing the forces on each box separately, combining the resulting equations to solve for the relevant quantities. In either problem, solution approaches coded as 'procedure' provide no evidence that the students applied a force analysis to a combined system of two or more boxes as a way of getting to their answer (Figure 2 provides a prototypical example).

Shortcut

Approaches fit this code if there was evidence that students analysed the net force acting on a system of two or more boxes. Explicit markers for this include the omission of intermediate tension forces and the value used for the mass being that of two or more units (Figure 3 provides a prototypical example).

Other

Solution approaches were coded into a third category if it was impossible to distinguish whether they were using a procedural or shortcut approach. This could be because Newton's laws were applied incorrectly, such that it was not clear whether separate boxes or combined boxes were being analysed (e.g. incorrectly concluding that since $F_{net} = 0$, then $T = 0$). Solution approaches that did not obviously follow from Newton's laws (e.g. work/energy theorem, only calculating the force of friction) also fit into this code.

Correctness

All solutions were coded as either correct or incorrect. Our correctness code was quite generous. To be coded as correct, the approach would need to, in principle, yield the correct answer. A simple arithmetic or typographic error did not preclude a solution from being coded correct. Trigonometric errors (exchanging *sin* and *cos*) are ignored if it is clear from the diagram that the student is attempting to solve for the correct components (either parallel or perpendicular to the surface of the ramp). Also, if a friction term μN was incorrectly included on Q2, the solution was coded as correct if it would yield the correct answer in the limit $\mu \rightarrow 0$. The purpose of the correctness code was to address our question of whether the diagram prompt influenced the correctness of their approach to solving the problems, not whether it induced algebraic/typographical/trigonometric errors or whether students misunderstood the problem scenario.

Appendix C: Summary of logistic regression predicting formal approach complaints in the evaluation phase.

To analyse the evaluation phase results, we performed a two-way logistic regression using condition (dummy coded: prompt = 0, control = 1) and procedural approach score (ranging from 0 to 2) to predict formal approach complaints (a binary code). The results are shown in Table C1. The model predicted 22% of the overall variance.

The significant interaction term indicates a key difference: for an increase in the procedural approach score, complaints are more likely to increase in the control group. Prompting overrides the association between formal approach complaints and prior problem-solving approach, leading to formal approach complaints for all students. This is especially reflected by the significant condition term, indicating a difference between prompt and control when procedural approach score is zero. The odds ratio of .02 indicates that, for students with a procedural approach score of zero, the odds that control students will make a formal approach complaint are 1/50th of the odds of prompted students.

Table C1. Results of 2-way logistic regression predicting formal approach complaints ($n = 102$).

Predictor	β	S.E.	Odds ratio	p
Intercept	−.95	.60		.11
Condition	−4.15	1.94	.02	.03*
Procedural approach score	0.04	0.41	1.04	.92
Condition × procedural approach score	2.35	1.07	10.5	.03*

Appendix D: Log-linear analysis of condition, diagram type, and approach in the problem-solving phase

We used a log-linear model selection analysis to find associations between the three primary factors in the study: condition, diagram type, and approach. The results of the log-linear analysis on the two problem-solving questions are shown in Tables D1 and D2, illustrating significant condition – diagram and diagram – approach associations.

Table D1. Results from the log-linear model selection analysis for Q1, starting from the overall model (condition × diagram × approach).

Effect for Q1 ($n = 133$)	(Change in) χ^2	df	Sig.
Condition × diagram × approach	4.03	4	.40
Condition × approach	.04	2	.98
Condition × diagram	29.1	2	<.001
Diagram × approach	38.9	4	<.001
Final model	4.07	6	.67

Table D2. Results from the log-linear model selection analysis for Q2, starting from the overall model (condition × diagram × approach).

Effect for Q2 ($n = 127$)	(Change in) χ^2	df	Sig.
Condition × diagram × approach	5.68	4	.22
Condition × approach	1.24	2	.54
Condition × diagram	31.7	2	<.001
Diagram × approach	33.6	4	<.001
Final model	6.92	6	.33