

# What do students do when asked to diagnose their mistakes? Does it help them?

## I. An atypical quiz context

Edit Yerushalmi and Elisheva Cohen

*Department of Science Teaching, The Weizmann Institute of Science, 76100 Rehovot, Israel*

Andrew Mason

*School of Physics and Astronomy, University of Minnesota, Minneapolis, Minnesota 55455, USA*

Chandralekha Singh

*Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, Pennsylvania 15213, USA*  
(Received 12 November 2010; revised manuscript received 10 January 2012; published 21 September 2012)

“Self-diagnosis tasks” are aimed at fostering diagnostic behavior by explicitly requiring students to present diagnosis as part of the activity of reviewing their problem solutions. Recitation groups in an introductory physics class of about 200 college students were distributed into a control group and three intervention groups in which different levels of guidance were provided for performing self-diagnosis activities. We investigated how well students self-diagnose their solutions in the different interventions and examined the effect of students’ self-diagnosis on subsequent problem solving in the different intervention groups. We found that in the context of an atypical quiz, while external support altered the self-diagnosis performance, the self-diagnosis score was not correlated with subsequent problem-solving performance on a transfer problem. We discuss possible explanations for our findings.

DOI: [10.1103/PhysRevSTPER.8.020109](https://doi.org/10.1103/PhysRevSTPER.8.020109)

PACS numbers: 01.40.Fk, 01.40.gb, 01.40.Ha

### I. INTRODUCTION

The purpose of this study was to shed light on the learning processes and outcomes associated with several alternative tasks we term “self-diagnosis tasks.” These tasks are a type of *formative assessment* tasks [1–3]. While traditional summative assessment emphasizes product and grades, these tasks emphasize process and feedback. They are aimed at fostering *intentional learning* [4] in the context of problem solving, i.e., ensuring that students will not focus their attention only on the specific task, but also on the more general learning goal of elaborating the solver’s conceptual understanding. They do so by installing, in the context of the assessment, an activity that encourages students to provide *self-explanations* [5] when reading and trying to make sense of problem solutions.

The activity of reviewing and diagnosing a solution that contains mistakes formerly made by the learner requires reading and interpreting a possibly deficient textual artifact from a new perspective. This artifact to some extent represents the mental model of the student when they first approached the problem on the quiz. Students might gain a new perspective on their own prior work merely by asking themselves reflective questions, in which they try to clarify what they did and why they did it. Novice students,

however, might need external support in the form of an instructors’ solved example or a diagnosis rubric that focuses their attention on possible mistakes.

In the study reported here, we focus on three different kinds of self-diagnosis tasks. The different activities share a similar basic structure: students get time and credit for presenting a diagnosis of their quiz problem solutions. College students in an introductory physics class are given their photocopied solutions and asked to identify and circle where they went wrong and explain the nature of the mistakes. The tasks vary in the external support (instructions and resources) students receive. In one task they receive minimal support—just the final answer to the problem—and they are allowed to use their notes and textbooks. In a second task, students diagnose the mistakes in their solution after the instructor presents on the whiteboard an outline of the solution, which is concise and “product oriented” [6], and skips some details of the derivations. The reasoning that underlies various solution steps is presented orally. Students are also asked to sort their mistakes according to major problem-solving steps described in the research literature [7–9] (e.g., “description,” “plan”). In a third task they compare their own solution to a detailed, “process oriented” [6] written solution. The written solution is not accompanied by oral explanations and includes both the reasoning and the detailed calculation.

The problems used in the study resemble to some extent “context-rich problems” [10] in the sense that they are represented as a story set in a real-world context rather than in an abstract form and are not broken into subproblems.

---

*Published by the American Physical Society under the terms of the [Creative Commons Attribution 3.0 License](https://creativecommons.org/licenses/by/3.0/). Further distribution of this work must maintain attribution to the author(s) and the published article’s title, journal citation, and DOI.*

Context-rich problems are suggested in research-based curricula [10] as an appropriate setting for instruction that is intended to move students toward more expertlike problem-solving behavior. These problems are represented in a way that differs from the more traditional textbook problems to which the solver might be more accustomed. Because of this, they prevent students from merely retrieving solutions to similar problems they have coped with in the past. To solve the problems, students need to sort through the problem narrative while identifying the physics concepts and principles needed to solve them and explore alternative solution paths. Thus, these problems provide an opportunity for reviewing the solution and for reflecting on and refining the solver's understanding of the concepts and principles they have been applying. The self-diagnosis tasks are designed to verify that students will indeed make use of this opportunity: students were first asked to solve the quiz problems and then diagnose their solution.

In this setting we attempt to answer the following questions:

- (1) How well do students self-diagnose and correct the mistakes in their solutions in different kinds of self-diagnosis tasks?
- (2) What is the effect of students' self-diagnosis in different kinds of self-diagnosis tasks on their subsequent problem solving?

Research in the context of studying mistaken solutions to probability problems indicates that such an activity is more advantageous for learners with a relatively high level of knowledge [11]. A solver's knowledge regarding a certain set of problems depends on how accustomed they are to such tasks. Students who have gained expertise in solving problems characterized by a particular problem situation encountered several times in the course might not know how to cope with problem situations that are less typical.

To determine in what way students' performance in self-diagnosis tasks in physics depends on their prior knowledge, we carried out the study first with an atypical problem and then with a typical one. A typical problem is one for which isomorphic problem(s) have been given as homework and for which the textbook, or the teaching assistant in the recitation group, presents solutions. An atypical problem is one for which no isomorphic problem(s) are given. This paper focuses on students' self-diagnosis performance in the atypical problem situation, where they might be expected to need further support, and in which the depth of self-diagnosis can prove crucial for transfer. In a subsequent paper we will present findings regarding students' performance on self-diagnosis in a typical problem situation and discuss the effect that the typicality of the problem situation has on students' performance on self-diagnosis and subsequent transfer.

We first present some background that constitutes the rationale for developing the self-diagnosis tasks studied

and informs their design. We then describe both the different self-diagnosis tasks and the time line for the different treatment groups in the experimental setup. Before presenting our findings, we describe the special analysis tools we developed to assess students' self-diagnosis.

## II. BACKGROUND

The idea that learning is promoted by reviewing solutions composed by the learners is common among physics instructors [12]. It is common practice to tell students that they should correct their solutions or compare them to a worked-out example. This idea is supported by research that compares the learning achieved by experts and novices from solving problems, either through the problem-solving process or after the fact when reading a worked-out example. Instructors suspect, however, that their students do not necessarily review their solutions in the way they would expect [12]. This concern is addressed by the research literature that focuses on formative assessment.

### A. Learning through problem solving: Experts versus novices

Expert problem solving is characterized by a strategic approach, in which the solver carries out a qualitative analysis of the situation and develops a plan to solve the problem [13–16]. The strategic approach enables the solver to envision the entire problem-solving process while at the same time focusing on specific decisions that need to be made at a particular point in the process. Experts are also recognized by continuous evaluation of their progress [17]. Their problem-solving process is iterative and they monitor their progress towards a solution by asking themselves implicit reflective questions such as, What am I doing? Why am I doing it? How does this help me? [18]. These questions serve to elaborate the solution between successive trials. While self-monitoring is directed mainly towards arriving at a solution, it might also involve self-diagnosis directed towards more general learning goals, such as elaborating conceptual understanding. In contrast to experts, novices often start solving a problem by quickly choosing an approach and then sticking to that approach even if it is unfruitful. They are also not likely to evaluate their final answer. Thus, there is a need to improve novice students' knowledge as well as their inclination to evaluate their progress and elaborate their understanding within the problem-solving process.

Researchers in physics education have shown that, in small experimental settings, having students follow an explicit expertlike problem-solving strategy significantly improves their problem-solving performance [19]. Curriculum developers have adapted materials and pedagogical techniques shown to be effective in these experimental settings to realistic classroom settings and institutional constraints [9,10,20–22]. Many of these instructional interventions share the common approach of

TABLE I. Distribution of students into control and intervention groups.

	Control A	Intervention B	Intervention C	Intervention D
Pre	36 students	30 students	28 students	25 students
Post	29 students	18 students	21 students	13 students

cognitive apprenticeship [23] in which students work collaboratively or with a computer [24,25], so that they must externalize and explain their thinking while they solve a problem. In many of these interventions students are provided with modeling of a problem-solving strategy that externalizes the implicit problem-solving strategies used by experts, and they are required to use it. Different strategies may involve similar steps (see Table I): (1) describe the problem, (2) plan and construct a solution, (3) check and evaluate the solution. The cognitive apprenticeship approach aims at mimicking the culture of expert practice. Accordingly, instructional interventions aligned with it often make use of realistic and challenging problems. Arriving at solutions for these kinds of problems is indeed facilitated by the use of an expertlike approach [10]. Such strategies have been shown to improve students' problem-solving skills (planning and evaluating rather than searching for the appropriate equation without evaluating) as well as their understanding of physics concepts [9,26,27].

### B. Learning from worked-out examples

Worked examples in standard textbooks do not explicitly justify the transition from step to step [5]. To make sense of the presented text, students must provide their own explanations about each component of the solution and the reasoning underlying the derivation. Content-relevant articulations that are formulated by students and state something beyond what the text explicitly states are termed "self-explanations" [5]. Students who self-explain more were found to learn more, even if their self-explanations were fragmented and sometimes incorrect. Two central mechanisms for constructing self-explanations have been suggested, namely, generation of inferences and self-repair [28]. In the first mechanism the learner fills in the gaps in the incomplete text. In the second mechanism the learner realizes and acknowledges a conflict between their deficient mental model and the scientific model underlying the worked example. In other words, learners (a) compare the worked-out example and their own solution and realize that there are *omissions*, i.e., differences that are critical for finding the right solution, and (b) acknowledge *violations*, or conflicts between a text sentence in the worked-out example and a belief that is embedded in their mental model. Resolving this conflict allows the solver to repair the flawed mental model. It was shown that the activity of self-explanation can be enhanced through interventions that require students to present their explanations [29,30],

and that encouraging students to provide justifications, in peer interaction or in human computer interaction, leads to significant learning gains [31].

### C. Supporting learning from problem solving through formative assessment

One challenge often posed by these approaches is that the assessment is a traditional one, focused on the product rather than the process, thus undermining the intended outcomes. The negative impacts of traditional assessment include overemphasis on grades and underemphasis on feedback to promote learning. Thus, traditional assessment approaches are lacking in formative assessment. Black and William [1] suggest that for formative assessment to be productive, students should be trained in self-assessment so that they can understand the main purposes of their learning and thereby grasp what they need to do to succeed. These authors claim that tests and homework can be invaluable guides to learning, but they must be clear and relevant to learning goals. The feedback should provide guidance on how to improve, and students must be given both opportunity and help to work on the improvement.

One such alternative instructional strategy involves self-correction tasks that are carried out after the solution has been completed and requires students to submit a corrected solution. However, in self-correction tasks, as well as in interaction with peers or a computer, self-diagnosis is not guaranteed to occur. Students may or may not self-diagnose the solution, namely, identify where they went wrong and realize the nature of the mistakes. One instructional strategy that bypasses this difficulty is to require students to present a diagnosis (identification and explanation of mistakes) as part of the activity of reviewing their own solutions. We shall call the tasks involved in such instructional strategy self-diagnosis tasks. Self-diagnosis tasks involve an *explicit requirement* to carry out *self-diagnosis activities* when given some feedback on the solution. A variety of self-diagnosis tasks are reported in the literature on self-assessment [1–3] and particularly in physics education [32,33].

Perkins and Swartz [34] define a scale for thinking processes on the basis of the individual's awareness of these processes: (1) *an implicit process*, in which the individual is not aware of the thinking process; (2) *a partially explicit process*, in which the individual explicitly identifies the activities (e.g., "I am making a decision now" or "I look for evidence"); (3) *a strategic process*, in which the individual plans and carries out an organized sequence of activities using tools for thinking and decision making ("this problem is complicated, so I will think of a simpler one"); and (4) *a reflective process*, in which individuals use critical thinking tools to improve their thinking. While self-correction tasks or cooperative work promote mainly the implicit or partially explicit process, self-diagnosis tasks promote the fourth level, reflective

diagnosis. The reflective process refers to what in the taxonomy of Bloom *et al.* [35] is termed the skill of evaluation, and is considered to be the highest order of learning skills.

Self-diagnosis with the aid of a worked-out example requires that two textual artifacts be interpreted in parallel: (a) the student's possibly deficient textual artifact and (b) the worked-out example (whether a teaching assistant's outline, an instructor's written example, or the textbook). While there is a similarity between these two contexts, there is also a difference that may change the characteristics of the learning processes and outcomes. In self-explaining a worked-out example, someone else has constructed the artifact that the learner reads. Thus, the text serves as a mediator between the mental model of the expert and that of the learner. As a result, there is a need to negotiate between the mental model of the learner and the intention of the writer, and this imposes additional cognitive load. The learners have to relate their own knowledge to the product of a process that has been carried out by someone else. By contrast, in the context of self-explaining an artifact produced by the learners, students interact with the outcomes of processes that they themselves have carried out, and thus they can interpret them more directly. Another difference has to do with the correctness and coherence of the artifact. A solution produced by an expert would be expected to be more correct and coherent than a solution produced by the student. Combining the two processes, i.e., interpreting both the worked-out example and the student-made solution, is extremely demanding. Accordingly, self-diagnosis tasks need to be accompanied by varied supports along several dimensions:

- (1) Instructions on how to carry out the diagnosis (e.g., the level of detail required in laying out possible deficiencies in a student's approach towards the solution and in its implementation; such instructions might include spoken guidelines, self-diagnosis rubrics, structure of worked-out example, etc.).
- (2) Resources available to the students while they are diagnosing their solutions (e.g., information provided about the correct solution, notebooks, and textbook).
- (3) Provision of customized feedback (e.g., diagnostic information about the solution, provided by the instructor).

The composition of the above can be used to calibrate different tasks for different levels of students. Consider, for example, the case where the resource for diagnosis is a worked-out example. Research in that context focuses on making analogies [36–38] and shows that many students do not know how to use a worked-out example to solve a transfer problem (i.e., one that is similar in required general procedure but different in detailed procedures). If the students' representation is organized around surface features [39], this will prevent them from retrieving and

implementing procedures from the worked-out example [40]. Research has shown [37] that medium- and high-achieving students benefit most from instruction that explicitly presents them with both the procedures and worked-out examples, as opposed to either the worked-out examples or the principles separately. This finding suggests guidelines on preparing worked-out examples that will be useful to students attempting self-diagnosis.

### III. EXPERIMENTAL SETUP

The purpose of the experiment was to compare the learning processes and outcomes in different self-diagnosis tasks. In each task, students carried out self-correction or self-diagnosis of their solutions after attempting to solve an atypical physics problem.

#### A. Study sample

The study participants were drawn from an introductory algebra-based physics course for premedical students ( $N \sim 200$ ), with one instructor and two teaching assistants. The course recitation groups were divided into two control groups (groups A, A') and four self-diagnosis intervention groups (B, B', C, D). Groups B and B' were slated to undergo the same intervention. Duncan analysis was performed on students' answers to the quiz to determine which groups were similar. Based on the results showing that groups A', B, C, and D were similar ( $P > 0.05$ ), we selected these four groups for different treatments, as detailed in the following. One teaching assistant worked with groups A and B and the other worked with groups C and D.

Each of the intervention groups (groups B, C, and D) carried out a different self-diagnosis task on the solution of a problem presented in a quiz during the previous week (hereafter defined as the "pre" problem). Initial training was provided using the quiz given in week 5, and the actual intervention took place using the quizzes given in the 6th week. At midterm (within 1 week after completion of the self-diagnosis task) the students were presented with a transfer problem (defined as the "post" problem), which was paired to the pre problem. The composition of the groups remained constant over this period. The post problem will be further explained later on, but for the moment we mention that several students who performed the pre problem did not attempt the post problem. In addition, some students who attempted the post problem were nevertheless excluded from the analysis of its performance because of their failure to answer the first part of the post problem correctly; a correct answer to the first part was a prerequisite for enabling the rest of the problem to serve as a transfer problem. Table I shows the distribution of students into control and intervention groups. Because of the drop in the number of students who were analyzed for the posttest, within each group we compared the pretest scores of students who participated in the posttest and of those who did not. For this comparison we used the



**Problem description:** Represent the problem in physics terms: Draw a sketch, list known and unknown quantities, target variable

**Solution construction:** Present the solution as a set of subproblems. In each subproblem write down:

- The unknown quantity you are looking for
- The physics principles you'll use to find it
- The process to extract the unknown

**Check answer:** Write down how you checked whether your final answer is reasonable

FIG. 1. Guidelines for presenting a problem solution according to a problem-solving strategy.

Wilcoxon nonparametric test (using Kruskal-Wallis statistics). No significant differences were found in any of these comparisons.

**B. Interventions and control**

Prior to the intervention, the students had become acquainted with context-rich problems (see later for details) as the type of problem used, since more than half of the weekly quizzes (12 quizzes altogether) required them to solve context-rich problems. Also, early in the semester, instructors had provided students with guidelines for presenting their problem solutions in a manner based on the steps of a problem-solving strategy [7,8] (see Fig. 1).

Initial training was provided using the quiz given in week 5. and the actual intervention took place using the quizzes given in the 6th week (hereafter referred to as “quiz 6”). Each intervention group followed a modified version of the intervention sequence. Students were given a sample incorrect solution of the “training problem” and diagnosed it according to their treatment group. Then, the instructor demonstrated how that incorrect solution should be diagnosed for the training problem. Table II shows the sequence followed in all intervention groups.

Problem A:  
 A friend tells a girl he has heard that if you sit on a scale while riding a roller coaster, the dial on the scale changes all the time. The girl decides to check the story and takes a bathroom scale to the amusement park. There she receives an illustration (see below), depicting the riding track of a roller coaster car along with information on the track (the illustration scale is not accurate). The operator of the ride informs her that the rail track is smooth, the mass of the car is 120 kg, and the car sets in motion from a rest position at the height of 15 m. He adds that point B is at 5 m height and that close to point B the track is part of a circle with a radius of 30 m. Before leaving the house, the girl stepped on the scale which indicated 55 kg. In the rollercoaster car the girl sits on the scale. Do you think that the story she has heard about the reading of the scale changing on the roller coaster is true? According to your calculation, what will the scale show at point B?

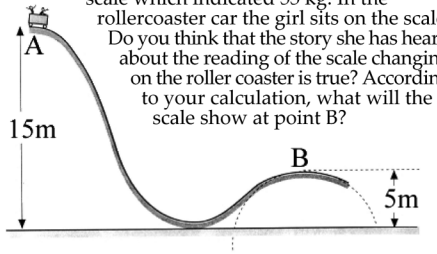


FIG. 2. The pre problem used in a self-diagnosis task.

*1. The pre problem*

Problem A, used for the first self-diagnosis task in the 6th quiz in the semester, is shown in Fig. 2. This problem is an atypical one in the course, as it involves a nonequilibrium application of Newton’s second law in the context of a nonuniform circular motion. This is an uncommon example of a quiz problem in an algebra-based introductory physics course. It is a context-rich problem, i.e., it has a context and motivation connected to reality, with no explicit cues (e.g., “apparent weight”), and contains more information than needed (e.g., the car’s mass). It requires the students to analyze the problem statement, determine which principles of physics are useful and what approximations are needed (e.g., no friction for a smooth track),

TABLE II. Experimental sequence. TA refers to the teaching assistant.

		Intervention groups			
		Control group A'	B	C	D
Initial training	Quiz 5	The training follows a modified version of the intervention sequence: Students are given, for demonstration purposes, an incorrect solution of a “training problem” and diagnose it according to the procedure specified by their intervention group. Then the TA demonstrates how the incorrect solution should be diagnosed for the training problem.			
Pre	Quiz 6	Students solve the quiz problem			
Intervention	Self-diagnosis of quiz 6	TA discusses solution of the pre problem	TA presents in class an outline of the solution. Students self-diagnose their solutions: circle mistaken parts, explain, and sort mistakes in a self-diagnosis rubric	TA provides a written worked-out example. Students self-diagnose their solutions: they circle mistaken parts and explain the mistakes	Students can use their notes and text books. Students write their self-diagnosis
Post	Midterm	Students are given a transfer problem paired to the pre problem			

General evaluation	Performance level	Explain what is missing		
Problem description	Full / Partial Missing	* In sketch * Known / unknowns		
Solution construction	Full / Partial Missing	* Subproblem's unknown * Principles used		
Check answer	Full / Partial Missing	* Possible checks for reasonability		
Circle and number mistakes you find in the solution. Fill in the following rubric				
Mistake #	Mark x if mistake is in:			Explain mistake
	Physics	Math	Other	
1				
...				
...				

FIG. 3. Self-diagnosis rubric.

and realize which subproblems are needed to solve the problem.

## 2. Resources and guidelines for self-diagnosis in the different groups

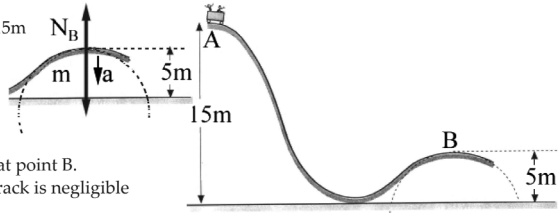
In the control group (recitation group A'), the instructor discussed the problem's solution with the students in the

recitation following the quiz, but they were not required to engage in a self-diagnosis task. The instructors in the three intervention groups (B, C, and D) gave their students a photocopy of their solutions in the recitation following the quiz and asked them to diagnose mistakes in their last week's quiz solution. Students were credited with 50% of their original quiz grade for completing the diagnosis. The credit was effort based. The instructors also motivated them by saying that self-diagnosis would help them learn.

In intervention group B the instructor discussed an outline of the correct solution with the students and they were required to fill in a self-diagnosis rubric (Fig. 3). The rubric was designed to direct students' attention to two possible types of deficiencies: deficiencies in approaching the problem in a systematic manner ("general evaluation") and deficiencies in the physics applied. The general evaluation part was intended to direct students' attention to the steps in a systematic solution process in which they had made mistakes, as well as to the presentation of the reasoning in a systematic manner.

The rubric structure was designed by teachers who had participated in a professional development workshop in which they had customized instructional innovations to promote self-monitoring in physics problem solving [41].

**1. Description of the problem**  
 Knowns: The height of release:  $h_A=15\text{m}$   
 Speed of the car at point A:  $v_A=0$   
 The height at point B:  $h_B=5\text{m}$   
 The radius at point B:  $R_B=30\text{m}$   
 The mass of the car:  $M=120\text{ kg}$   
 The mass of the girl:  $m=55\text{kg}$   
 Target quantity:  $N_B = \text{Normal force at point B.}$   
 Assumptions: the friction with the track is negligible



**2. Constructing the solution**  
**Plan:** During the motion of the girl along the curved track, the magnitude of her velocity as well as the radial acceleration changes from point to point. If the radial acceleration changes, we infer that the net force acting in the radial direction on the girl changes as well. The net force on the girl is the sum of 2 forces acting on her: the force of gravity and the normal force. To calculate the normal force at point B we can use Newton's 2<sup>nd</sup> law  $\Sigma \vec{F} = m \vec{a}$ ; however we will need to know the acceleration at this point. To calculate the centripetal acceleration at B, we need to know the speed at point B.  
 We calculate the velocity of the girl at point B using the law of conservation of mechanical energy between point of departure A and point B (the mechanical energy is conserved since the only force that does work is the force of gravity, which is a conservative force; the normal force does no work because it is perpendicular to the velocity at every point on the curve).

**Subproblem 1: calculating the velocity at point B**  
 We'll set ground as the reference for potential energy and compare the total mechanical energies of the girl and car at points A and B:  $E_A = E_B$   
 Since the speed is zero at point A the kinetic energy is zero at that point. Therefore we get:  
 $PE_A = KE_B + PE_B$   
 $\Rightarrow (m + M)gh_A = (m+M)gh + \frac{1}{2} (m+M)V_B^2$   
 $\Rightarrow gh_A = gh + \frac{V_B^2}{2}$   
 We can calculate the speed at B in the following way  
 $V_B^2 = 2gh_A - 2gh_B = 2(10\text{m/s}^2)(15\text{m}-5\text{m}) = 200\text{m}^2/\text{s}^2$

**Subproblem 2: calculating the normal force at B**  
 Using Newton 2<sup>nd</sup> law  $\Sigma \vec{F} = m \vec{a}_R$   
 $N_B - mg = -ma_R$   
 $N_B - mg = -ma_B^2/R_B$   
 $N_B = mg - ma_B^2/R_B$   
 $\Rightarrow N_B = (55\text{kg})(10\text{m/s}^2) - (55\text{kg})(200\text{m}^2/\text{s}^2)/(30\text{m})$   
 $\Rightarrow N_B = 18333\text{N}$   
 Final result: when the car crosses Point B at the track, the scale indicates 183.3 N

**3. Reasonability check of the final result:**  
 limiting cases of the parametric solution for  
 $N_B = mg - mV_B^2/R_B$  At rest ( $v=0$ ):  $N = mg$ ;  
 On horizontal surface ( $R \rightarrow \text{infinity}$ ):  $N = mg$

FIG. 4. Worked-out example for the pre problem, aligned with guidelines.

In intervention group C the students were provided with a worked-out example (Fig. 4) that the instructor handed out during the self-diagnosis activity. This solution followed the guidelines for presenting a problem solution (Fig. 2). The students were asked to circle their mistakes and explain on their photocopied solutions what they had done wrong in that part.

In intervention group D students received minimal guidance. They were asked to circle their mistakes and explain on their photocopied solutions what they had done wrong in that part, aided by their notes and books only, without being provided with the solution.

### 3. The post problem

One of the problems that was presented at the midterm functioned as a post (transfer) problem. This was used to determine the ability of students to transfer the understanding gained when diagnosing the pre problem in order to solve a problem in a somewhat similar context (motion along a vertical circular path in a nonequilibrium situation). The midterm took place within a week from the recitation in which students had self-diagnosed their solutions of the pre problem. The post problem is portrayed in Fig. 5.

Both problems employ the same physical principles, i.e., Newton's second law applied in a nonequilibrium situation involving centripetal acceleration and conservation of mechanical energy (EC). Furthermore, both problems require recognition of similar target variables (in the form of either a normal force  $N$  or a tension force  $T$ ) and intermediate variables (centripetal acceleration  $a_c$  and velocity  $v$  at the maximum or minimum point on a circular trajectory). The problems differ in context and also in that in solving the midterm problem, students must first realize that they should focus on the lowest point on the circular trajectory.

Table III shows how the midterm problem was designed to be a transfer problem to the pre problem.

While the post problem employs the same physical principles and requires recognition of similar intermediate and target variables as those of the pre problem, it can serve as a transfer problem to the pre problem only if students

A family decides to create a tire swing in their backyard for their son Ryan. They tie a nylon rope to a branch that is located 16 m above the ground, and adjust it so that the tire swings 1 m above the ground. To make the ride more exciting, they construct a launch point that is 13 m above the ground, so that they don't have to push Ryan all the time. You are their neighbor, and you are concerned that the ride might not be safe, so you calculate the maximum tension in the rope to see if it will hold. (a) Where is the tension greatest? (b) Calculate the maximum tension in the rope, assuming that Ryan (mass 30 kg) starts from rest from his launch pad. Is it greater than the rated value of 750 N? (c) Name two factors that may have been ignored in the above analysis, and describe whether they make the ride less safe or more safe.

FIG. 5. The post (midterm) problem.

TABLE III. Comparison of pre and post problems. The FBD column refers to the relative orientations of acting forces in the free body diagram.

	Principles	Variables	FBD	Context	Details
Pre	EC	$v$	$\uparrow N \downarrow F_g$	Roller coaster	$\downarrow a_c \uparrow N$
Post	2nd law	$a_c N/T$	$\uparrow T \downarrow F_g$	Tire swing	$\uparrow a_c \uparrow N$

realize, while attempting to solve part (a) of the post problem, that the tension is greatest at the bottom of Ryan's path. Students who failed to answer this part correctly were excluded from the analysis.

## IV. ANALYSIS TOOLS

Analysis of data requires a robust method of grading that will allow assessment of students' solutions as well as students' self-diagnosis. We developed a scoring rubric so that only minor changes to its structure will be needed to adapt the rubric to different kinds of problems.

### A. Structure of the rubric

In constructing the rubric, we integrated top-down and bottom-up approaches [42]. In the top-down approach we constructed an *a priori* representation of the "ideal knowledge" underlying an expert approach to the problem and looked for the extent to which each student's approach included certain elements of the ideal knowledge. We developed a scoring rubric that enabled us to identify the gap between the student's knowledge reflected in their solution and the ideal knowledge needed to solve the problem appropriately.

In representing the expert ideal knowledge in the rubric, there were both generic and specific elements to consider. Specific elements have to do with the physical principles the student must invoke and apply in order to solve a specific problem. Thus, we constructed a "physics" category, which was divided into two subcategories: *invoking* the principles of physics and *applying* these principles. Taken together, each row in each subcategory therefore represented every physical principle that a student would need to invoke and apply to correctly solve the specific problem.

General elements have to do with the presentation of the solution, i.e., how well it is communicated and justified. As students in the course were asked to follow a strategic problem-solving approach, we considered the solution to be "ideally" communicated and justified if the student presented it according to that approach. This approach can be found in several instruments developed by other researchers [33,43]. Thus, we constructed in the rubric a "presentation" knowledge category that includes three subcategories. In the *problem description* subcategory we have to judge whether the student has presented a helpful description of the problem's situation in terms of physics concepts and principles, e.g., by drawing a diagram to help

visualize the problem. The *planning and solution* subcategory requires us to judge whether the student has constructed a good plan for solving the problem with regard to the target quantity and intermediate problem steps needed to obtain this quantity. In the *evaluation* subcategory we must determine whether the student has checked the reasonability of the answer so as to make sure that the problem has been correctly solved (e.g., units, limiting cases, etc.). These three presentation subcategories are essentially general and do not have to be changed from problem to problem; only the specific criteria based upon the subcategories need to be changed.

However, to determine how the students are actually thinking and the possibly incorrect mental models that they use to solve the problem, a bottom-up approach must be considered simultaneously. To that end, we went over students' work and identified common mistakes in approaches to solving the problem. We represented these common approaches in the rubric under "novice incorrect ideas." The rubric has additional rows in the physics subsection that tracks whether a student has invoked an inappropriate principle that is not applicable to the problem or has inappropriately applied principles that were correctly invoked. Such analysis allows us to identify students who realize that they have made some specific mistake even though they cannot correct it.

The work of each student was evaluated in three ways. The first evaluation was done according to the researcher's diagnosis of the student's quiz solution (RDS). The second evaluation was done according to the student's self-diagnosis of their solution (SDS). Here we used the rubric<sup>1</sup> to interpret the student's self-diagnosis. The third evaluation was made according to the researcher's judgment of this student's self-diagnosis (RSD); here we compared the researchers' and the student's diagnosis of the student's solution. To represent these three ways, we constructed three columns in the rubric.

After the categories were coded, each of the three evaluations was scored. In the RDS and the RSD columns, we assign "+" if a student performs correctly or identifies a mistake defined by some subcategory. We assign "-" if the student performs incorrectly or fails to identify a mistake or identifies it incorrectly. Each "+" is worth 1 point and each "-" is worth 0. If a student is judged to have gotten something partially correct, then the grader may assign ++/- (i.e., 0.66), +/- (i.e., 0.5), or +/ - -

<sup>1</sup>Initially we planned to differentiate, in evaluating this self-diagnosis, between three levels of diagnosing mistakes in the physics category: (1) realizing which principles or concepts had not been invoked or applied correctly, (2) explaining why that application was wrong, and (3) providing a correction by showing how these principles or concepts could have been applied correctly. Eventually, we did not apply this differentiation because it was difficult to differentiate between (1) and (3), and most students skipped level (2) in their diagnosis.

(i.e., 0.33). We assign the term "NA" if the student could not reasonably address a subcategory given the prior work done. In the SDS column, if a student correctly diagnoses a mistake we assign "-" (since this is the grade the student gives to their solution). If a student does not refer to a mistake they have made, we assign "×," and a 1-point score is assigned in this category because we assume that the student treated the mistake as a correct thing to do.

The content validity of the rubric, namely, the extent to which it allows us to map the student's solution to the expert ideal knowledge as well as to the novice incorrect ideas, was determined by four experts in physics education who perceived it as measuring an appropriate performance of the solution and self-diagnosis.

The analysis was done by two researchers, and any disagreements were discussed and resolved. The inter-rater reliability achieved was 80% for 20% of the students graded by both researchers before discussion, and almost 100% after discussion.

## B. Scoring rubric for the pre problem

Table IV shows the rubric adapted to the pre problem (Fig. 2), along with scoring of one student who was assigned to intervention group B (outline plus self-diagnosis rubric provided as scaffolding tools). Figure 6 represents the student's solution to the problem and their attempt at self-diagnosis.

In the following section we explain how the structure of the rubric explained above is adapted to the pre problem and demonstrate how we used this rubric to assess an example of a student solution and self-diagnosis (the problem used and the question asked are presented in Fig. 2).

The rubric addresses both the ideal knowledge and the novice incorrect ideas. Under the heading "Physics principles," the ideal knowledge requires students to understand that the target variable is the normal force the scale exerts on the girl. To calculate the normal force at point B, the students will have to define an appropriate system (the "girl") to invoke and apply to it Newton's second law:  $\Sigma \vec{F} = m \cdot \vec{a}$ . They will have to realize that  $\vec{F}$  is the net force (sum of the force of gravity and the normal force). To calculate the acceleration at this point they will have to invoke the expression for centripetal acceleration:  $a_R = v_B^2/R$ . The intermediate variable (the speed of the cart at the top of the circular bump) can be found using the law of conservation of mechanical energy,  $PE_i + KE_i = PE_f + KE_f$ , between the point of departure and point B at the top of the bump (which is justified because all forces doing work are conservative forces). We found that the common incorrect ideas in the physics part had to do with defining the system inappropriately and/or inconsistently, calculating KE and PE without reference to energy conservation, calculating Newton's second law while ignoring the circular motion or the normal force, and referring to centripetal



TABLE IV. Self-diagnosis grading rubric adapted to pre problem. In italics, grading of an individual student from group B (Fig. 6). The last three column respectively refer to a researcher diagnosis of student’s solution (RDS), student’s self-diagnosis of solution (SDS), researcher’s assessment of student’s self-diagnosis (RSD). (NA indicates an item being not applicable.)

General task	Specific criteria	RDS	SDS	RSD	
<b>Physics principles</b>					
Invoking principles	1. Conservation of mechanical energy	+	×	NA	
	2. Nonequilibrium applications of Newton’s 2nd law in centripetal motion ( <i>student does not invoke Newton 2nd law</i> )	–	–	+	
	3. Defining the system appropriately and consistently ( <i>student defines the system as the girl plus car rather than girl</i> )	–	–	+	
	<b>Incorrect ideas:</b> inappropriate principle: “–” marked if inappropriate principle is used in student’s solution or diagnosis ( <i>student uses gravitational law</i> )	–	–	+	
Applying principles	1. Conservation of mechanical energy	–	–	+	
	2. Nonequilibrium applications of Newton’s 2nd law	NA	NA	NA	
		<b>Incorrect ideas:</b> referring to centripetal force as a physical force <b>Incorrect ideas:</b> ignoring normal force (etc.)	–	–	+
Algebra	Algebraic manipulation	+	×	NA	
<b>Presentation</b>					
Description	1. Invokes a visual representation	FBD, acceleration vector, axis, defining PE = 0, radius of the circle	–, –, –, +, +	–, ×, ×, ×, ×	+ , – , – , NA, NA
Plan or solution construction representing the problem as a set of subproblems	2. Clear or appropriate knowns ( <i>student missed some</i> )		+ + /–	+	+ + /–
	1. Appropriate target quantity chosen ( <i>student chose <math>F_g</math></i> )		–	–	+
	2. Did not write down surplus equations or intermediate variables ( <i>student did not relate the energy to the gravitational force calculations presented</i> )		–	×	–
	3. Appropriate intermediate variables explicitly stated		+ /–	+ /–	+
	4. Explicitly stating in words or a generic form the principles used to solve for this intermediate variables ( <i>Not reflected in the student’s solution</i> )		–	×	–
Evaluation (check)	1. Writing down the units		+ /–	×	+ /–
	2. Checking the reasonability of the final answer (units, limiting cases, etc.)		–	–	+

force as a physical force; i.e., some students, instead of perceiving the centripetal force as the radial component of the net force that brings about the centripetal acceleration, treated it as a force in its own right, i.e., as if it relates to some interaction between two objects, whether long-range interactions (such as gravitational interaction) or contact interaction (such as normal force). To reflect the ideal knowledge as well as common incorrect ideas, we identified two rows in the “invoking principles” section that allow us to judge whether the student has referred to the required principles (conservation of mechanical energy

and nonequilibrium application of Newton’s second law in circular motion) and two corresponding rows in the “applying principles” subcategory to judge whether the student had applied them correctly. We added a row to mark whether the student had defined the system appropriately. We also added an *algebra* category to the rubric to reflect calculation mistakes made during the problem-solving process, e.g., forgetting a coefficient when rewriting an equation. As solutions consist of several subproblems intended to determine subgoals, it is difficult to differentiate between ideas that work together to find

Quiz:6 P0110 10/10/09

General evaluation:

	Performance level	? Explain what is missing
Problem description	Full / Partial / Missing	<ul style="list-style-type: none"> <li>Did not look for the normal force.</li> <li>No free body diagram drawn.</li> </ul>
Solution construction	Full / Partial / Missing	<ul style="list-style-type: none"> <li>Did not write out the steps that were to be taken to solve.</li> <li>used wrong equations.</li> </ul>
Check answer	Full / Partial / Missing	<ul style="list-style-type: none"> <li>The is no evidence of an answer check.</li> </ul>

• Circle and number mistakes you find in the solution  
• Fill in the following rubric

Diagnosis of the mistakes:

Mistake #	:Mark X if mistake is in			Explain mistake	Instructor feedback
	Physics	Math	Other		
1	X	X		used wrong equation needed $N = mg - mac = m \frac{v^2}{r}$	
2	X			needed equation $PE_A = PE_B + KE_B$	
3	X			used total mass instead of just girl to find the force	
4	X			Did not separate into $E_A$ and $E_B$	

Quiz:6 P0110

A friend told a girl that he had heard that if you sit on a scale while riding a roller coaster, the dial on the scale changes all the time. The girl decides to check the story and takes a bathroom scale to the amusement park. There she receives an illustration (see below), depicting the riding track of a roller coaster car along with information on the track (the illustration scale is not accurate). The operator of the ride informs her that the rail track is smooth, the mass of the car is 120 kg, and that the car sets in motion from a rest position at the height of 15 m. He adds that point B is at 5m height and that close to point B the track is part of a circle with a radius of 30 m. Before leaving the house, the girl stepped on the scale which indicated 55kg. In the rollercoaster car the girl sits on the scale. Do you think that the story she had heard about the reading of the scale changing on the roller coaster is true? According to your calculation, what will the scale show at point B?

$m_{car} = 120 \text{ kg}$   
 $m_{girl} = 55 \text{ kg}$   
 $h = 15 \text{ m}$   
 $g = 9.8 \text{ m/s}^2$   
 $v = ?$

$PE = mgh$   
 $KE = \frac{1}{2}mv^2$

$PE = mgh$   
 $PE = (m_1 + m_2)gh$   
 $PE = (120 + 55)(9.8)(15 \text{ m})$   
 $PE = 25725$

$mgh = \frac{1}{2}mv^2$   
 $2gh = v^2$   
 $\sqrt{2gh} = v$   
 $\sqrt{2(9.8)(15)} = v$   
 $\sqrt{294} = 17.15 \text{ m/s}$

$F = G \frac{m_1 m_2}{r^2}$   
 $G = 6.67 \times 10^{-11}$   
 $m_1 = 120 \text{ kg}$   
 $m_2 = 55 \text{ kg}$   
 $r = 30 \text{ m}$

$F = 6.67 \times 10^{-11} \left( \frac{120 \times 55}{(30)^2} \right)$   
 $F = 4.67 \times 10^{-11} (7.333)$   
 $F = 48.9 \text{ N}$

A  
 $PE = mgh$   
 $PE = (m_1 + m_2)gh$   
 $PE = (120 + 55)(9.8)(15 \text{ m})$   
 $PE = 25725$

$KE = \frac{1}{2}mv^2$   
 $KE = \frac{1}{2}(175 \text{ kg})(17.15)^2 = 25725.7$

①  
 $F = G \frac{m_1 m_2}{r^2}$   
 $G = 6.67 \times 10^{-11}$   
 $m_1 = 120 \text{ kg}$   
 $m_2 = 55 \text{ kg}$   
 $r = 30 \text{ m}$

$F = 6.67 \times 10^{-11} \left( \frac{120 \times 55}{(30)^2} \right)$   
 $F = 4.67 \times 10^{-11} (7.333)$   
 $F = 48.9 \text{ N}$

②  
 Yes, I think that the story she had heard about the reading of the scale changing on the roller coaster is true.  
 At point B, there will be a F of 48.9N on the scale.

FIG. 6. Solution and self-diagnosis by a student in group B. Circled number 14 is the student’s code number. Circled numbers 1 and 2 are references labeled in the student’s self-diagnosis. Student’s mistakes are (a) defining the system or mass as the girl plus car rather than girl, (b) invoking universal gravitational law instead of Newton’s second law to find  $F_g$  rather than the normal force, (c) invoking energy conservation in an incomplete form, (d) using the wrong height in calculating the potential energy, (e) regarding presentation, the student did not provide explicit subproblems and did not check their answer. Self-diagnosis: The student marks two mistakes on the solution sheet: (1) invoking the gravitational universal law, (2) defining the system or mass as the girl plus car rather than girl. In the self-diagnosis text the student refers to the following: (a) the student was not looking for the normal force (target variable) and how it should be calculated, the student notices they did not sketch a free-body diagram; (b) the student states they needed to use EC law, and how it should have been applied; c) regarding presentation, they realize they did not write the steps taken in order to solve the problem, and did not check the answer.

out some intermediate variables. Accordingly, we did not allocate a separate subcategory for each concept that the students had to invoke. Rather, we grouped ideas that are required within the two major subproblems comprising the solutions: (a) calculating the normal force at point B using Newton’s second law and (b) calculating the velocity of the girl at point B using the law of conservation of mechanical energy between the point of departure A and point B. In this way, invoking the expression for centripetal acceleration, or the idea of a net force, was considered within

the row labeled “Nonequilibrium applications of Newton’s second law.”

Under the heading Presentation, the *description* should include a visual representation of the forces acting on the girl at the point where her weight needs to be calculated (FBD), denoting where  $PE = 0$ , the radius of the circle, the acceleration vector at point B, and definition of the axes. The students have to write down the known variables such as the velocities and heights at points A and B. In *planning*, an appropriate target quantity (the normal force) and

TABLE V. Final sample grades of an individual student.

Category	Grading
RDS	Physics score, 0.17; presentation score, 0.27
SDS	Physics score, 0.17; presentation score, 0.65
RSD	Physics score, 1.00; presentation score, 0.54

intermediate variables (speed and acceleration at point B) must be mentioned. In *evaluation* the student needs to show that the reasonability of the answer was checked.

Table IV shows that a student from group B, whose solution and self-diagnosis are presented in Fig. 6, invoked conservation of energy during the original quiz and therefore did not address it during self-diagnosis. In the researcher's judgment column (RSD), "NA" would then be entered. This student did not invoke Newton's second law, and was therefore assigned "-" in the RDS column. However, the student self-diagnosed this problem, and was therefore assigned "-" in the SDS column (since SDS is the grade the students would give their own solutions as measured by their self-diagnosis) and "+" in the RSD column for diagnosing the problem.

Table V displays how the grades given in Table IV are interpreted as an overall grade, i.e., the average of all possible criteria that the student correctly addressed. Thus, scores range between 0 and 1. Note that under Presentation we initially graded each part separately (i.e., *description*, *planning*, and *evaluation*), and then averaged those grades. In the Physics principles part we calculated a regular average, which did not, however, include the algebraic manipulation.

## V. HYPOTHESIS

Before describing the findings, we present our hypothesis regarding the effects of the various interventions. We expect the various interventions to differ in their impact (a) on the average self-diagnosis grade and consequently on the post (transfer) problem grade in the different groups and (b) on the gap between low-achieving and high-achieving students in the different groups, as measured by the correlations between the pre problem, the self-diagnosis (SD), and the post problem grades within groups.

We expect the effect to manifest itself in the physics principles aspect rather than in presentation, since presenting a solution in an organized and planned manner reflects an approach to communication that we believe is built over time. On the other hand, it has been shown that even a singular self-diagnosis exercise can improve the physics understanding, as students need to realize how they have misapplied specific concepts or principles [44]. Intervening once or twice is not sufficient to change a habit. Moreover, many teachers do not insist that their students present the solution in an organized way, so the

students do not perceive presentation as a learning goal. Therefore, we would not expect the self-diagnosis of the presentation part in the pre problem to affect the presentation part in the post problem.

Hereafter we will refer only to the physics principles aspect. We expect that the greater the external support, the better the average self-diagnosis performance will be, and consequently the better the outcome will be in solving the post problem. This means that we expect group B (which received the most support) to self-diagnose better and then do better on the transfer problem, and group D (which received the least support) to achieve the least in self-diagnosis and hence on the transfer problem.

We expect that the greater the external support, the more "meaningful"<sup>2</sup> the intervention will be in the sense that it will help the lower-achieving students to perform a meaningful diagnosis of their mistakes that will influence their achievement later on. Hence, we expect, in the case of the meaningful intervention, a positive correlation between the self-diagnosis and the post problem. We expect that if the low achievers improve and reduce the gap between them and the better students, the correlation between the pre problem and its self-diagnosis will be insignificant; this also implies that we will expect insignificant correlations between the pre and post problems, since the formerly weak students will improve if they have actually learned from the self-diagnosis.

Where there is less external support we expect the intervention to be "weak" in the sense that it will not reduce the gap between low and high achievers. While allowing the high-achieving students to display their natural diagnostic tendency, it will not aid the low achievers to reflect on and diagnose their solutions. We therefore expect positive correlations between the pre problem and its self-diagnosis, between the pre and post problems, and between the self-diagnosis and post problems.

For the control group (A'), we expect to obtain positive correlations between the pre and the post problem, since there will be no intervention aimed at reducing the gaps between the low and high achievers. Our within-group correlation expectations are summarized in Table VI.

We also look at the possibility of a "superficial" intervention, i.e., one in which the diagnosis is not meaningful; students do not firm up their knowledge by doing the diagnosis and their achievement on transfer problems is not affected. This might happen if students, despite realizing that relevant principles or concepts were not invoked or applied correctly during the self-diagnosis, do not make an

<sup>2</sup>Research shows that successful problem solvers provide more and better self-explanations when studying worked-out examples. Accordingly, we expect high-performing students to generate better self-diagnoses under conditions of minimal intervention. The term "meaningful" is used to suggest greater benefit.

TABLE VI. Expected correlations within groups regarding the physics principles aspect of the physics section of the rubric (NA indicates an item being not applicable; NS, nonsignificant difference; NS\*, nonsignificant difference due to low achievers reducing the gap between themselves and high achievers).

	Control	Weak	Intervention	
			Superficial	Meaningful
Pre versus self-diagnosis correlation	NA	Positive	NS ( $P > 0.05$ )	NS* ( $P > 0.05$ )
Self-diagnosis versus post correlation	NA	Positive	NS ( $P > 0.05$ )	Positive or NS
Pre versus post correlation	Positive	Positive	Positive	NS* ( $P > 0.05$ )
Difference in average gain (post versus pre) between intervention and control groups		NS	NS	Significantly higher

effort to repair their knowledge structure and understand what was wrong with the way the principles or concepts were applied. If students do not provide such explanations, the self-diagnosis grade in itself does not indicate whether the diagnosis was accompanied by a superficial or a meaningful learning process. In the case of a superficial intervention, we expect insignificant correlation between initial performance and self-diagnosis on the pre problem as well as between the self-diagnosis and the post problem, since this self-diagnosis does not allow for transfer to occur and manifest itself in students' performance on the post problem. Accordingly, we expect correlations between the pre and the post problem to be positive in this case, since the situation is the same as if the students have not engaged in self-diagnosis themselves.

Another possibility is that simply engaging in self-diagnosis, even if the student does not articulate their self-diagnosis very well, is enough to be helpful to the student; thus, the intervention is meaningful. In this case we expect insignificant correlations between initial performance on the pre problem and the self-diagnosis, between the self-diagnosis and the post problem, and between the pre and post problem. If the intervention was indeed meaningful, we expect that the average gain from pre to post in the intervention group would be significantly higher than in the control group.

## VI. FINDINGS

In describing our findings, we first show the achievements in the physics principles part: in the pre (quiz 6) self-diagnosis and then in the post (transfer) problem at midterm. We compare the groups' averages and then the correlations within the groups. We then follow the same procedure in describing the Presentation results.

### A. Physics principles aspect—Group averages

#### 1. Performance on the pre problem

The basis for the self-diagnosis is the student's own solution, so if a student had solved the problem correctly, self-diagnosis of mistakes would not be required. However, as shown in Table VII, the mean physics grades for the pre-problem in all intervention groups were low. Evidently, the problem was sufficiently challenging that all of these students made some mistakes.

#### 2. Performance on the self-diagnosis done on the pre problem

The self-diagnosis grades for the physics principles part are presented in Table VIII. For the physics self-diagnosis part, analysis of variance (ANOVA) revealed significant differences between all the groups ( $P < 0.0001$ ): group B did the best, while group D did the worst (Tables VIII and IX). This result is in line with our hypothesis, i.e., the greater the external support the better the self-diagnosis, as group B received the maximum support. The teaching assistant's outline provided this group with a solution to the problem while still requiring thought by the students in examining the details. Moreover, the self-diagnosis rubric given to students in this group provided structure to help them understand what was needed. Group D, in which the intervention simulated the common diagnostic context in which students refer to the answer in the back of the book and search their books and notes, received the least support.

Table IX presents a more detailed analysis of students' performance of the self-diagnosis of the pre problem. The table shows how well students were able to diagnose deficiencies they had made both in invoking the correct physics principles and in applying them. We focus on the

TABLE VII. Average physics grades (normalized to 1) for pre solutions for the different groups.

	Control group A'	Intervention group B Solution outline, rubric	Intervention group C Worked-out example	Intervention Group D Notes and text books
Mean	0.33	0.37	0.36	0.45
Standard error	0.04	0.04	0.04	0.04



TABLE VIII. Average physics grades (normalized to 1) for self-diagnosis of the pre quiz 7 solutions.

	Group B	Group C	Group D
Mean	0.74	0.56	0.22
Standard error	0.05	0.06	0.05

most common deficiencies in both of these subcategories. In particular, many students overlooked conservation of energy and only addressed the application of Newton's second law (see Table IV, which records the principles needed to solve the problem).

In Table IX, "Total" refers to the overall percentages of students who had mistakes in their pre problem in a given subcategory: + refers to those overall percentages of students who made a correct diagnosis (e.g., 50% of the 39% of students in group B who made mistakes in invoking the physics principle self-diagnosed it correctly); likewise, +/- refers to the percentages who made a partially correct diagnosis; and - refers to the percentages who made no diagnosis. Most of the students who had not invoked conservation of energy in their solution were able to self-diagnose this deficiency, either completely (about half in each of groups B and C and one-third in group D) or partially (about half). Those students who invoked Newton's second law but made mistakes in applying it did worse: in groups B, C, and D, respectively, no more than 25%, 20%, and 4% of those who made mistakes in application were able to self-diagnose them completely.

In Figs. 6–8 we present the self-diagnoses made by students from groups B, D, and C, respectively.

The self-diagnoses of the students in Figs. 7(b) and 8 are circled with dots. These cases were chosen as they represent the common mistakes. Figure 6 (shown earlier, in Sec. IV B) and Figs. 7 and 8 demonstrate the differences in self-diagnosis between the different intervention groups. The self-diagnosis made by the student in group B (Fig. 6) is much more detailed, especially in mentioning missing things in the presentation of the solution, such as not having a free-body diagram, not checking the answer, and not finding the correct target variable. None of this is mentioned in the self-diagnosis attempts of the students from group C [Fig. 7(b)] or D (Fig. 8), i.e., from groups that had not been provided with a self-diagnosis rubric.

We can also see the differences between diagnosing errors in "invoking" and in "applying" of physics principles. None of the three students invoked EC, but they all diagnosed this deficiency. On the other hand, the student in group D applied Newton's second law incorrectly and did not diagnose her deficient application of this law, in contrast to the student from group C, who did diagnose a similar application deficiency.

### 3. Performance on the post problem

We now address the effect of students' self-diagnosis on solving the post problems in the different intervention groups. According to our hypothesis, group B, which received the most support, will self-diagnose best and will then do best on the post problem, while group D, who received the least support, will achieve the least in the self-diagnosis and hence on the post problem. The actual achievements in the post problem are recorded in Table X, which presents the physics grades for all groups at the midterm. The table shows that the physics grades for all groups were poor, and ANOVA showed that groups did not differ significantly from each other ( $P > 0.05$ ).

### B. Physics principles aspect—Correlations within groups

Next, we consider our hypothesis regarding the extent to which the interventions are meaningful. Table XI presents the correlations between pre and post performance and between self-diagnosis and post performance.

In line with our hypothesis (see expected correlations within groups presented in Table VI), the findings shown in Table XI could not be matched to the weak, the superficial, or the meaningful intervention (Table XII). Thus, as shown in Table XII, we found a correlation between pre and post performances in the control group A', but not in the intervention groups B, C, or D (Table XII). These results suggest that the interventions were meaningful, since the distribution in students' grades in the pre problem was not carried over to the grades on the post problem.

If the intervention was indeed meaningful, the nonsignificant correlation in the intervention groups can be assumed to be due to low-achieving students performing relatively better on the post than on the pre problem. To examine this hypothesis we compared the performance of

TABLE IX. Analysis of self-diagnosis subcategories.

Group	Subcategory							
	Invoking energy conservation				Applying nonequilibrium applications of Newton's 2nd law			
Diagnosis performance	+	+/-	-	Total	+	+/-	-	Total
Group B	6 (50%)	6 (50%)	0	12/31 (39%)	6 (25%)	17 (71%)	1(4%)	24/31 (77%)
Group C	11 (52%)	9 (42%)	1 (5%)	21/28 (75%)	4 (20%)	13 (65%)	3 (15%)	20/28 (70%)
Group D	3 (30%)	4 (40%)	3 (30%)	10/25 (40%)	1 (4%)	14 (56%)	10 (40%)	25/25 (100%)

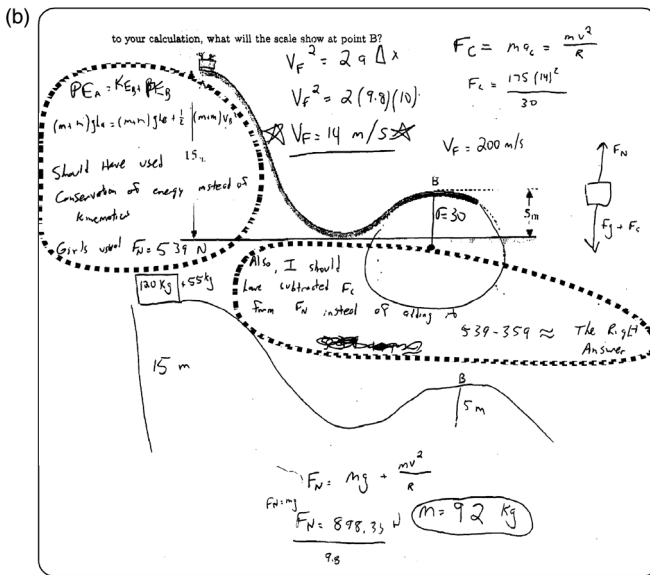
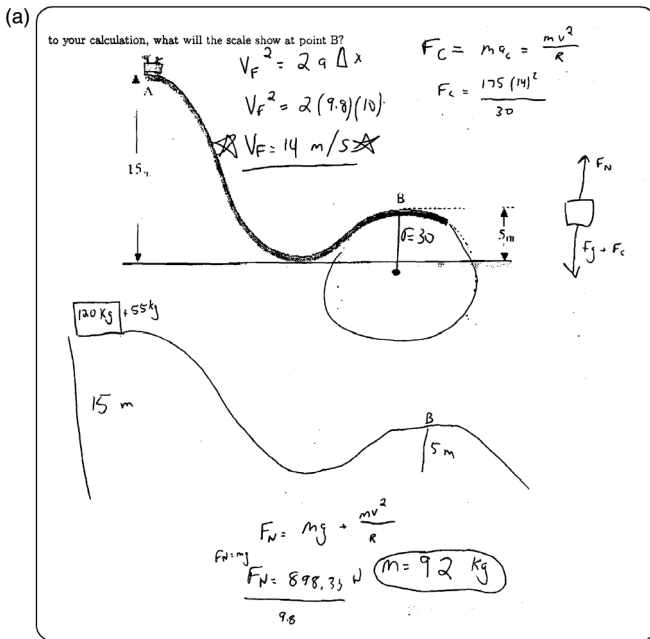


FIG. 7. (a) Pre problem solution by a student in group C. Mistakes: (1) the masses of the girl and the car are taken together as the relevant mass of the system, (2) use of kinematics instead of EC, (3) Newton’s 2nd law is incorrectly applied (wrong signs), (4) the target variable is defined as a new mass, which is yielded from  $F_N = mg$ , and (5) the student does not justify his choices. (b) Self-diagnosis attempt by the same student. References to mistakes cited in (a) are (2) the EC, “should have used EC instead of kinematics,” and shows the correct formula for EC, and (3) the mistake in signs is addressed, “I should have subtracted  $F_C$  from  $F_N$  instead of adding.” The wrong formula  $F_N = mg$  is not mentioned.

students whose pre grade was above the median of the whole group to the performance of students whose pre grade was below the median. We found that in group A the average performance of students whose pre grade was

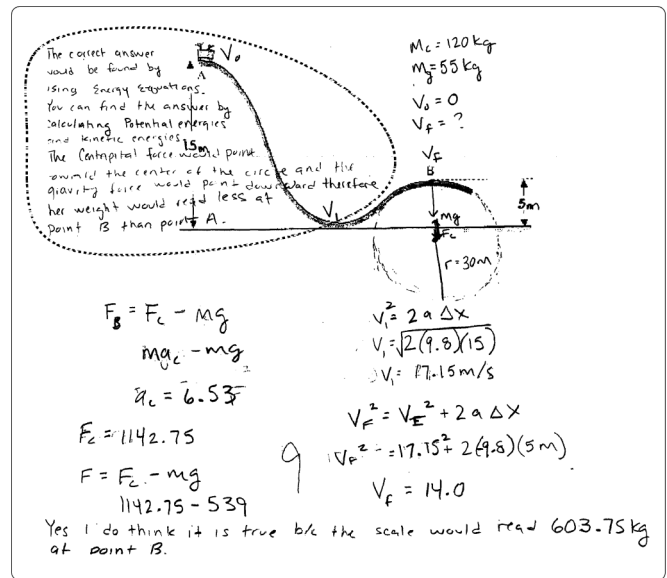


FIG. 8. Pre problem solution and self-diagnosis by a student in group D. Mistakes are (a) use of kinematics instead of energy conservation (EC), (b) wrong signs in Newton’s 2nd law. In the self-diagnosis, the student (a) refers to the fact that she had used kinematics instead of EC, and describes in words how EC should have been applied.

above the median (0.59) was indeed significantly higher than that of the students whose pre grade was below the median (0.37) ( $P < 0.01$ ). For all other groups this difference between pre and post achievement was not significant; i.e., the low achievers had managed to reduce the gap between them and the better students.

Nevertheless, ANOVA showed that the differences in average gain between the four groups were not significant. Also not significant was the correlation between the average self-diagnosis grade and the average post grade in each group, suggesting that the intervention was only a superficial one, since it did not help those who self-diagnosed better to succeed better on the post problem.

The problem can be demonstrated visually by examining the work of two students. Both achieved relatively good grades in their self-diagnosis but had different results on the post problem. Figure 8 showed an example of a student whose grade was poor on the pre problem. Figure 9 seems to show that the student did not learn from his self-diagnosis; he performed poorly on the post problem even though his self-diagnosis performance had been relatively good.

TABLE X. Average physics grades (normalized to 1) for post solutions for the different groups.

	Group A'	Group B	Group C	Group D
Mean	0.46	0.53	0.35	0.51
Standard error	0.04	0.06	0.05	0.05

TABLE XI. Correlations between pre and self-diagnosis scores, between self-diagnosis and post scores, and between pre and post scores—physics aspect.

Pre versus self-diagnosis	Correlation	Not applicable	-0.04	0.06	-0.38
	<i>P</i> value		0.83	0.77	0.06
Self-diagnosis versus post	Correlation	Not applicable	0.35	0.14	0.11
	<i>P</i> value		0.16	0.55	0.71
Pre versus post	Correlation	0.41	0.14	0.25	0.16
	<i>P</i> value	0.03	0.57	0.26	0.61

TABLE XII. Conflict between actual and expected correlations. (NA indicates an item being not applicable; NS, nonsignificant difference; NS\*, nonsignificant difference due to low achievers reducing the gap between themselves and high achievers.) The checkmark stands for expected correlations that were confirmed by the analysis, and the X stands for expected correlations that were not confirmed by the analysis.

	Control	Weak	Intervention Superficial	Meaningful
Pre versus self-diagnosis correlation	NA	X positive	√ NS ( <i>P</i> > 0.05)	√ NS* ( <i>P</i> > 0.05)
Self-diagnosis versus post correlation	NA	X positive	√ NS ( <i>P</i> > 0.05)	X Positive or √ NS
Pre versus post correlation	√ positive	X positive	X positive	√ NS* ( <i>P</i> > 0.05)
Difference in average gain (post versus pre) between intervention and control groups	NA	√ NS	√ NS	X Significantly higher

A different picture is presented in Fig. 10, which shows an example of a student who seems to have learned from his self-diagnosis. Whereas his grade on the pre problem solution had been poor [Fig. 10(a)], he had performed relatively well on the self-diagnosis [Fig. 10(b)] and performed well on the post problem solution [Fig. 10(c)]. Thus, the two students whose work is shown in Figs. 9 and 10 had obtained poor grades in the pre problem and relatively good grades in their self-diagnoses, but different grades in their post problem scores at midterm. One possible explanation for this result is that the self-diagnosis grade does not reflect how focused the students were on learning when they had diagnosed their solutions. Thus, a

similar self-diagnosis grade could represent not only students who had gone through a meaningful learning process in which they explained their mistakes to themselves, but also those who had gone through a superficial learning process in which they had merely identified their mistakes without thinking about what caused them.

**C. Presentation aspect**

Presentation grades for the pre (quiz 6 problem) performance, as shown in Table XIII, are quite similar across different groups.

The self-diagnosis grades for presentation are presented in Table XIV. ANOVA shows that these grades are about the same for groups C and D and significantly better for group B (*P* < 0.0001). This may be explained by the support that group B had received, since the rubric they needed to fill out in order to diagnose had specifically referred to presentation issues. Thus, modeling of a well-presented solution had not sufficed to make students aware of deficiencies in presentation of the solution. Only the explicit requirement in a rubric (given to group B during self-diagnosis) had enabled students to recognize such deficiencies. This means that students did not pay attention to the presentation in the diagnosis unless instructed to do so, as with group B. For that reason, even those students who did poorly in the presentation had done even worse in realizing that they had done poorly.

Table XV records the mean presentation grades for all groups on the post problem. ANOVA showed that differences between the groups were not significant (*P* > 0.05). Intervention had no effect on the presentation performance, even though group B had done better on the self-diagnosis

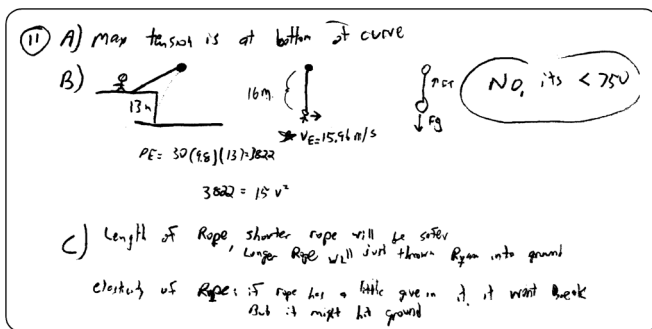
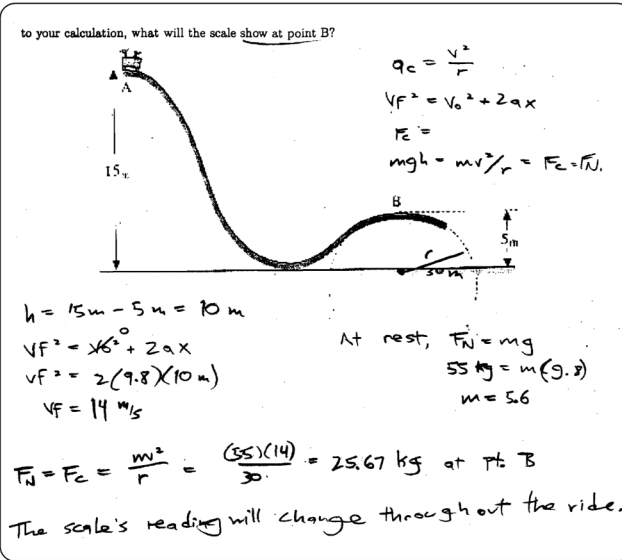


FIG. 9. Post problem solution by a student from group C. Mistakes: (a) Newton’s 2nd law, though invoked, is incorrectly applied, (b) incorrect height is used in establishing PE, while on the quiz he did not invoke EC. In his self-diagnosis the student had referred to the lack of EC—“I should have used EC instead of kinematics”—and had showed the correct formula for EC. He had also, in his self-diagnosis, corrected the wrong application of Newton’s 2nd law—“I should have subtracted  $F_C$  from  $F_N$ .”

(a) to your calculation, what will the scale show at point B?



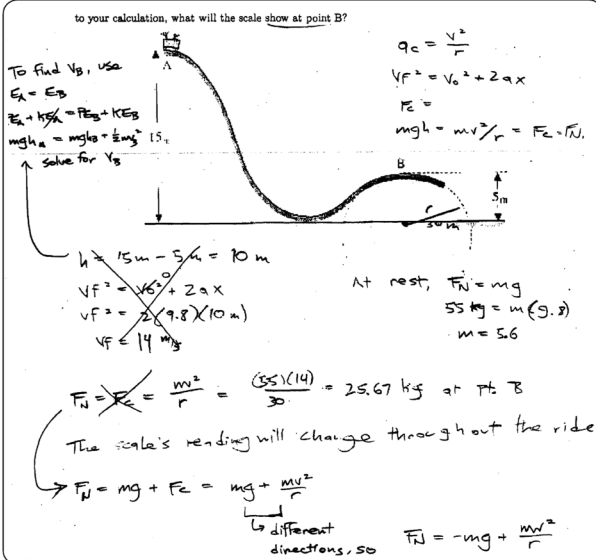
$q_c = \frac{v^2}{r}$   
 $v_f^2 = v_0^2 + 2ax$   
 $F_c =$   
 $mgh = mv^2/r = F_c = F_N$

$h = 15m - 5m = 10m$   
 $v_f^2 = v_0^2 + 2ax$   
 $v_f^2 = 2(9.8)(10m)$   
 $v_f = 14 m/s$

At rest,  $F_N = mg$   
 $55 kg = m(9.8)$   
 $m = 56$

$F_N = F_c = \frac{mv^2}{r} = \frac{(55)(14)}{30} = 25.67 \text{ kg at Pt B}$   
 The scale's reading will change through out the ride.

(b) to your calculation, what will the scale show at point B?



To find  $v_b$ , use  $E_A = E_B$   
 $E_A + K_A = E_B + K_B$   
 $mgh_A = mgh_B + \frac{1}{2}mv^2$

$q_c = \frac{v^2}{r}$   
 $v_f^2 = v_0^2 + 2ax$   
 $F_c =$   
 $mgh = mv^2/r = F_c = F_N$

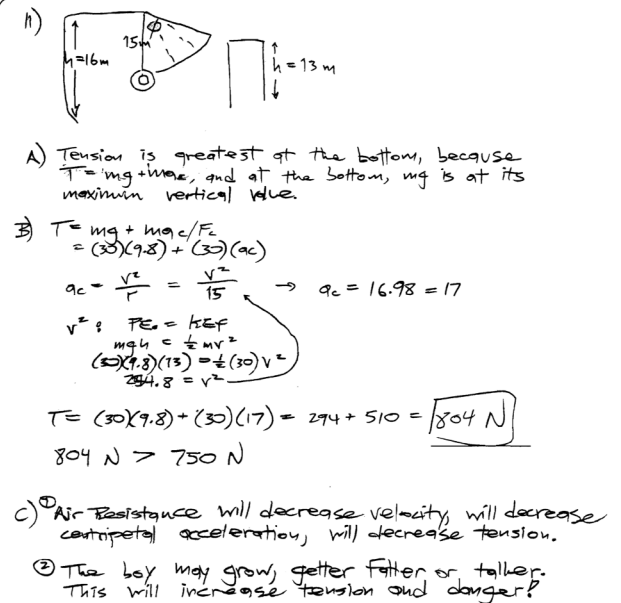
$h = 15m - 5m = 10m$   
 $v_f^2 = v_0^2 + 2ax$   
 $v_f^2 = 2(9.8)(10m)$   
 $v_f = 14 m/s$

At rest,  $F_N = mg$   
 $55 kg = m(9.8)$   
 $m = 56$

$F_N = F_c = \frac{mv^2}{r} = \frac{(55)(14)}{30} = 25.67 \text{ kg at Pt B}$   
 The scale's reading will change through out the ride.

$F_N = mg + F_c = mg + \frac{mv^2}{r}$   
 different directions, so  $F_N = -mg + \frac{mv^2}{r}$

(c)



$h = 16m$   
 $15m$   
 $h = 13m$

A) Tension is greatest at the bottom, because  $T = mg + ma_c$ , and at the bottom,  $mg$  is at its maximum vertical value.

B)  $T = mg + ma_c/F_c$   
 $= (30)(9.8) + (30)(a_c)$   
 $a_c = \frac{v^2}{r} = \frac{v^2}{15} \rightarrow a_c = 16.98 = 17$   
 $v^2$ :  $F_c = KcF$   
 $mgh = \frac{1}{2}mv^2$   
 $(30)(9.8)(13) = \frac{1}{2}(30)v^2$   
 $294.8 = v^2$

$T = (30)(9.8) + (30)(17) = 294 + 510 = 804 \text{ N}$   
 $804 \text{ N} > 750 \text{ N}$

C) ① Air Resistance will decrease velocity, will decrease centripetal acceleration, will decrease tension.  
 ② The boy may grow, get fatter or taller. This will increase tension and danger!

FIG. 10. (a) Pre problem attempt by another student from group C. Mistakes: (1) kinematics are used instead of energy conservation, (2)  $F_N$  is equated to  $F_C$ , (3) the target variable is identified as a new mass yielded from  $F_N = mg$ . (b) Self-diagnosis attempt by the same student. References to mistakes cited in (a) are (1) refers to the problem in invoking EC, “To find  $v_b$ , use  $E_A = E_B$ ,” and shows how to use EC, (2) corrects the incorrect application of Newton’s 2nd law and writes the correct description of the law, (3) identifies the incorrect target variable and isolates  $F_N$  as the correct one. Note that the previous error of finding a new mass is not erased. (c) This student’s attempt on the midterm (post) problem. The wrong height is used in addressing the potential energy of the system, but there is an improvement relative to the errors described in (a), as follows: (1) energy conservation is used correctly, (2) Newton’s 2nd law is addressed more properly, and (3) the target variable is properly identified. Here, the force of tension on the rope is analogous to the normal force in the quiz problem.

exercise. We suspect that presentation was something that the students did not consider important; therefore, their self-diagnosis of deficiencies in it had not affected their achievements in the post problem.

Table XVI presents the correlations within groups for presentation. There are no correlations between the post problem grades and self-diagnosis grades for all intervention groups.

Since group B students fared as poorly as the other groups with regard to the average presentation grade on the post problem despite their better performance on presentation self-diagnosis, we must consider the possibility that this was attributable to the “one-time” nature of the intervention. Presentation guidelines were provided at the beginning of the experiment and were not repeated in other instructional contexts. Given the time constraints of teaching



TABLE XIII. Average group presentation grades for pre problem solution.

	Group A'	Group B	Group C	Group D
Mean	0.33	0.40	0.37	0.35
Standard error	0.02	0.03	0.02	0.02

TABLE XIV. Average group presentation grades for self-diagnosis of the pre problem.

	Group B	Group C	Group D
Mean	0.42	0.10	0.12
Standard error	0.03	0.02	0.02

TABLE XV. Average group presentation grades for post problem.

	Group A'	Group B	Group C	Group D
Mean	0.42	0.44	0.42	0.47
Standard error	0.02	0.02	0.03	0.03

assistants in a large college classroom, students did not receive good feedback on their performance in the diagnosis. Moreover, the course was not directed in other ways towards teaching students to provide well-presented reasoning in their solutions. It would be of interest to examine whether applying the intervention consistently throughout the semester would help develop presentation skills.

## VII. DISCUSSION

In this study we compared the learning processes and outcomes in four groups of students. Three were intervention groups, in which the students carried out self-diagnosis tasks on their solutions after attempting to solve a physics problem. The interventions differed in the resources and instructions that the students received. Students in a control group carried out no self-diagnosis tasks after attempting to solve the problem but were presented with worked-out example and common mistakes by the teaching assistant. The focus in this study was on the context of an atypical problem situation in an algebra-based introductory physics course dealing with Newton's second law in a vertical circular motion.

The first research question was, *How well do students self-diagnose their solutions in the different interventions?* We found that self-diagnosis performances tended to be incomplete. Students were better at realizing deficiencies in their solutions with regard to *invoking* rather than to *applying* the physics principles. Students who had failed to invoke some principles could easily diagnose such deficiencies. In contrast, students had trouble diagnosing their deficiencies in applying the physics principles. A comparison of the average self-diagnosis grades of the different groups shows that providing students with a self-diagnosis rubric or a detailed model solution improved their self-diagnosis of the physics underlying the solution.

Students' performances on diagnosing the presentation of the solution tended to be poor. The worked-out example emphasized its presentation according to guidelines aligned with problem-solving strategies (Fig. 1). Yet, even the modeling of a well-presented solution (given to group C) did not suffice to bring those students to recognize deficiencies in their presentations. Only the explicit requirement specified by the self-diagnosis rubric given to group B made those students acknowledge their deficiencies to some extent, mainly in the planning and checking subcategories.

The second research question was, *What is the effect of students' self-diagnosis on subsequent problem solving in the different intervention groups?* We found that, whereas the external support made a difference to the self-diagnosis, it was not correlated with performance on the post problem. In particular, students who scored higher on self-diagnosis did not score higher on the post problem. Moreover, there was no within-group correlation between the self-diagnosis physics grades and the midterm physics grades for any of the groups.

The post problem was designed as a transfer problem in that it was similar in general procedure to the pre problem, but differed in detailed procedures. However, in agreement with the findings of an independent categorization study [44], a majority of students at the introductory level did not associate the two problems as belonging to the same category. Thus, students did not recognize the analogous general procedure underlying both problems. The interventions provided students with an opportunity to study the quiz problem deeply. Thus, they could recognize the underlying general procedure needed to solve the problem and later transfer what they had learned in studying the pre

TABLE XVI. Within-group correlations—Presentation aspect.

		Group A'	Group B	Group C	Group D
Pre versus self-diagnosis	Correlation	Not applicable	0.28	0.32	-0.02
	<i>P</i> value		0.13	0.09	0.94
Self-diagnosis versus post	Correlation	Not applicable	-0.21	-0.16	0.08
	<i>P</i> value		0.41	0.48	0.79
Pre versus post	Correlation	0.05	0.14	0.07	0.38
	<i>P</i> value	0.79	0.57	0.76	0.20

problem to the post problem. Our finding that students' self-diagnosis performance did not correlate with their performance on the post problem suggests that this intervention did not suffice to make students recognize the underlying general procedure. This conclusion is strengthened by the fact that ANOVA revealed nonsignificant differences in the average gains between the four groups, meaning that the average gains in the three intervention groups were not significantly higher than that of the control group.

However, whereas in the control group we found a correlation between the average physics grades for the pre and the post problems, no such correlation was found in the intervention groups. This implies that the intervention might have helped the lower-achieving students to reduce the gap between themselves and the better students.

One possible explanation for this lack of consistency in the findings is that the analysis rubric we devised for grading students' self-diagnosis performance did not differentiate between a meaningful self-diagnosis, which would lead to self-repair of students' mental models, and a superficial self-diagnosis, which would not. For a self-repair [28] process to take place, students need to realize differences between their solution and the instructor's solution that are of relevance for finding the right solution, and must also acknowledge conflicts between the beliefs underlying their solution and the scientific model underlying the worked example. If students realize, while engaged in self-diagnosis, that certain principles or concepts were not invoked or applied correctly, but do not provide further explanation regarding the nature of the conflict between their thinking and the scientific model underlying the instructor's solution, then the researcher cannot tell from their self-diagnosis grade if the learning process that accompanied the diagnosis was meaningful or superficial.

Thus, the better average grades for self-diagnosis achieved by groups B and C than those obtained by the others might be explained by the ease with which students from groups B and C were able to compare the instructor's

solution with their solution and realize that they differed in the principles that should have been invoked. Group D students were unable to do so since solved problems that would be similar to those in the quiz, i.e., the pre problem, did not appear in their textbooks.

There are two ways in which the analysis of students' self-diagnosis might be improved.

- (1) Change the diagnosis instructions to better explicate the nature of the conflict between the student's solution and the scientific model. For example, include generalization instructions such as "Learn from your mistakes by generalizing beyond the specific problem" [33], or ask students to devise "warning signs" that will prevent them from repeating similar mistakes in the future [45].
- (2) Change the self-diagnosis rubric so that students not only recognize differences between their solutions and the instructor's, but also go beyond that to acknowledge conflicts.

Our results suggest there is a need for further study of self-diagnosis tasks. In a companion paper (Ref. [46]) we will compare students' self-diagnosis performance in a typical and an atypical problem situation. The findings of such a study would be expected to shed light on how students' performances in self-diagnosis tasks depend on their prior knowledge.

## ACKNOWLEDGMENTS

We wish to thank Professor Jeremy Levy, the physics faculty member who gave his valuable class time to participate in this study. We thank Yetty Varon, the statistician of the Science Teaching Department at The Weizmann Institute of Science, for her help in analyzing the data. We appreciate the support of the University of Pittsburgh and The Weizmann Institute of Science, Department of Science Teaching. This work was primarily supported by Israel Science Foundation Grant No. 1283/05 and National Science Foundation Grant No. DUE-0442087.

- 
- [1] P. Black and D. Wiliam, Inside the black box: Raising standards through classroom assessment, *Phi Delta Kappan* **80**, 139 (1998).
  - [2] P. Black and D. Wiliam, Assessment and classroom learning, *Assess. Educ. Princ. Pol. Pract.* **5**, 7 (1998).
  - [3] P. Black and C. Harrison, Feedback in questioning and marking: The science teacher's role in formative assessment, *Sch. Sci. Rev.* **82**, 55 (2001).
  - [4] C. Bereiter and M. Scardamalia, in *Knowing, Learning, and Instruction: Essays in Honor of Robert Glaser* (Lawrence Erlbaum, Hillsdale, NJ, 1989), pp. 361–392.
  - [5] M. Chi, M. Bassok, M. Lewis, P. Reimann, and R. Glaser, Self-explanations: How students study and use examples in learning to solve problems, *Cogn. Sci.* **13**, 145 (1989).
  - [6] T. Van Gog, F. Paas, and J.J.G. van Merriënboer, Effects of studying sequences of process-oriented and product-oriented worked examples on troubleshooting transfer efficiency, *Learn. Instr.* **18**, 211 (2008).
  - [7] G. Polya, in *How to Solve It* (Princeton University Press, Princeton, NJ, 1945).
  - [8] F. Reif, Millikan Lecture 1994: Understanding and teaching important scientific thought processes, *Am. J. Phys.* **63**, 17 (1995).

- [9] 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).
- [10] P. Heller and M. Hollabaugh, Teaching problem solving through cooperative grouping. Part 2: Designing problems and structuring groups, *Am. J. Phys.* **60**, 637 (1992).
- [11] C. S. Große and A. Renkl, Finding and fixing errors in worked examples: Can this foster learning outcomes?, *Learn. Instr.* **17**, 612 (2007).
- [12] E. Yerushalmi, C. Henderson, K. Heller, P. Heller, and V. Kuo, Physics faculty beliefs and values about the teaching and learning of problem solving. I. Mapping the common core, *Phys. Rev. ST Phys. Educ. Res.* **3**, 020109 (2007).
- [13] J. Larkin, Processing information for effective problem solving, *J. Eng. Educ.* **70**, 285 (1979).
- [14] J. Larkin and F. Reif, Understanding and teaching problem solving in physics, *Eur. J. Sci. Educ.* **1**, 191 (1979).
- [15] J. Larkin, in *Cognitive Skills and Their Acquisition*, edited by J. Anderson (Lawrence Erlbaum Associates, Hillsdale, NJ, 1980), pp. 311–334.
- [16] D. Maloney, in *Handbook of Research on the Teaching and Learning of Science*, edited by D. Gabel (MacMillan, New York, 1994).
- [17] J. Larkin, J. McDermott, D. Simon, and H. Simon, Expert and novice performance in solving physics problems, *Science* **208**, 1335 (1980).
- [18] A. H. Schoenfeld, in *Handbook for Research on Mathematics Teaching and Learning*, edited by D. Grouws (MacMillan, New York, 1992), pp. 334–370.
- [19] F. Reif and J. Heller, Knowledge structure and problem solving in physics, *Educ. Psychol.* **17**, 102 (1982).
- [20] E. F. Redish, *Teaching Physics with the Physics Suite* (Wiley, Hoboken, NJ, 2003).
- [21] A. Van Heuvelen, Learning to think like a physicist: A review of research-based instructional strategies, *Am. J. Phys.* **59**, 891 (1991).
- [22] W. J. Leonard, R. J. Dufresne, and J. P. Mestre, *Am. J. Phys.* **64**, 1495 (1996).
- [23] A. Collins, J. Brown, and S. Newman, in *Knowing, Learning, and Instruction: Essays in Honor of Robert Glaser* (Ref. [4]), pp. 453–494.
- [24] F. Reif and L. A. Scott, Teaching scientific thinking skills: Students and computers coaching each other, *Am. J. Phys.* **67**, 819 (1999).
- [25] K. VanLehn, C. Lynch, L. Taylor, A. Weinstein, R. Shelby, K. Schulze, D. Treacy, and M. Wintersgill, Minimally invasive tutoring of complex physics problem solving, in *Proceedings of Intelligent Tutoring Systems: 6th International Conference, ITS 2002, Biarritz, France and Sebastián, Spain*, edited by S. A. Cerri, G. Gouarderes, and F. Paraguacu (Springer, Berlin, 2002), pp. 367–376.
- [26] K. Cummings, J. Marx, R. Thornton, and D. Kuhl, Evaluating innovation in studio physics, *Am. J. Phys.* **67**, S38 (1999).
- [27] T. Foster, Ph.D. thesis, University of Minnesota, 2000 (unpublished).
- [28] M. Chi, in *Advances in Instructional Psychology*, edited by R. Glaser (Lawrence Erlbaum Associates, Hillsdale, NJ, 2000), pp. 161–238.
- [29] M. Chi, N. de Leeuw, M. Chiu, and C. LaVancher, Eliciting self-explanations improves understanding, *Cogn. Sci.* **18**, 439 (1994).
- [30] K. Bielaczyc, P. Pirolli, and A. Brown, Training in self-explanation and self-regulation strategies: Investigating the effects of knowledge acquisition activities on problem solving, *Cogn. Instr.* **13**, 221 (1995).
- [31] V. Aleven, K. R. Koedinger, and K. Cross, Artificial intelligence in education, open learning environments: New computational technologies to support learning, exploration, and collaboration, in *Proceedings of the 9th International Conference on Artificial Intelligence in Education, Le Mans, France*, edited by S. P. Lajoie and M. Vivet (IOS Press, Amsterdam, 1999), pp. 199–206.
- [32] C. Henderson and K. A. Harper, Quiz corrections: Improving learning by encouraging students to reflect on their mistakes, *Phys. Teach.* **47**, 581 (2009).
- [33] E. Etkina, A. Van Heuvelen, S. White-Brahmia, D. T. Brookes, M. Gentile, S. Murthy, D. Rosengrant, and A. Warren, Scientific abilities and their assessment, *Phys. Rev. ST Phys. Educ. Res.* **2**, 020103 (2006).
- [34] D. Perkins and R. Swartz, in *If Minds Matter: A Foreword to the Future*, edited by A. L. Costa, J. Beflanca, and R. Fogarty (Skylight Publishing, Palatine, IL, 1992), Vol. 2, pp. 53–69.
- [35] B. S. Bloom, M. D. Engelhart, E. J. Furst, W. H. Hill, and D. R. Krathwohl, *Taxonomy of Educational Objectives: The Classification of Educational Goals; Handbook I: Cognitive Domain* (Longmans, New York, 1956).
- [36] M. L. Gick and K. J. Holyoak, Schema induction and analogical transfer, *Cogn. Psychol.* **15**, 1 (1983).
- [37] B. Eylon and J. Helfman, in *Science Teaching in Israel: Origins, Development, and Achievements*, edited by A. M. Mayer and P. Tamir (Israel Science Teaching Center, Jerusalem, 1984), pp. 259–271 (in Hebrew).
- [38] Y. Neuman and B. Schwarz, Is self-explanation while solving problems helpful? The case of analogical problem-solving, *Br. J. Educ. Psychol.* **68**, 15 (1998).
- [39] M. Chi, P. J. Feltovich, and R. Glaser, Categorization and representation of physics problems by experts and novices, *Cogn. Sci.* **5**, 121 (1981).
- [40] B. Eylon and F. Reif, Effects of knowledge organization on task performance, *Cogn. Instr.* **1**, 5 (1984).
- [41] E. Yerushalmi and B. Eylon, Teachers' approaches to promoting self-monitoring in physics problem solving by their students, in *Proceedings of the International GIREP Conference on Physics Teacher Education Beyond 2000: Selected Contributions, Barcelona, 2000*, edited by R. Pinto and S. Surrinach (Elsevier, Paris, 2001).
- [42] M. Chi, Quantifying qualitative analyses of verbal data: A practical guide, *J. Learn. Sci.* **6**, 271 (1997).
- [43] J. Docktor and K. Heller, Robust assessment instrument for student problem solving, in *Proceedings of the 82nd NARST Annual International Conference, 2009* (NARST, Garden Grove, CA, 2009).
- [44] C. Singh, Categorization of problems to assess and improve proficiency as teachers and learners, *Am. J. Phys.* **77**, 73 (2009).
- [45] P. Labudde, F. Reif, and L. Quinn, Facilitation of scientific concept learning by interpretation procedures and diagnosis, *Int. J. Sci. Educ.* **10**, 81 (1988).
- [46] E. Yerushalmi, E. Cohen, A. Mason, and C. Singh, following paper, What do students do when asked to diagnose their mistakes? Does it help them? II. A more typical quiz context, *Phys. Rev. ST Phys. Educ. Res.* **8**, 020110 (2012).