PHYSICAL REVIEW SPECIAL TOPICS - PHYSICS EDUCATION RESEARCH **8**, 020110 (2012)

What do students do when asked to diagnose their mistakes? Does it help them? II. A more typical 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 11 January 2012; published 21 September 2012)

"Self-diagnosis tasks" aim at fostering students' learning in an examination context by requiring students to present diagnoses of their solutions to quiz problems. We examined the relationship between students' learning from self-diagnosis and the typicality of the problem situation. Four recitation groups in an introductory physics class (~ 200 students) were divided into a control group and three intervention groups in which different levels of guidance were provided to aid students in their performance of self-diagnosis activities. The self-diagnosis task was administered twice, first in an atypical problem situation and then in a typical one. In a companion paper we reported our findings in the context of an atypical problem situation. Here we report our findings in the context of a typical problem situation and discuss the effect of problem typicality on students' self-diagnosis performance and subsequent success in solving transfer problems. We show that the self-diagnosis score was correlated with subsequent problem-solving performance only in the context of a typical problem situation, and only when textbooks and notebooks were the sole means of guidance available to the students for assisting them with diagnosis.

DOI: 10.1103/PhysRevSTPER.8.020110 PACS numbers: 01.40.Fk, 01.40.gb, 01.40.Ha

I. INTRODUCTION

"Self-diagnosis tasks" are formative assessment tasks [1] that aim at fostering student learning in an examination context by requiring students to present their own diagnosis (in which they identify where they went wrong and explain the nature of the mistakes) as part of the activity of reviewing their quiz solutions [2–4]. These tasks are intended to induce students to generate *self-explanations* involving *self-repair* [5], meaning that in reflecting on their solution to a problem when they self-diagnose it, students will recognize and resolve conflicts between their possibly flawed mental model and the scientifically acceptable model. Research on students' study of worked-out examples has indeed shown that students who self-explain more learn more, and that good self-explainers are those who readily detect conflicts while learning from a sample solution [6,7]. Also, research on self-reflection as recorded in weekly reports [8], in which students reflected on how they learned specific physics content, showed correlation between their conceptual gains and their ability to articulate what they had learned.

We carried out a two-part study, the second part of which is reported in this paper. In the first part of the study, which we described in detail in a companion paper [9], recitation groups in an introductory physics class were divided into three intervention conditions and a control condition. The intervention groups performed different kinds of self-diagnosis tasks, involving different levels of support during the intervention, to determine which level helped students to self-diagnose best. In the control group students were merely presented with a worked-out solution to the previous week's quiz problem. The different self-diagnosis tasks shared a similar basic structure: students were given photocopied solutions of their quiz problem (hereafter termed the pre problem) and were given time and credit for presenting a diagnosis (identifying and circling where they went wrong and explaining the nature of the mistakes). The problems used in that study were *atypical* in the sense that no isomorphic problems had been previously worked on, either in the textbook or in the recitation group. The self-diagnosis tasks varied in the external support (instructions and resources) that students received. In one intervention, the instructor first presented on the whiteboard a concise and "product-oriented" [10] outline of the solution, which skipped some details of the derivations. The instructor added more reasoning underlying the derivation orally. The students were then asked to diagnose their mistakes while also sorting them according to major

---

problem-solving steps described in the research literature [11–13] (e.g., ‘‘description,’’ ‘‘plan,’’ etc.). In a second intervention the students compared their own solution to a detailed, ‘‘process-oriented’’ [10] written solution that was not accompanied by oral explanations. In a third intervention the students were given the final answer to the problem and allowed to use their notes and textbooks to help them diagnose their mistakes. In the subsequent midterm examination the instructor gave all students a transfer problem (hereafter termed the post problem) with the aim of determining the effect of students' self-diagnosis, when carried out via different tasks, on subsequent problem solving,

The results of the first part of the study [9] showed that both the pre and the post problems were difficult for students: not a single student was awarded full credit on the quiz problem, and only a few were able to answer the post problem correctly. We found that low achievers in the intervention groups, unlike those in the control group, had managed to reduce the gap between themselves and the better students. We also found that the average self-diagnosis performance improved as the degree of external support was increased. However, the self-diagnosis performance did not correlate with students' performance in the post problem.

Bransford and Schwartz [14] postulated that transfer of knowledge from the situation in which it was acquired to new situations is optimal if the activities that students engage in include both elements requiring innovation and elements requiring efficiency. In their model, efficiency means rapid retrieval and accurate application of the appropriate knowledge and skills for solving a problem, while innovation refers to adaptive capacity, meaning the ability to rearrange one's environment and thinking to handle new types of problems. In their model, efficiency and innovation are represented on two orthogonal coordinates (shown in Fig. 1, model of Schwartz *et al.* [15]). If an intervention is focused only on efficiency, the students' cognitive engagement will be diminished and they will not develop the ability to transfer the acquired knowledge to new situations. Similarly, if the intervention is focused only on innovation, students may struggle to connect what they are learning with their prior knowledge, so that both learning and transfer will be inhibited. These authors proposed that transfer would be enhanced if the learning activity was focused on moving along a diagonal trajectory in the two-dimensional space of innovation and efficiency.

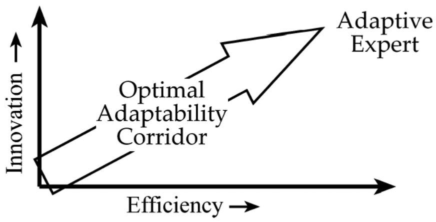


FIG. 1. Model of Schwartz *et al.* [15] for the roles of ‘‘efficiency’’ and ‘‘innovation’’ in transfer.

The self-diagnosis tasks used in both parts of our study were designed to promote *innovation*, as it provided an opportunity for students to examine their ideas and change them when necessary. According to the Bransford and Schwartz model, the poor performance on post problems in the first part of our study [9] might be a result of the intervention's main focus on *atypical* problems that required the students to be highly innovative, as they were not accustomed to these types of problems. Focusing on problems in which students had gained expertise, and from which they could thus more readily retrieve the knowledge required, would augment the *efficiency* component. Thus, interventions that placed due focus on both components would allow students to make the connection between what they were learning and their prior knowledge, so that both learning and transfer would be improved.

In the context of probability problems, researchers [16] found that studying mistaken solutions is indeed advantageous for learners with high levels of knowledge. They showed that learners with poor prior knowledge can benefit to some extent when the errors in the mistaken solution are highlighted. Another example is the expertise-reversal effect: at the initial stages of skill acquisition, learning from worked-out examples is more effective than problem solving [17,18]. When learners have acquired more expertise, such worked examples are less effective [10]. At this stage students derive more benefit if they learn from practice problems on their own, followed by isomorphic examples [19].

To establish in what way students' performance in self-diagnosis tasks in physics depends on their prior knowledge, we carried out an additional experiment [20,21], whose context this time was a *typical* problem situation. By ‘‘typical’’ problem we mean one for which isomorphic problem(s) have been previously given as homework, and for which solutions are presented by the textbook or by the teaching assistant in the recitation group. It therefore strengthens the *efficiency* component in the self-diagnosis exercise.

The additional experiment carried out in the second part of our study, and described in this paper, was based on the same guidelines as those described in the companion paper [9]. As in the first part of the study, here too we investigated the learning processes and outcomes associated with three different ‘‘self-diagnosis’’ tasks that varied with respect to the support students received. The problems we selected here, however, were more typical and demanded less innovation from the students. We then compared the results of the first part of the study, in which students' self-diagnosis had related to an atypical problem situation, to the results of the second part, where students' self-diagnosis related to a typical problem situation. In particular we ask the following:

(1) How well do students self-diagnose and correct mistakes in their solutions in different kinds of self-diagnosis tasks involving a *typical* problem?
(2) What is the effect of students' self-diagnosis in different kinds of self-diagnosis tasks on subsequent problem solving involving a *typical* problem?
(3) How does the relationship between students' self-diagnosis performance and subsequent problem solving in the case of a *typical* problem compare with that in the case of an *atypical* problem?

II. EXPERIMENTAL SETUP

As mentioned above, the purpose of the experiment in this part of our study was to compare, in the context of a typical problem situation, the effect of students' self-diagnosis in different kinds of self-diagnosis tasks on their subsequent problem solving. Recitation groups in an introductory physics class were divided into three intervention conditions and a control condition. In the control group the teaching assistant discussed in class a worked-out solution to the previous week's quiz problem (hereafter termed the pre problem). In the intervention groups the students self-diagnosed their solutions to the pre problem. The specific self-diagnosis tasks carried out in each intervention group differed from those in the other intervention groups with respect to the external support (instructions and resources) that the students received. Division of the recitation groups into intervention and control groups was the same as in the first part of the study. Thus, in the second part the various groups not only underwent different treatments, but had also undergone different past experiences. Therefore, in order to provide an overall picture, in this paper we describe their experiences in the course of both parts of the study.

A. Experimental sequence

Initial training was provided using the quiz given in week 5. Each intervention group followed an appropriately modified version of the intervention sequence. Students were given an incorrect solution of the "training problem" and attempted to self-diagnose it according to their treatment group. The teaching assistant then demonstrated how the incorrect solution to the training problem should be diagnosed.

In all groups, the 6th and 7th quizzes in the course served as pre problems. Unlike the quiz 6 problem of the first experiment, the present quiz 7 problem was a typical one. A typical problem differed from an atypical one in that in the former case isomorphic problems had been previously given as homework and the students had encountered solutions to those problems either in the textbook or in recitation class discussions.

One week after students first attempted to solve the pre problem presented in quiz 6, the students in the intervention groups self-diagnosed their quiz solutions. In the control groups, the teaching assistant discussed a worked-out solution to the quiz 6 problem in class. A similar sequence was carried out for the pre problem presented in quiz 7. On the midterms the students were given transfer problems to solve (hereafter termed post-atypical or post-typical), which were paired with the respective pre problems presented in quiz 6 or quiz 7. Because of course constraints, intervals between the self-diagnosis activities and the post tests differed with problem type: three days in the case of the atypical problem, and three weeks in the case of the typical problem.

Table I shows the sequence followed in all intervention groups.

B. Study sample

The study participants were the same as those in the first part of the study. They were drawn from an introductory algebra-based physics course for premedical students ($N \sim 200$), with one instructor and two teaching assistants. As our study sample we focused on four recitation groups, control group A' and three intervention groups (B, C, and D). Similarity of the four groups was verified by a Duncan analysis performed on students' answers to the quiz in the first part of the study [9] ($P > 0.05$). One teaching assistant worked with groups A and B, and the other worked with groups C and D. The composition of the groups remained constant (no more than one student moved between

TABLE I. Experimental sequence.

Date	Stage in study		Related course materials
Mon, week 5	Initial training		Quiz 5
Mon, week 6	1st part	Pre-atypical	Quiz 6
Mon, week 7		Treatment, atypical	Self-diagnosis of students' solution to quiz 6
Thu, week 7		Post-atypical	2nd midterm—isomorphic problem to the pre problem
Mon, week 7	2nd part	Pre-typical	Quiz 7
Mon, week 8		Treatment, typical	Self-diagnosis of students' solution to quiz 7
Thu, week 11		Post-typical	3rd midterm—isomorphic problem to the pre problem

TABLE II. Distribution of students into control and intervention groups. TA refers to the teaching assistant.

	Control A'	Intervention B	Intervention C	Intervention D
TA 1	41	31	26	23
TA 2	40	27	21	22

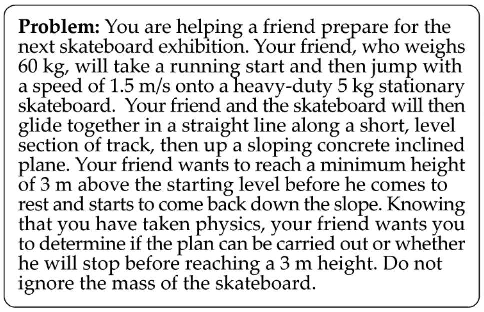
recitation groups, and students who moved were eliminated from the study). Several students who performed the pre problem (quiz 7) did not attempt the post problem. Students who carried out 75% of the tasks in the first part of the study and participated in either the training or the self-diagnosis sessions were included in the analysis in the second part of the study. Table II shows the distribution of students (by number) into control and intervention groups in quiz 7.

C. Resources and guidelines in the different groups

In all groups students first solved a realistic, motivating quiz problem. The problem used in quiz 7 is shown in Fig. 2.

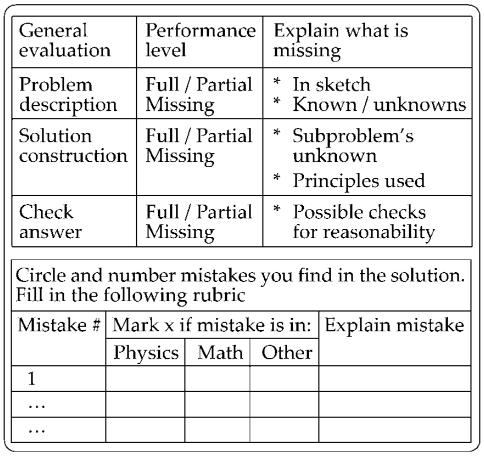
This is a typical example of a quiz problem in an algebra-based introductory physics course, as it involves energy and mass conservation. To solve it, students will need to understand that the target variable is the maximum height that the skateboard can reach, or the speed of the skateboard at the height of 3 m. To calculate maximum height, they will have to invoke energy conservation before and after the climb, which is justified because all forces doing work are conservative forces: $PE_i + KE_i = PE_f + KE_f$. They will need to find the intermediate variable, which is the speed v of the friend plus skateboard after the jump. To calculate this speed they will have to invoke the expression for momentum conservation: $mv_i = (m + M)v_f$.

In control group A' the instructor discussed the problem's solution with the students, but they were not required to engage in a self-diagnosis task. In each intervention group, during the recitation following the quiz the instructor gave the students an ungraded photocopy of their solution and asked them to diagnose their mistakes. Students were credited with 50% of their original quiz grade for completing the diagnosis. The credit was effort based. The instructor also motivated them by explaining how self-diagnosis would help them learn.



Problem: You are helping a friend prepare for the next skateboard exhibition. Your friend, who weighs 60 kg, will take a running start and then jump with a speed of 1.5 m/s onto a heavy-duty 5 kg stationary skateboard. Your friend and the skateboard will then glide together in a straight line along a short, level section of track, then up a sloping concrete inclined plane. Your friend wants to reach a minimum height of 3 m above the starting level before he comes to rest and starts to come back down the slope. Knowing that you have taken physics, your friend wants you to determine if the plan can be carried out or whether he will stop before reaching a 3 m height. Do not ignore the mass of the skateboard.

FIG. 2. The pre (quiz 7) problem used in a self-diagnosis task (written by the University of Minnesota [22]).



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

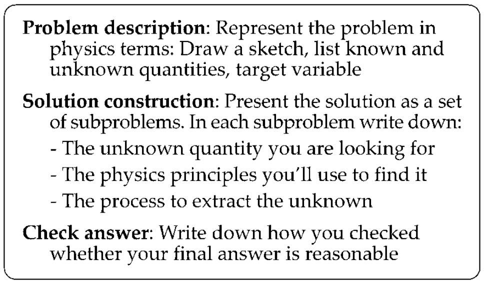
Mistake #	Mark x if mistake is in:			Explain mistake
	Physics	Math	Other	
1				
…				
…				

FIG. 3. Self-diagnosis rubric.

The different interventions were as follows.

In intervention group B, the instructor presented an outline of the correct solution and students were required to circle mistaken parts in their solution and fill in a self-diagnosis rubric (shown in 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 was intended to direct students' attention to the steps in a systematic solution process in which their mistakes occurred, as well as to the presentation of reasoning in a systematic manner.

In intervention group C the students were provided with a worked-out example handed out by the instructor during the self-diagnosis activity. This solution was based on guidelines for presenting a problem solution in a way that follows the steps of a problem-solving strategy (shown in Fig. 4) that was given to the students early in the semester. Figure 5 shows the worked-out example used.



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

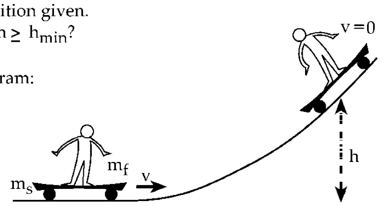
FIG. 4. Guidelines for presenting a problem solution according to a problem-solving strategy.

1) Description of problem

Knowns:
- Friend's mass m_f=60 kg
- Skateboard mass m_s=5 kg
- Friend's speed when jumping on skateboard v_f=1.5 m/s
- Desired minimum height h_{min}=3 m

Target quantity: actual height h reached under the initial condition given.
- is $h \geq h_{min}$?

Diagram:



Assumption: ignore retarding effects of friction and air resistance

2) Constructing the problem

Plan: We notice that the problem has two distinct components:
- The friend jumping over the skateboard and coming to rest with respect to skateboard. This is completely inelastic collision. We must find speed of cart with your friend after collision.
- The system consisting of the friend + skateboard goes up the ramp then stops at height h when kinetic energy is zero.

- We note that we can use conservation of momentum for the first part to find speed of skateboard + friend.
- Then we can use conservation of mechanical energy for the second part to find height h at which the skateboard stops.

Execution of the plan:
Subproblem A: calculating the speed v of friend + skateboard after inelastic collision
Since momentum of system of friend + skateboard is conserved,

$$\mathbf{P}_0 = \mathbf{P}_f$$
$$m_f \cdot v_f + m_s \cdot v_s = (m_f + m_s) \cdot v$$

Since initial speed of skateboard is zero,

$$m_f \cdot v_f = (m_f + m_s) \cdot v \Rightarrow v = m_f \cdot v_f / (m_f + m_s)$$

Subproblem B: calculating the maximum height reached before coming to a stop momentarily
From conservation of mechanical energy:

$$K_0 + U_0 = K_f + U_f$$

Mechanical energy on horizontal surface | Mechanical energy on the incline when skateboard stops momentarily

$U_0 = m \cdot g \cdot h_0 = 0$ if we choose reference height to be on a horizontal surface.
$K_f = 0$ at highest point since speed is zero.

$$1/2 \cdot m \cdot v^2 = m \cdot g \cdot h \quad \text{where } m = m_f + m_s$$
$$v^2 = 2 \cdot g \cdot h$$
$$h = v^2/2g = [m_f \cdot v_f/(m_f + m_s)]^2/2g =$$
$$= [60_{kg} \cdot 1.5_{m/s}/(60+5)_{kg}]^2/2 \cdot 10_{m/s^2} = 0.096 \text{ m}$$
$$h < h_{min} \text{ friends plan fails}$$

3) Reasonability check:
- Unit is correct for h
- We can caluclate the speed with which system of friend + skatboard should move horizontally to reach 3 m height
  $v = [2 \cdot g \cdot h]^{1/2} = [2 \cdot 10 \cdot 3]^{1/2} = 7.75_{m/s}$
  which is much faster than friend's speed.

FIG. 5. Worked-out example for the pre problem (quiz 7), aligned with the guidelines.

Fred Flintstone just got off work, and exits in his usual way, sliding down the tail of his dinosaur and landing in his car (see Figure). Given the height of the dinosaur (h=10 m), it's not hard to calculate his speed v as he enters his vehicle.
Conservation of energy yields the following equation: $mgh=1/2\ mv^2$, where m=100 kg is Fred's mass and v is his speed. Algebraic manipulation yields v=sqrt(2gh)=14 m/s. Judging from the picture shown in Fig. 1E, the angle at which Fred enters the car is approximately 45°. (a) If the mass of the car is M=200 kg, find the speed with which Fred is driving in the last frame (Fig. 1F), assuming he hasn't used his feet to pedal. (Remember also that there are no fossil fuels since there are no fossils yet.) (b) Assuming that there is no friction or air resistance, determine the maximum height H that Fred and his car can travel without extra pushing.

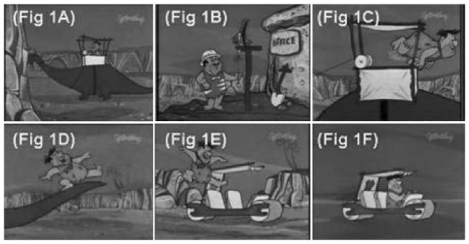


FIG. 6. The post (3rd midterm) problem.

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

The midterm problem shown in Fig. 6 functioned as a post problem to assess students' ability to transfer the understanding gained when diagnosing the quiz 7 problem in order to solve a problem in a somewhat similar context. This problem is similar to the quiz 7 problem in that it employs the same physical principles, i.e., conservation of linear momentum and conservation of mechanical energy. Thus, the two problems are isomorphic with regard to the physics principles involved.

Table III shows the comparison between the pre (quiz 7) problem and the post (3rd midterm) problem.

TABLE III. Comparison of pre (quiz 7) and post (3rd midterm) problems.

Activity	Principles	Variables	Context	Details
Quiz 7	Energy and momentum	v_c, h	Skateboard	v_i: horizontal component only
3rd midterm	conservation		Car	v_i: horizontal and vertical components

TABLE IV. Self-diagnosis grading rubric adapted to pre (quiz 7) problem. The last three columns would be filled in similarly for a sample student as they are for Table IV in the companion paper [9]; here, there is no student sample. The titles for the last three columns are as follows: RDS refers to the researcher diagnosis of student's solution; SDS, student's self-diagnosis of solution; RSD, researcher's assessment of student's self-diagnosis.

General task	Specific criteria		RDS	SDS	RSD
Invoking principles	1. Conservation of mechanical energy				
	2. Momentum conservation				
	3. Defining the system appropriately and consistently				
	Inappropriate principle: marked “−” if inappropriate principle is used in student's solution or diagnosis				
Applying principles	1. Conservation of mechanical energy	**Incorrect idea (example):** deficient analysis of problem situation—initial speed (etc.)			
	2. Momentum conservation	**Incorrect idea (example):** deficient analysis of problem situation—speed of skateboard before collision is not zero (etc.)			
Algebra	Algebraic manipulation				

III. ANALYSIS TOOLS

We adapted the scoring rubric described in detail for the first part of the study ([9], Sec. IV B) to suit the different problem discussed in this part. The original rubric incorporated both generic and specific elements. Specific elements were represented in the “Physics” category and involved the physical principles the student needed to invoke and apply in order to solve the specific problem. General elements were represented in the “Presentation” category and expressed how well the solution was communicated and justified. The overall aim was to examine the effects of self-diagnosis on subsequent problem solving in the case of atypical and of typical problems. Problem typicality is a function of the extent to which students have been exposed to isomorphic problems. We did not expect (and indeed did not find) that problem typicality would impact the presentation, and we therefore excluded the presentation category from the adapted rubric shown in Table IV.

As with our construction of the original rubric, here too we first took a top-down approach, where we created a representation of the “ideal knowledge” underlying an expert approach to the problem. This representation was intended to allow us to examine the extent to which each student's approach included certain elements of the “ideal knowledge.” Later we added a bottom-up approach [23], in which we went over the students' work and identified common mistakes in their approach. We represented these mistaken approaches in the rubric under the subcategory “incorrect ideas.”

The rubric was divided into two major subcategories: invoking a physical principle and applying that principle. Each row in each subcategory was therefore constructed so as to cover, when taken altogether, every physical principle that a student would need to invoke and apply in order to correctly solve the specific problem.

The work of each student was evaluated in three ways. The first evaluation was carried out according to the researcher's diagnosis of the student's quiz solution (RDS). The second work was evaluated according to the student's self-diagnosis of their solution (SDS). The third evaluation was based on 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 evaluations, 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 assigned “+” if a student had performed correctly or identified a mistake defined by some subcategory. We assigned “−” if the student had performed incorrectly, or failed to identify a mistake, or identified it incorrectly. Each “+” is worth 1 point and each “−” is worth 0. If a student was judged to have gotten something partially correct, then the grader would assign $++/-$ (i.e., 0.66), $+/-$ (i.e., 0.5), or $+/--$ (i.e., 0.33). We assigned “NA” if the student could not reasonably be expected to address a subcategory given the prior work done. In the SDS column, if a student had correctly diagnosed a mistake, we assigned “−” (since this was the grade awarded by the student to their solution). If a student did not refer to a mistake made, we assigned “×,” and awarded a score of 1 point in this category because we assumed that the student had treated the mistake as correct. An example of the scoring of one student's solution and self-diagnosis can be found in Fig. 6 and Table IV of the companion paper ([9], Sec. IV B). The validity of the rubric, i.e., the extent to which it indeed

TABLE V. Average physics grades (normalized to 1) for pre (quiz 7) 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	0.54	0.45	0.41	0.50
Standard error	0.04	0.03	0.04	0.04

allowed us to map the student's solution to the "ideal knowledge" of an expert as well as to the "incorrect ideas" of a novice, was determined by four experts in physics education, all of whom perceived the rubric as appropriately measuring a student's performance of the solution and self-diagnosis. The completed rubric for each student was analyzed by two researchers and any disagreements were discussed and resolved. Inter-rater reliability was 80% for a sample of 20% of the students graded by both researchers before discussion, and almost 100% after discussion.

IV. FINDINGS

In our companion paper [9] we reported that in the context of an atypical problem, the average self-diagnosis performance improved as the degree of external support increased. However, students' self-diagnosis performance did not correlate with their performance in transfer problems. In the following we will first present the findings in the context of the more typical problem(s) that characterized the current part of the study. We will then discuss the findings of the second part of the study, in relation to the findings of the first part, in order to shed light on how the different setting of a more typical problem—in which we expect students to have had prior accessibility to similar problems—affects the students' ability to learn from reflecting on their solution through a self-diagnosis task. We compare group averages as well as correlations within each group.

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

	Group B	Group C	Group D
Mean	0.57	0.59	0.63
Standard error	0.06	0.07	0.07

A. Group averages, typical problem

As shown in Table V, the mean grades for the pre problem (quiz 7) in all control (A') and intervention (B, C, D) groups were low. Even though the problem was a typical one, all students made mistakes. We therefore included the entire study population in the analysis.

Table VI presents a comparison of students' self-diagnosis performance in the different groups.

It might be expected that the self-diagnosis performance would improve as the level of external support increases. However, analysis of variance (ANOVA) (with physics grades in the pre problem as covariate) (Table VII) showed no differences between the groups ($P > 0.05$). This suggests that when self-diagnosing a solution to a typical problem, students made as effective use as possible of whatever resources and tools they were given. Even group D students, who were allowed to use only their notes and textbook, achieved fairly reasonable self-diagnoses.

Table VII presents a more detailed analysis of students' self-diagnosis performances in quiz 7. The table shows how well students were able to diagnose their deficiencies, both in invoking the correct physics principles and in applying them. When we examined the students' most common deficiencies in the two subcategories of invoking and applying principles (see Table IV, which specifies the principles needed to solve the problem), we observed that many students had overlooked conservation of momentum and addressed only conservation of energy (see Table IV, which specifies the principles needed to solve the problem). This finding is consistent with a reported study on solving a ballistic pendulum problem, where a majority of students invoked either the conservation of mechanical energy principle or the conservation of momentum principle, but not both, because they were completely focused on only one of the two principles [24].

TABLE VII. Analysis of self-diagnosis subcategories. "+" represents a correct diagnosis, "+/−" represents a partially correct diagnosis, and "−" represents an incorrect diagnosis or no diagnosis performed. "Total" refers to the total percentages of students who made mistakes in either subcategory. The students who got a subcategory correct are not included. SD refers to the student's self-diagnosis.

Subcategory group	Invoking momentum conservation				Applying energy conservation			
SD performance	+	+/−	−	Total	+	+/−	−	Total
Group B	8 (28%)	13 (44%)	8 (28%)	29/31 (94%)	4 (57%)	3 (43%)	0	7/31 (22%)
Group C	8 (40%)	7 (35%)	5 (25%)	20/27 (73%)	2 (100%)	0	0	2/27 (7%)
Group D	12 (55%)	5 (22%)	5 (22%)	22/23 96%	1 (50%)	1 (50%)	0	2/23 (9%)

TABLE VIII. Average physics grades (normalized to 1) for post (3rd midterm) solutions for the different groups.

	Group A’	Group B	Group C	Group D
Mean	0.62	0.66	0.70	0.74
Standard error	0.04	0.04	0.07	0.04

In line with the above, we examined the ability of students to diagnose deficiencies in invoking momentum conservation, as well as in applying conservation of energy. Table VII shows the number of students who successfully diagnosed their mistakes in each subcategory, expressed as a percentage of those who made mistakes in that subcategory.

Most of the students who did not invoke conservation of momentum in their solution diagnosed this deficiency in their self-diagnosis. The few students who made mistakes in applying energy conservation were able to diagnose their deficiency. In group C, in which students had received the fully worked-out solution during their self-diagnosis activities, all students who were incorrect with regard to the energy conservation principle were able to self-diagnose correctly.

We now address the effect of students’ self-diagnosis on solving the post problems. Table VIII, which presents the scores of each intervention group on the midterm quiz, shows that all the scores were poor.

ANOVA (with physics solution grade as covariate) showed that the groups did not differ significantly from each other ($P > 0.05$). Thus, although the different groups had received different types of support when self-diagnosing, they did not obtain different results either for their self-diagnosis scores or for the post problem.

B. Correlations within groups, typical problem

Next, we examined how students’ self-diagnosis and subsequent performance of transfer problems depends on their prior achievements.

As mentioned in the companion paper [9] we expect that the greater the external support, the more “meaningful” 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. This way, low achievers will reduce the gap described [6,7] between them and high-achieving students in the inclination to self-explain and self-repair. We expect, therefore, in the case of meaningful intervention, a positive correlation between the self-diagnosis and the post problem. We expect that if the low achievers improve, reducing the gap between them and the better students, the correlation between the pre problem and its self-diagnosis will be insignificant; this also implies an expectation of nonsignificant correlations between the pre and post problems, since students who were formerly weak will improve if they have actually learned from the self-diagnosis. For the control group (A’), i.e., without any external support for self-diagnosis, we expect to obtain positive correlations between the pre and the post problems, since there will be no intervention aimed at reducing the gaps between the low and high achievers.

Table IX presents the correlations between scores in quiz 7 (pre) and self-diagnosis scores, between self-diagnosis scores and post-test problem scores, and between quiz scores and post-test problem scores. The table shows that for the control group (group A) the pre-test and post-test scores were indeed positively correlated, and also that there were no significant correlations between pre-test and post-test for the intervention groups. The table also shows that only group D satisfied all the other expectations for a meaningful intervention, and, in particular, the moderately strong positive correlation between the self-diagnosis and the post-test scores. Thus, group D, in which students had received the least support for self-diagnosis, was the only group in which students who performed better on self-diagnosis of physics errors also performed better on the post problem in the midterm exam.

C. Comparison of students’ performance on typical and atypical problems

In the first part of the study [9], which dealt with an atypical problem situation, we reported that students performed poorly on the pre problem (the mean physics score in quiz 6 over all groups was about 0.35; see Table VII in Sec. VI A of the companion paper [9]) and were more capable of invoking a correct principle than of applying the principle correctly (see Table IX in the companion paper [9]). In comparison, in the second part of the study, as described in this paper, even though the mean grades in the

TABLE IX. Correlations between pre and self-diagnosis scores, between self-diagnosis and post scores, and between pre and post scores.

		Group A’	Group B	Group C	Group D
Pre versus self-diagnosis	Correlation	Not applicable	0.34	0.30	0.21
	P value		0.06	0.13	0.34
Self-diagnosis versus post	Correlation	Not applicable	−0.26	0.29	0.54
	P value		0.18	0.20	0.01
Pre versus post	Correlation	0.54	0.13	0.29	0.20
	P value	00003	0.53	0.20	0.38

pre problem (quiz 7) for all intervention groups were low (the mean physics score over all groups was about 0.45), they were higher than the quiz 6 grades (see Table V). The most common mistake made by the students concerned invoking the principle of momentum conservation. Thus the problem, being a more typical one, was somewhat easier.

In the context of the atypical problem we found that self-diagnosis performance improved with the level of scaffolding provided, so that the self-diagnosis performance of group B, which received the most support (outline solution + self-diagnosis rubric), was the best, and that of group D (minimal guidance, only notes and books) was the worst (see Table VIII in Sec. VI A of companion paper [9]). We also found that students were better at identifying mistakes in invoking principles than in applying those principles. In comparison, in the second part of the study, as described here, we found that the achievements of the different groups in self-diagnosis were similar, regardless of the external support provided in the different self-diagnosis tasks (see Table VI). We concluded that in the case of a typical problem, students were able to self-diagnose their mistakes even when the external support was minimal. The relatively good self-diagnosis performance of students who were not provided with a sample problem solution can be explained in terms of the accessibility of related worked-out examples in their textbooks. Indeed, one of the solved examples that was used in the course textbook [25] was about the ballistic pendulum. It is known [26–29] that when students' representation of a problem situation is organized around surface features, this prevents the students from retrieving and implementing procedures from a worked-out example. In the case of the typical problem, although it involved a person jumping on a skateboard and climbing up a hill, which is somewhat different from the ballistic pendulum problem that appears in the textbook, the two problems evidently share enough surface features to make the textbook worked-out example retrievable.

Another finding regarding students' self-diagnosis was that, in the context of an atypical problem, students did better in realizing their deficiencies in invoking relevant principles than in realizing application deficiencies (see Table VII in this paper and Table IX in the companion paper [9]). For example, in quiz 6, many students who treated centripetal force as a physical force would later dismiss it in their self-diagnosis as a minor "math" mistake.

With regard to the transfer of what students learned from reflecting on their quiz problem solution to a paired problem, we found in the context of the atypical problem that while their self-diagnosis performance improved with the increase in external support, their performance on the post problems did not correlate with their self-diagnosis performance. ANOVA showed that group C students did not differ from group D students in their performance on the midterm transfer task, even though the former group was

TABLE X. Comparison of students' performance in a self-diagnosis task and subsequent post (transfer) problem in the context of an atypical and a typical problem situations. SD refers to the students' self-diagnosis. NA means there is no applicable correlation for the item in question.

Problem		Atypical					Typical			
Pre-test performance		Quiz 6: poor overall performance In general students were able to invoke relevant principles. The main difficulties were in applying them carefully.					Quiz 7: medium overall performance The most common mistake was in invoking the relevant principles. Difficulties in careful application of principles were less evident.			
Self-diagnosis performance		Group		B	C	D	Group	B	C	D
		Mean		0.73	0.57	0.24	Mean	0.61	0.62	0.56
		Standard error		0.05	0.05	0.05	Standard error	0.06	0.06	0.06
		Students were better at realizing their deficiencies in invoking relevant principles than at realizing application deficiencies.					Group D improved significantly. Students were better at realizing their deficiencies in invoking relevant principles than at realizing application deficiencies.			
Post-test performance		2nd midterm—poor performance group D ≈ group C ≈ group B					3rd Midterm—medium performance group D ≈ group C ≈ group B			
Correlations	Group		A'	B	B	C	A'	B	C	D
	Pre vs SD	Correlation	NA	−0.04	0.06	−038	NA	0.34	0.30	0.21
		P value	NA	0.83	0.77	0.06	NA	0.06	0.13	0.34
	SD vs post	Correlation	NA	0.35	0.14	0.11	NA	−0.26	0.29	**0.54**
		P value	NA	0.16	0.55	0.71	NA	0.18	0.20	**0.01**
	Pre vs post	Correlation	**0.41**	0.14	0.25	0.16	**0.54**	0.13	0.29	0.20
		P value	**0.03**	0.57	0.06	0.61	**0.0003**	0.53	0.20	0.38

provided with a solution and the latter had to work out, without the assistance of a written solution, what had gone wrong and why (see Table X in the companion paper [9]).

In the current part of the study, which focuses on a typical problem situation, we also found that the groups did not differ significantly in their performance on the transfer problems (see Table VIII).

Finally, we compared how students' self-diagnoses and subsequent performance of transfer problems depend on their prior achievements on the pre-quiz. We looked at the correlations between the pre-test scores in the quiz and the self-diagnosis scores, between the self-diagnosis scores and the post-test problem scores, and between the quiz physics scores (pre) and the transfer (post) problem scores (Table IX in this paper, Table XI in the companion paper [9]). In both the atypical problem setting and the typical problem setting, the intervention groups did not have the strong pre-versus-post correlations found to exist for the control group (A'). In other words, the distribution of students' grades in quiz 7 was not carried over to the midterm grades, implying that the interventions were helping some of the students, possibly the lower achievers, more than others. On the other hand, there was a nonsignificant correlation between self-diagnosis and post grades for all groups except for D. This calls into question the effect of the self-diagnosis task on learning, as those who self-diagnosed better did not do better on the post problem at the midterm. The main difference between the two settings was seen in group D, for which, in the context of a typical problem, the self-diagnosis grades were significantly correlated with the post-test performance.

Table X summarizes the comparison between the two settings with regard to both the group averages and the within-group correlations.

V. DISCUSSION

Table X shows the observed correlations in the three intervention groups for both atypical and typical problem settings. The results are puzzling. On the one hand, there are no significant correlations between the pre and post attempts in any of the intervention groups in either setting, whereas a positive correlation exists between these two variables in the control group in both settings. This trend was observed whether the interval between the self-diagnosis activity and the post-test was three days (as in the case of the atypical problem) or three weeks (as in the case of the typical problem).

This suggests that the interventions reduced the gaps between low and high achievers, whereas in the control groups the gaps between low and high achievers remained unchanged. This result was further supported by evidence [9] that the interventions helped low-achieving students to perform relatively better on the post-test than on the pretest. On the other hand, there are no significant correlations between self-diagnosis and post-test performances for any of the intervention groups, except in one case—group D in the context of a typical problem.

In the companion paper focused on the atypical problem situation [9], we explained this inconsistency by pointing out that the analysis rubric we devised for grading students' self-diagnosis performance did not differentiate between a meaningful self-diagnosis (i.e., one that leads to self-repair of students' mental models) and a superficial self-diagnosis [9]. For a self-repair process [5] to take place, students need to acknowledge conflicts between the beliefs underlying their solution and the scientific model underlying a correct solution, as represented, for example, in the instructor's solution. When self-diagnosing their solutions, students rarely articulated explanations concerning the nature of the conflict between their thinking and the correct scientific model. Therefore, we could not assess from their self-diagnoses the extent to which they might be accompanied by self-repair. As a result, the rubric focused on whether students realized, while self-diagnosing, that certain principles or concepts had not been invoked or correctly applied. Students in the intervention groups who were given an outline of the solution (group B) or a worked-out example (group C) could easily realize that their solution differed in the principles that should have been invoked, hence, their better average grades for self-diagnosis as compared to group D students. But this apparent superiority might not have been accompanied by a meaningful learning process in which students recognized how these principles and concepts had been misinterpreted. As most students did not provide such explanations, the self-diagnosis grade by itself did not allow us to know if the diagnosis was accompanied by a superficial or a meaningful learning process.

That being the case, we would expect to find a nonsignificant correlation between the performance on the pre-test and the performance in self-diagnosis, since an adequate self-diagnosis could be provided by the less achieving students as well, and thus the gap between them and the high achievers would be reduced. We would also expect to find no correlation between the self-diagnosis performance and the post-test performance, since this self-diagnosis is not necessarily accompanied by self-repair, and accordingly does not allow for the occurrence of transfer, as reflected in the performance on the midterm problems.

In what way is group D in the second (typical) setting different? Group D students had first to find and identify information in their notes or textbooks that was relevant for the problem to be diagnosed. For example, they needed to access, without assistance, solutions related to similar problems. They had to recognize an analogy between their own solution and the textbook solution to a similar problem. They then had to struggle to self-diagnose their mistakes. It is likely that these students experienced more cognitive involvement in their self-diagnostic activity than group B or C students, who possessed readily accessible worked-out examples as the scaffold for self-diagnosis.

This might explain the different behavior of the students in group D. In other words, the more the external support, the less we could differentiate between students whose self-diagnosis was or was not accompanied by self-repair. When students received an outline or a detailed worked-out example (group B or C, respectively), they might have ‘‘copy pasted,’’ thus, providing an ‘‘acceptable’’ diagnosis without actually self-repairing their mental model, and we could not distinguish between those who carried out self-repair and those who did not. Students who received minimal guidance (group D) had to struggle on their own, and as a result, students who provided an ‘‘acceptable’’ diagnosis would commonly be those who had self-repaired their mental model. This more meaningful diagnosis was reflected in the midterm test.

An important question remains: why did group D behave like all other groups in the ‘‘typical problem’’ part of the study, i.e., why did their pre-test and post-test performances show a nonsignificant correlation? In the first part of our study [9], group D struggled with an atypical quiz problem, which was beyond their zone of proximal development. Their self-diagnosis grades were very poor. In the second part of the study, however, group D students were able to self-diagnose their mistakes in the quiz problem with the aid of textbook and notes, and their self-diagnosis scores were therefore comparable to those of the other experimental groups (B and C).

This finding is in line with the model of Bransford and Schwartz [14] with regard to the roles of ‘‘efficiency’’ and ‘‘innovation’’ in transfer. The self-diagnosis task has to allow for *innovation* by providing an opportunity for students to evaluate and refine their ideas. The innovative component can only be fruitful, however, when adequate consideration is given also to the task’s *efficiency* component, by allowing for students to connect what they are learning with their prior knowledge. According to the model [14], students’ poor transfer in the case of the atypical problems might be a result of the intervention’s main focus on problems that require students to be highly innovative, as they are not accustomed to these types of problems. This result is also consistent with findings [16] that in the context of probability problems, learners with low prior knowledge benefit more from learning from worked-out examples than learning from incorrect solutions, whereas learners with greater prior knowledge benefit more from learning from incorrect solutions. Focusing on typical problems, for which students have gained expertise and can more readily retrieve required knowledge, would expand the task’s efficiency component. However, such problems might lower the level of mental activity required. We suggest that group D represents a satisfactory balance between the innovation and the efficiency components. Since group D students were not provided with a solution, these students, unlike the others, had to carry out a mentally demanding task.

To summarize, we conclude that an intervention which requires students to self-diagnose their solutions can induce students to generate self-explanations involving self-repair. Taken together, the two parts of this study point to the importance of providing problems commensurate with students’ prior knowledge so that self-diagnosis will be within their zone of proximal development. In terms of the Bransford and Schwartz model [14], we propose that the ability to transfer the knowledge constructed in a self-diagnosis exercise to the solution of new problems will be enhanced if students are, on the one hand, challenged by the need to realize their mistakes and to acknowledge how their thinking conflicts with the scientific view, and, on the other hand, able to retrieve the prior knowledge required to perform the self-diagnosis task.

ACKNOWLEDGMENTS

We wish to thank Professor Jeremy Levy for his valuable cooperation and help 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] C. Henderson and K. A. Harper, Quiz corrections: Improving learning by encouraging students to reflect on their mistakes, Phys. Teach. **47**, 581 (2009).

[3] E. Yerushalmi, C. Singh, and B. Eylon, in *Proceedings of the Conference on Physics Education Research*, edited by L. McCullough, L. Hsu, and C. Henderson (AIP, Melville, NY, 2007).

[4] 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).

[5] M. Chi, in *Advances in Instructional Psychology*, edited by R. Glaser (Lawrence Erlbaum Associates, Hillsdale, NJ, 2000), pp. 161–238.

[6] 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).

[7] M. Chi, N. de Leeuw, M. Chiu, and C. LaVancher, Eliciting self-explanations improves understanding, Cogn. Sci. **18**, 439 (1994).

[8] D. B. May and E. Etkina, College physics students' epistemological self-reflection and its relationship to conceptual learning, Am. J. Phys. **70**, 1249 (2002).

[9] E. Yerushalmi, E. Cohen, A. Mason, and C. Singh, preceding paper, What do students do when asked to diagnose their mistakes? Does it help them? I. An atypical quiz context, Phys. Rev. ST Phys. Educ. Res. **8**, 020109 (2012).

[10] 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).

[11] G. Polya, in *How to Solve It* (Princeton University Press, Princeton, NJ, 1945).

[12] F. Reif, Millikan Lecture 1994: Understanding and teaching important scientific thought processes, Am. J. Phys. **63**, 17 (1995).

[13] 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).

[14] J. D. Bransford and D. L. Schwartz, in *Review of Research in Education*, edited by A. Iran-Nejad and P. D. Pearson (American Educational Research Association, Washington, DC, 1999), Vol. 24, pp. 61–100.

[15] D. L. Schwartz, J. D. Bransford, and D. Sears, in *Transfer of Learning from a Modern Multidisciplinary Perspective*, edited by J. P. Mestre (Information Age Publishing, Greenwich, CT, 2005).

[16] C. S. Große and A. Renkl, Finding and fixing errors in worked examples: Can this foster learning outcomes?, Learn. Instr. **17**, 612 (2007).

[17] R. K. Atkinson, S. J. Derry, A. Renkl, and D. W. Wortham, Learning from examples: Instructional principles from the worked examples research, Rev. Educ. Res. **70**, 181 (2000).

[18] J. Sweller, J. J. G. van Merriënboer, and F. G. Paas, Cognitive architecture and instructional design, Educ. Psychol. Rev. **10**, 251 (1998).

[19] J. Reisslein, R. K. Atkinson, P. Seeling, and M. Reisslein, Encountering the expertise reversal effect with a computer-based environment on electrical circuit analysis, Learn. Instr. **16**, 92 (2006).

[20] E. Yerushalmi, A. Mason, E. Cohen, and C. Singh, in *Proceedings of the 2009 Physics Education Research Conference, Ann Arbor, MI, 2009*, edited by M. Sabella, C. Henderson, and C. Singh, AIP Conf. Proc. (AIP, Melville, NY, 2009).

[21] A. Mason, E. Cohen, C. Singh, and E. Yerushalmi, in *Proceedings of the 2009 Physics Education Research Conference, Ann Arbor, MI, 2009* (Ref. [20]).

[22] UMPERD archives, http://groups.physics.umn.edu/physed/Research/CRP/on-lineArchive/crcecm.html#cem, problem no. 20.

[23] M. Chi, Quantifying qualitative analyses of verbal data: A practical guide, J. Learn. Sci. **6**, 271 (1997).

[24] C. Singh and D. Rosengrant, Multiple-choice test of energy and momentum concepts, Am. J. Phys. **71**, 607 (2003).

[25] J. Cutnell and K. Johnson, *Physics* (John Wiley and Sons, New York, 2007), 7th ed., pp. 206–209.

[26] M. Chi, P. J. Feltovich, and R. Glaser, Categorization and representation of physics problems by experts and novices, Cogn. Sci. **5**, 121 (1981).

[27] B. Eylon and F. Reif, Effects of knowledge organization on task performance, Cogn. Instr. **1**, 5 (1984).

[28] 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).

[29] M. L. Gick and K. J. Holyoak, Schema induction and analogical transfer, Cogn. Psychol. **15**, 1 (1983).