

## Choosing the right solution approach: The crucial role of situational knowledge in electricity and magnetism

Elwin R. Savelsbergh\*

*Freudenthal Institute for Science and Mathematics Education, Utrecht University,  
P.O. Box 80000, NL-3508 TA Utrecht, The Netherlands*

Ton de Jong

*University of Twente, P.O. Box 217, NL-7500 AE Enschede, The Netherlands*

Monica G. M. Ferguson-Hessler

*Eindhoven University of Technology, P.O. Box 513, NL-5600 MB Eindhoven, The Netherlands*

(Received 31 March 2010; published 28 March 2011)

Novice problem solvers are rather sensitive to surface problem features, and they often resort to trial and error formula matching rather than identifying an appropriate solution approach. These observations have been interpreted to imply that novices structure their knowledge according to surface features rather than according to problem type categories. However, it may also be the case that novices do know problem types, but cannot map the problem at hand to a known type, because they fail to create a sufficiently well-elaborated problem representation. This study aims to distinguish between these explanations. In this study novice physics students at high and low levels of proficiency completed two problem-sorting tasks from the domain of electricity and magnetism, one with and one without elaboration support. Results confirm that these students do distinguish problem types in accordance with their required solution approaches, and that their problem-sorting performance improves with elaboration support. Therefore, it was concluded that their major difficulty lies in the process of matching concrete problems to a proper category. Within-group analysis revealed that the performance of proficient novices clearly improved with elaboration support, whereas the effect for less proficient novices remained inconclusive. The latter finding is explained from the less proficient novices' problem representations being too fragmented to integrate new information. These results suggest that, in order to promote schema-based problem solving, instruction in the domain of electricity and magnetism should be based not so much on restructuring the conceptual knowledge base but rather on enriching situational knowledge.

DOI: [10.1103/PhysRevSTPER.7.010103](https://doi.org/10.1103/PhysRevSTPER.7.010103)

PACS numbers: 01.40.Fk

### I. INTRODUCTION

It is an often heard complaint that novice problem solvers will skip most of the problem analysis and start calculating right away. Many instructors strive to teach their students more expertlike approaches, [1] such as analyzing the problem and devising a solution plan [2], but in general such interventions turn out to be little effective [3,4].

The experts' behavior has been explained from their knowledge being organized in problem schemas, i.e., knowledge structures organized around a problem type with an associated solution approach. Once the right problem type has been identified, the expert will also know an outline for a solution approach [5]. It seems worthwhile to

promote the formation of powerful problem schemas among physics learners. However, in order to do so effectively, one needs to know what exactly is missing from the novices' knowledge structure.

The classical problem-sorting experiment by Chi, Feltovich, and Glaser might be among the most well-known studies to discriminate between novice and expert knowledge structures [6]. In their study, Chi *et al.* let several experts and novices sort a set of mechanics problems. The experts would spontaneously sort these problems according to their solution approaches (i.e., second law, conservation of energy, conservation of momentum), whereas the novices would sort according to surface features (angular motion, springs, inclined planes). Although such studies provide convincing evidence for the quality and structure of expert problem schemas, the interpretation of the novice results is much less conclusive [7]. For instance, whereas Chi *et al.* [6] concluded that novices did not have their knowledge organized in problem type schemas, Cooper and Sweller [8] found that novices already start to induce problem categories after trying only a few problems.

\*e.r.savelsbergh@uu.nl

*Published by American Physical Society under the terms of the Creative Commons Attribution 3.0 License. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.*

Moreover, whereas Chi *et al.* [6] concluded that novices had “sufficiently elaborate declarative knowledge about the physical configurations of a potential problem,” many studies indicate that novices form only limited representations of the problems they are working on [9–12]. To complicate matters, there is clear evidence for qualitative differences between more and less proficient novices [9,13–16], so, even if they both fail to identify a proper solution approach, it still may be for different reasons.

It should be noted that studies about problem schemas were conducted in different domains. For instance, the experimental task of Chi *et al.* [6] was drawn from the domain of mechanics, whereas the studies of de Jong and Ferguson-Hessler [9] were conducted in the field of electricity and magnetism. Although studies in both fields confirm that expert problem solving knowledge tends to be organized in a limited number of problem solving schemas, it is not self-evident that more specific findings about problem schemas generalize across domains. First, there may be considerable structural differences between domains: in mechanics the deep features of a problem tend to be associated with a small number of conservation principles (conservation of energy, momentum, angular momentum), which can be sharply distinguished from superficial features such as spatial symmetries. By contrast, in the field of electricity and magnetism, spatial symmetries are crucial to the choice of an appropriate solution approach. Moreover, mechanics differs from other physics topics, such as electricity and magnetism, in its situational features being more concrete, and in involving more everyday knowledge [17–19]. On the one hand, students’ experiences in the physical environment may help them to develop a clear representation of a mechanics situation, whereas, on the other hand, their experience-based expectations may also turn out to be “misconceptions” leading them in the wrong direction [20,21]. Therefore, in order to generalize educational implications, the experimental findings must be validated across multiple domains.

Educational implications are clearly different depending on which explanation holds true: If novices rely on alternative category structures, we may aim to restructure their knowledge, by teaching knowledge hierarchies [22], or by inducing cognitive conflicts [23,24]. By contrast, if the novice categories are created “on the spot,” building on a poor problem representation, instruction should target students’ understanding of the situations and of the situational features that make a solution approach into a useful one [10,25–27].

The purpose of this study is to find out to what extent each of both factors (that is, poor representation of the problem situation and missing knowledge of problem type schemas) hampers the ability of novices at different levels of proficiency to identify solution approaches in the domain of electricity and magnetism.

### A. Identifying a solution approach: Roles of elaborations and schemas

Every problem solving effort must begin with creating a representation for the problem, a problem space in which the search for a solution can take place [28]. First, upon reading a new problem, one forms a text based representation [29]. Keywords in the text based representation may immediately suggest a solution approach, and both experts and novices make productive use of such keywords to determine a solution approach [6,30–32].

Starting from the text based representation, information has to be selected, knowledge from memory has to be added, and the various elements have to be connected to form a structured mental representation of the situation [29,33]. Even if one has gained a correct understanding of the situation, this may not be the optimal representation: the same problem may be easy or difficult, depending on the way we describe it. For good reason, one of Polya’s famous problem solving heuristics is, “Could you restate the problem? Could you restate it still differently?” [34]. Thus, while for some problems a surface representation of the situation may provide sufficient guidance to find an adequate solution approach, most problems will require elaborations in order to identify the deep features of the problem.

Following VanLehn [35], we define an elaboration as “an assertion that is added to the state without removing any of the old assertions or decreasing their potential relevance.” In many cases, a series of elaborations will be necessary before the problem is “well evolved.” Some elaborations will be made on the basis of common sense [36], but many will also require domain knowledge, as in the following problem:

Given two copper cylinders  $O_1$  and  $O_2$ , with radii  $r_1$  and  $r_2$ , the relation between the radii being given by  $r_2 = 3r_1$ . Both cylinders are centered along the  $x$  axis.  $O_2$  is grounded;  $O_1$  carries a charge  $q_1$  per meter. Compute the potential at the surface of  $O_1$ .

In the mind of a problem solver, this problem can evoke many potentially relevant elaborations:

The cylinders are concentric— $O_1$  is at the inside of  $O_2$ —the situation is symmetric under rotations around the  $x$  axis—the situation has translational symmetry along the  $x$  axis—copper is a conductor—in a conductor the potential is equal everywhere—at the surface of a conductor, the field is perpendicular to the conductor— $O_2$  shields its inside from external fields and vice versa—the potential at  $O_2$  is zero—the total charge enclosed in a surface that lies completely in the metal of  $O_2$  amounts to zero—there is a surface charge of  $q_1$  per meter at the inside of  $O_2$ —the potential runs continuously across this surface charge.

Each elaboration considered by itself provides little direction, and hardly any of them could be identified as providing *the* crucial step. However, taken together, these

elaborations, also called “construction rules,” lead to a “physics representation” [5,37].

In order to build such a representation, one must have a sense of the features that matter. As Hestenes notes in comparing physics problem solving to chess playing, “the chess player’s attention is automatically confined to the chessboard, where the patterns are to be found. But in physics a student who hardly knows what the game is about is likely to attend to the wrong things” [25,38]. Therefore, teaching a “vocabulary of patterns” is a core element of Hestenes’ modeling approach to physics teaching. To a novice, identifying a solution approach for a physics problem can thus be a real insight problem, in the sense that he will find few cues as to whether he is close to seeing the approach, until the approach has actually been identified [39,40].

In addition, if the newly inferred information cannot be integrated in a coherent problem representation, working memory will soon be overloaded with isolated facts. In many domains, it has been found that experts have a superior recall of problem situations, which indicates that they perceive an integrative structure in the situation [41,42]. While proficient novices were found to develop a coherent mental model of the situation, less proficient novices tend to rely on a text based representation of the problem, which provides little support to integrate elaborations [9,14,15]. In line with this finding, there is clear evidence that less proficient students are little inclined to elaborate on the problems they encounter [9,14,15,43], and that their problem solving performance improves if they are forced to perform a structured analysis [44].

The expert’s knowledge is much richer, based on past experience with similar problems. This is reflected in their knowledge being structured in problem schemas [6,31,45,46]. Such a problem schema would be a coherent body of all knowledge relevant to a particular “basic solution approach,” which includes relevant theory, procedures, memories of prototypical problems and features, and conditions that must be met in order to make the approach useful. If a problem schema is sufficiently rich with prototypical problems and situational features, it will be easily matched when studying a new problem. Once the problem has been recognized as similar to a known type, the schema permits a much more targeted analysis, and furthermore directs the steps to be taken for solving the problem. Of course, even if one recognizes the appropriate problem type and the associated solution approach, one may still fail. That is to say, *knowing* a schema is not a matter of black and white. Nevertheless, at least for proficient novices, it has been found that the identification of a proper schema leads to better problem solving performance [16].

To sum up, there is clear evidence that experts elaborate on a problem to build a rich representation, which then easily matches a known schema. Novices build a much poorer representation, which does not easily trigger a schema. Nevertheless, it may be the case that they do

know schemas that might be helpful if they could only make the right match. Because observations of problem solving behavior, and even intervention studies such as Dufresne *et al.* [44], can provide only limited insight into participants’ knowledge structures, various experiments have been conducted to assess these knowledge structure more directly. In the next section we will review the evidence on a most basic element of schema knowledge, namely, whether novices do know problem types according to their solution methods.

## B. Assessment of schema knowledge: Evaluating the evidence

The most commonly used method to assess subjects’ perceptions of similarities between problems is a sorting task [47,48]. In comparison to a full problem solving task, problem sorting provides a more direct test of whether subjects are able to identify a basic solution approach, independent of whether they have the skill to execute the chosen approach. There have been alternative approaches to assess schema knowledge, such as problem editing [49] and writing down the first step of the solution under time pressure [50], but these techniques are more sensitive to the quality of the schema content, rather than to the category knowledge *per se*. Interview techniques have also been used, but the outcomes of such studies tend to be less directly linked to problem solving [21,51].

Chi *et al.* [6] used a problem-sorting task to demonstrate that novices (i.e., undergraduates who had just completed a relevant introductory course) categorize problems according to a qualitatively different structure than experts do. To further characterize the difference, they administered a second categorization task. In this case they deliberately constructed the set of problems such that the similarities suggested by the surface features would be at odds with the deep structure of the problems. The conclusion of this second experiment was that experts sorted according to deep structure, whereas the novices attended to the surface structure of the problems. With due caution, they conclude that “although it is conceivable that the categories constructed by novices do not correspond to existing internal schemata, but rather represent only problem discriminations that are created on the spot during the sorting tasks, the persistence of the appearance of similar category labels across a variety of tasks gives some credibility to the reality of the novices’ categories even if they are strictly entities related” [6]. By contrast, in a study among students of similar level, de Jong and Ferguson-Hessler [52] found that proficient students did sort knowledge elements according to problem types, and that only the less proficient students tended to sort according to surface features. Hardiman *et al.* [31], using a similarity judgment task, found that novices are able to identify the same similarities experts see, although they are more easily distracted by salient surface features, and they often use the wrong principles.

Hardiman *et al.* also found proficient novices to be more likely to base their judgment on physics principles than less proficient novices. Similar results were reported by Zajchowski [16], based on think-aloud protocols of a problem solving task.

The results of de Jong and Ferguson-Hessler [52], Hardiman *et al.* [31], and Zajchowski [16] suggest that at least more proficient novices do know a category structure based on problem types. For less proficient students, the evidence is less conclusive. However, even for these students, if they fail to demonstrate their category knowledge in a categorization task, it does not necessarily imply that they do not have such knowledge. In particular, it should be noted that both Chi *et al.* [6] and Hardiman *et al.* [31] obtained their results with a set of problems where surface features and deep structure were crossed. While this is an effective approach to demonstrate novices' sensitivity to surface features, it makes the outcomes less representative of what happens in a realistic problem solving setting, where a problem solver finds valuable cues with regard to the solution approach at all stages of interpreting the problem.

Given these considerations, we expect that, in the domain of electricity and magnetism, novices' failure to identify proper solution approaches should be attributed to a poor problem representation more than to a lack of category knowledge. In order to test this claim we will use a problem-sorting task where we manipulate the quality of the problem representations by providing simple first elaborations. Our central hypothesis is that providing novices with this form of "elaboration support" may lead to a more expertlike problem-sorting performance. We also expect that the usefulness of the given elaborations will depend on the quality of one's problem representation thus far. If one has a too incoherent problem representation, elaborations will remain isolated facts with little added value. Therefore, our second hypothesis is that the effect of given elaborations will be stronger for more proficient novices.

## II. METHOD

### A. Design

The hypotheses were tested in a within-subjects design where more and less proficient participants completed two different problem-sorting tasks, one with and one without elaborations. Performance on both tasks was judged by comparison to experts' problem sorting.

### B. Participants

Expert participants were three physics faculty who were, or had been, instructors in introductory and/or advanced electrodynamics courses.

Novice participants were 80 first-year university physics students who had just completed their first introductory course on electrodynamics. In The Netherlands, preuniversity education extends until the age of 18, and over 80% of the physics students are 18 or 19 years old at the beginning of their first year. The population is rather homogeneous in other respects as well: there are 80%–90% males, and over 90% are of Dutch ancestry.

Because the population within a single university was too small to provide a sufficient number of participants, participants were recruited from two different universities (hereafter "University A" and "University B"). In the Dutch educational system, there are no major status differences between the diplomas of different universities, and practical factors, such as vicinity, are the dominant factors in determining students' choice of university. Both student populations and curriculum are similar enough across both universities to regard them as samples from a single population.

To make sure that the students had spent some time on the topic, the experiment was conducted after the final test had been taken. First-year students were randomly selected from the faculty's phone directory and approached by telephone until the desired number of participants was reached. Participants were paid for their participation.

TABLE I. Mean test scores of more and less proficient groups. Note that all test scores are on a scale of 1 to 10, higher scores are better, 6 is the threshold between pass and fail.

		Test scores		
		National exam	University tests A	University tests B
More proficient students				
University A	<i>M</i> (SD)	8.93 (.59)	7.13 (1.50)	
	<i>n</i>	15	15	
University B	<i>M</i> (SD)	8.90 (.66)		7.87 (1.15)
	<i>n</i>	20 <sup>a</sup>		21
Less proficient students				
University A	<i>M</i> (SD)	7.25 (.59)	4.58 (1.07)	
	<i>n</i>	22	22	
University B	<i>M</i> (SD)	7.27 (.43)		5.92 (.88)
	<i>n</i>	13		13

<sup>a</sup>One participant got admittance on a foreign diploma.

TABLE II. Distribution of the problems over topics according to the experimenters.

Set 1: Electricity	Number of problems	Set 2: Magnetism	Number of problems
Gauss' law	6	Ampère's law	6
Image charges	5	Dipole approximation	5
Dipole approximation	5	Induction	5
Coulomb's law or superposition	4	Biot-Savart's law	4

TABLE III. Examples of electricity problems from two different problem categories. Note that for each problem there was an elaborated version and a nonelaborated version. The nonelaborated version only gave the normal printed text, the elaborated version also gave the italicized text.

Gauss' law problems	Coulomb's law with superposition
E10 Compute the field between two concentric spherical shells with the inner shell carrying a uniform charge $Q_1$ and the outer shell $Q_2$ <i>The field inside a uniformly charged spherical shell does not depend on the charge on the shell</i>	E5 A point charge $q$ is at the origin and a second point charge $-q$ is at a distance $\vec{R}$ from the origin. Compute the net electric field at a point $\vec{r}$ ( $ \vec{r}  \approx  \vec{R} $ ) <i>The net electric field is the sum of all contributions</i>
E17 Compute the field of a planar charge distribution that extends to infinity <i>The field of an infinitely large planar charge distribution has a field component perpendicular to the plane only</i>	E9 Compute the field at the axis of a charged ring. Charge $Q$ , radius $r$ <i>At the axis of a charged ring, only the field component along the axis remains</i>

In order to classify the participants as less or more proficient, test scores were collected both from the national high school final examination in physics and mathematics and from their previous university tests for classical mechanics, relativistic mechanics, electricity and magnetism, calculus and algebra. Because university test scores are not directly comparable across different institutions, each university had its own performance subscale (University A: 6 items, Cronbach  $\alpha = 0.92$ ; University B: 9 items, Cronbach  $\alpha = 0.93$ ). The scores on the national high school examination could be used to rescale the mean and variance of the university test scores in order to compute an overall student ranking.

Out of the 80 participants, 9 had either omitted a card or had mentioned a card more than once on one of their problem-sorting forms. As a consequence, the number of valid observations for the elaboration effect is  $N = 71$ . A median split on the overall student ranking was used to create sufficiently different groups of more and less proficient participants, with 36 students remaining in the more proficient group and 35 students in the less proficient group. As shown in Table I, the grade level on the university test scores differs by about two points on a scale of 10 between the two groups. The less proficient group on average scored below threshold on the university tests, whereas the more proficient group scored well above. A chi-square test confirmed that the distribution of students from the two universities over proficiency groups was not significantly skewed,  $\chi^2(1, N = 71) = 3.19$ ,  $p = 0.74$ .

### C. Materials

Two sets of 20 problems each were developed. One set consisted of problems from the field of electricity, the other of problems on magnetism. Our aim was to construct both sets in such a way that there would be four essentially different solution types in each set. Problems that required a combination of multiple approaches were not included. We did not include catch problems of types the students would never encounter in their practice problems, and we did not manipulate surface features to suggest a systematically different ordering. Within these constraints, we took care that the solution type could not be inferred from the topical area alone. The design procedure started from a larger set of problems that was presented to several teaching staff. Problems that were classified inconsistently were removed from the set. After this procedure, a set of 23 electricity problems and a set of 22 magnetism problems were left. Finally, in order to reduce the problems to two sets of 20 problems each, and to validate our intended categories, we had the three expert participants sort both sets, after which we kept the problems on which agreement was best. The design of the final sets is presented in Table II.

In order to test for the effect of elaborations, two versions were needed for each problem set, one with and one without elaboration. The elaborations were designed to be close to the original problem statement and to avoid keyword patterns that could promote a keyword matching

TABLE IV. The experimental setup.

Group	A	B	C	D
$n$	20	20	20	20
First set	Electricity, with elaboration	Electricity, no elaboration	Magnetism, with elaboration	Magnetism, no elaboration
Second set	Magnetism, no elaboration	Magnetism, with elaboration	Electricity, no elaboration	Electricity, with elaboration

strategy. Examples of problem cards with elaborations are presented in Table III (for full collection of problem cards, see the Appendix).

#### D. Procedure

Participants were to sort both sets of problems with each problem presented on a “problem card.” Half of the novice participants received elaboration support on the electricity problems, the other half on the magnetism problems. The design was counterbalanced to cancel out effects of order and task version (Table IV).

Before doing the first set, participants read the instructions. Unlike in the study by Chi *et al.* [6], our participants were explicitly instructed that problems were to be sorted according to solution approach. This was further illustrated using the example of how a cook might answer when asked about similarities between several dishes. The explicit instruction was included because our goal was to find out whether participants are able to categorize problems according to solution approach, not whether they would do so under all circumstances. The instruction went on with the request to read all problem cards in the first set, prior to doing any sorting. When the sorting was done, the participant would assign a name to each category.

The first set took approximately 50 minutes to complete. After a short break participants did their second set, which took about 40 minutes on average.

### III. RESULTS

#### A. Expert validity

To verify the agreement between the experimenters’ sorting and those by the three experts, we recoded the sorting data into a list of problem pairs. Each pair could have one of two values: “together” or “apart” (see the Appendix). With

the data in this format, we could compute an inter-rater reliability. Because we had four raters, we used an intraclass correlation coefficient [53]. Values were  $R = 0.77$  for the electricity problems and  $R = 0.80$  for the magnetism problems (two-way random effects, average value). Finally, with respect to the number of stacks, the external experts created about six stacks on average (electricity:  $M = 5.67$ ,  $SD = 1.15$ ; magnetism:  $M = 6.33$ ,  $SD = 0.58$ ), while in the design we had set on four clusters. Looking at the stack labels the experts gave (see the Appendix), it turns out that most of the labels could be matched to one of the experimenters’ categories, but that some of the experts made further subdivisions according to geometry or mathematical technique. For instance, one expert distinguished between Ampère’s law with surface current and Ampère’s law with spatial current. These subdivisions did not reveal any consistent pattern across individuals, however.

Taken together, these findings indicate that the combined judgments of experts and experimenters provide a reliable judgment of problem similarity. For most problem pairs there is agreement as to whether they “belong together” or not (see the Appendix for a full overview). For a few problem pairs, opinions vary. When it comes to judging the “expert-likeness” of students’ sortings, these ambiguous combinations will be left out of consideration.

#### B. Nature of novices’ categorizations

The next question is whether the novices’ problem sortings reveal the intended category structures. To answer this question, we considered the numbers of stacks, the clusters that emerged in a cluster analysis, and the types of stack names students gave. In order to make the results directly comparable to those of the experts, only the nonelaborated sorting tasks were included in this analysis.

TABLE V. Average numbers of stacks for all groups.

		Electricity	Magnetism
Experts ( $n = 3$ ) <sup>a</sup>	$M$ (SD)	5.67 (1.15)	6.33 (0.58)
Proficient students ( $n = 36$ )	$M$ (SD)	5.97 (1.50)	5.64 (1.51)
Less proficient students ( $n = 35$ )	$M$ (SD)	5.94 (1.73)	5.69 (1.49)

<sup>a</sup>The problem sets presented to the experts originally consisted of 23 problems (electricity) and 22 problems (magnetism), respectively. For this table, in order to make figures comparable across participant groups, only the problems that were kept in the final design have been taken into account.

The students, like the experts, on average created more stacks than the four intended in the design (Table V). The number of stacks was not significantly different across participant groups (electricity:  $F(2, 71) = 0.050$ ,  $p = 0.951$ ; magnetism:  $F(2, 71) = 0.303$ ,  $p = 0.739$ ). The number of valid observations is slightly lower than the number of participants, because some students had handed in incomplete results, for instance, by omitting a problem number from their written results.

To reveal patterns across participants, we used a hierarchical cluster analysis. Based on the sorting data, the procedure first computes the dissimilarity, or “distance,” between each pair of problems. After that, the two problems that are closest are taken together to form a cluster. In the next step the problems or clusters with the next smallest distances are taken together, and so on, until all problems

are linked together. The analysis was performed using the statistical software package SPSS. As a distance measure we used “Euclidean distance,” which is the most commonly used. As a linkage method, we used “between groups average linkage” as a robust multipurpose method [54]. The outcomes were interpreted using a dendrogram plot, in which the cluster structure is represented as a branching tree, with problems that were placed together more frequently being represented by a more close connection.

Figures 1(a)–1(d) present hierarchical cluster analyses for electricity and for magnetism problems, both for more and less proficient students. For the proficient students, if we consider the topmost four clusters in both sets, the clusters are in line with the design, except for two problems in the electricity set (E14 and E17) and two in

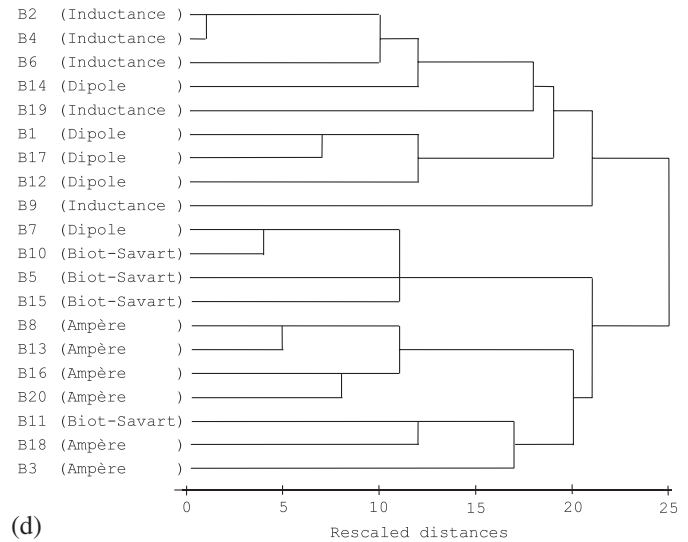
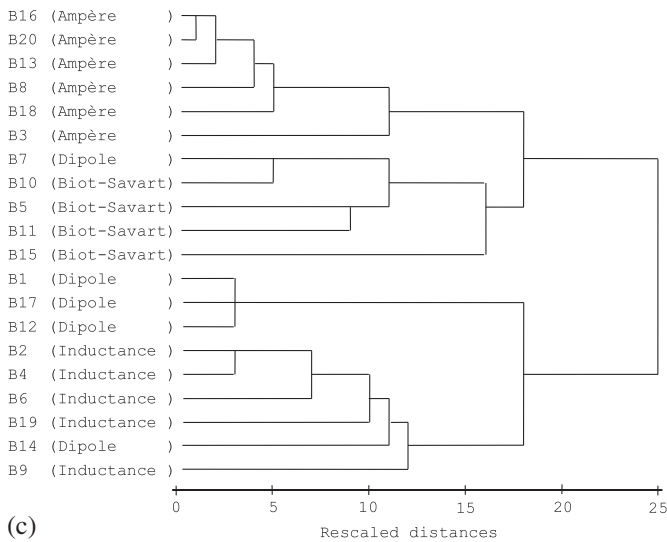
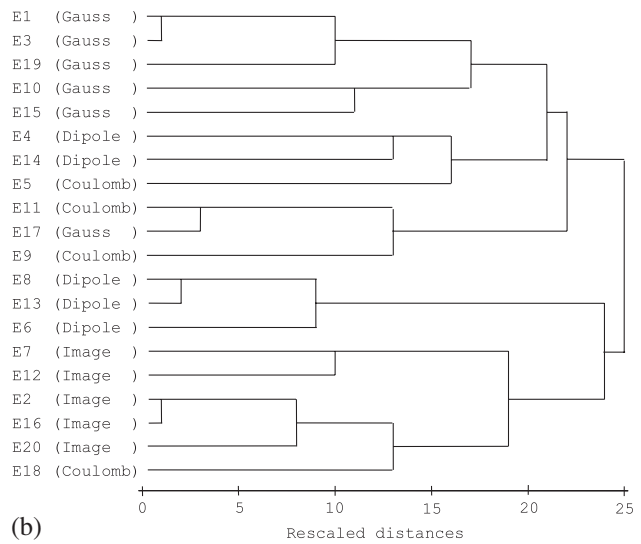
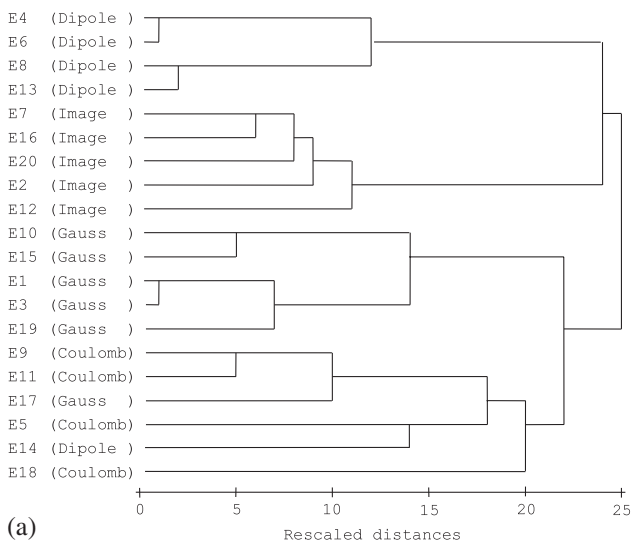


FIG. 1. (a) Cluster analysis, electricity, proficient students ( $n = 19$ ). (b) Cluster analysis, electricity, less proficient students ( $n = 16$ ). (c) Cluster analysis, magnetism, proficient students ( $n = 17$ ). (d) Cluster analysis, magnetism, less proficient students ( $n = 19$ ).

TABLE VI. Stack labels by a more and a less proficient student.

	Electricity	Magnetism
More proficient	Gauss Image charges Dipoles Non-Gaussian field computations	Dipoles Rotations in magnetic field Ampère's circuit law Biot-Savart integration Fluxes
Less proficient	Coulomb Dipole Capacitor Poisson (integral) Gauss Conductor	Flux Sum field Symmetry Magnetic moment Magnetic field

the magnetism set (B7 and B14). For the less proficient students, the appropriate number of clusters is less evident, and not all problem categories emerge from the analysis equally clearly. For the electricity problems, a five cluster interpretation yields three mismatched problems (E5, E17, and E18), with the “Coulomb” problems being dispersed over all clusters and the “dipole” category split in two. For the magnetism set, a four cluster interpretation yields two mismatched problems (B7 and B11) and a mix up of inductance and dipole problems. However, at a lower level in the tree, subgroups of inductance and dipole problem emerge as distinct clusters again. Thus, both for more and for less proficient novices, it turns out that in both sets most problems that were intended to be similar are linked together more closely than problems that were intended to be different.

In order to triangulate the outcomes of the cluster analysis, we looked into the names participants gave their problem stacks (see Table VI for an example). This analysis was conducted by two of the authors separately, after which differences were resolved by discussion. We distinguished labels that indicated a solution method (frequent examples are: Gauss' law, Gauss' law in differential form, image charge, dipole approximation, Ampère's law, Biot-Savart's law, sometimes with added specifications about symmetry, algebraic procedure. etc.); labels that only mentioned objects or quantities (e.g., moving charge, field), a geometrical property (e.g., planar symmetry), or an algebraic procedure (e.g., integration); and noncontent labels (e.g., insight, difficult, standard problem). In line with the outcomes of the cluster analysis, the majority of the labels indicated a solution method. On average the more proficient participants had a higher proportion of their labels indicating solution methods (64%) than the less proficient participants (49%),  $F(1, 78) = 7.3$ ,  $p = 0.008$ ,  $\eta^2 = 0.086$ .

### C. Quantifying the similarity to expert sorting

In order to quantify the gradual differences between novice and expert sortings, we had to express the (dis)

similarity of an individual sorting to an expert sorting in a single number. To this end, the most objective measure is directly based on combinations of individual problems. If the experimenters and at least two of the experts had put a pair of problems together, the combination was judged expertlike. If neither the experimenters nor any of the experts had put the two problems together, the combination was judged expert-unlike. All other combinations, which were made by some of the experts but not by all, were neglected in computing the expert-likeness. Thus, for each participant we had two scores per sorting:

$N_{i, \text{like}}$ : number of expertlike combinations subject  $i$  made,

$N_{i, \text{unlike}}$ : number of expert-unlike combinations subject  $i$  made.

For each set of cards we had two normalization parameters:

$N_{\text{max-like}}$ : maximum number of expertlike combinations,

$N_{\text{max-unlike}}$ : maximum number of expert-unlike combinations.

For the electricity problems  $N_{\text{max, like}} = 32$  and  $N_{\text{max, unlike}} = 123$ , leaving 35 possible combinations to be neglected. For the magnetism set,  $N_{\text{max, like}} = 18$ ,  $N_{\text{max, unlike}} = 131$ , and 41 to be neglected. The resulting “expert-likeness” score  $E$  for subject  $i$  was calculated from the following formula:

$$E_i = \frac{N_{i, \text{like}}}{N_{\text{max-like}}} - \frac{N_{i, \text{unlike}}}{N_{\text{max-unlike}}}.$$

A “perfect” sorting would yield the maximum score, “1”, a random sorting should give an outcome close to zero. To test the score's sensitivity to random variations, and to compare the performance of the subjects to chance, we generated a set of 1000 random sortings. In these sortings the number of stacks per sorting was distributed binomially with the average number of clusters set to 5.9, which corresponds to the average number in the real data (cf. Table V). Both the scores for real participants and the artificially generated scores are presented in Table VII.



TABLE VII. Means and standard deviations of the expert-likeness score per experimental group for real data and for computer-generated random data.

		Electricity		Magnetism	
		Without elaboration	With elaboration	Without elaboration	With elaboration
Experts	$M$ (SD)	0.73 (0.22)		0.83 (0.09)	
	$n$	3		3	
Proficient students	$M$ (SD)	0.36 (0.22)	0.58 (0.21)	0.46 (0.25)	0.57 (0.25)
	$n$	19	17	17	19
Less proficient students	$M$ (SD)	0.31 (0.14)	0.23 (0.14)	0.31 (0.20)	0.48 (0.21)
	$n$	16	19	19	16
Random data	$M$ (SD)	0.0012 (0.08)		0.0019 (0.10)	
	$N$	1000		1000	

The data confirm that the random variations are small compared to the real scores for both the electricity and the magnetism cards, and that both groups of novices do considerably better than chance. To ensure that the assumptions underlying the analysis of variance method (ANOVA) were met, we verified that neither the Kolmogorov Smirnov test of normality nor the Levene test for the homogeneity of variances indicated any significant deviations.

#### D. Effects of presenting elaborations

We expected problem-sorting performance to depend on student level, on the presence of elaborations, and on the interaction between both factors. We had planned to assess the effects of elaborations within subjects through a repeated-measures ANOVA. Because the data (Table VII) suggest that the effects of elaborations might be different for electricity and magnetism problem sets, a third experimental factor, “task version,” was included in the analysis. This factor produced significant interactions, which implies that the two tasks cannot be regarded as parallel tasks. Therefore, we will provide separate analyses for the electricity and the magnetism problems.

For the electricity problems, we found a significant positive main effect for student level, no significant main effect for the presence of elaborations, and a significant interaction between student level and the presence of elaborations, with a medium effect size (Table VIII). Within-group analysis revealed a significant positive effect of elaborations on proficient students’ scores,  $F(1, 67) = 11.6$ ,  $p = 0.001$ , and no significant effect on less proficient students scores,  $F(1, 67) = 2.92$ ,  $p = 0.092$ .

TABLE VIII. ANOVA with electricity problem-sorting score as dependent variable.

	$df$	$F$	$p$	$\eta^2$
Proficiency ( $P$ )	1	22.10	<.001	0.248
Elaboration ( $E$ )	1	2.26	0.138	0.033
$P \times E$	1	12.50	0.001	0.157
Within cells	67	(Mean square error = 0.032)		

TABLE IX. ANOVA with magnetism problem-sorting score as dependent variable.

	$df$	$F$	$p$	$\eta^2$
Proficiency	1	6.24	0.015	0.085
Elaboration	1	7.20	0.009	0.097
$P \times E$	1	0.37	0.546	0.005
Within cells	67	(Mean square error = 0.047)		

For the magnetism problems we found a significant positive main effect for student level, a significant positive main effect for the presence of elaborations, with a medium effect size, and no interaction between student level and the presence of elaborations (Table IX).

## IV. DISCUSSION AND CONCLUSION

Both the cluster analyses and the analysis of stack names confirmed that, for the domain of electricity and magnetism, physics students already start to know solution types in the initial phases of their studies. Nevertheless, the expert similarity scores indicate that students quite often fail to identify a suitable solution approach. The expert-likeness scores also confirm that the proficient novices did significantly better than the less proficient novices.

Our central hypothesis was that performance on the categorization task would improve if participants were supported to elaborate on the problem by getting a first elaboration to start with. For the magnetism problems, the expected main effect was confirmed. For the electricity problems, the difference was in the same direction, although the effect was not significant. Because the main effects were not significantly different across both tasks ( $t = 1.28$ ,  $p = 0.2$ ), we conclude, with some caution, that the hypothesis was confirmed.

Our final hypothesis was that the gain in performance would be greater for the more proficient novices, based on the idea that, in order to be useful, elaborations need to be

integrated in a coherent problem representation. The hypothesis was confirmed for the electricity problems, where the proficient students performed better with elaboration than without, whereas for the less proficient students there was no such effect. However, for the magnetism problems the interaction effect was not significant and the trend was in the opposite direction. Findings with regard to the interaction effect were significantly different across both tasks ( $t = 2.40$ ,  $p = 0.02$ ). Taken together, in this study proficient students performed significantly better with elaborations, regardless of task version, whereas for the less proficient students the evidence is inconclusive.

As a limitation of the current study it should be noted that, although the problem-sorting method is suitable to establish the effects of given elaborations, it provides only little insight into the ways these elaborations affect the reasoning process, and why one elaboration might be more effective than another. Nevertheless, given the significant difference between both task versions, we turned to our stimulus materials again in search of an explanation. Upon close inspection we found that in the magnetism problem set some elaborations contained words that might have supported a keyword strategy (for instance, elaborations for the “induction” problems containing the word flux). In the electricity set, we had been more successful in avoiding such keywords. As a consequence, we speculate that the electricity elaborations would only be useful if they could be integrated in a coherent problem representation. An alternative explanation for the effect of giving elaborations might be that they do not so much provide new information, but rather stimulate a greater depth of processing. To test this hypothesis a follow-up study could provide “placebo elaborations” (e.g., repeating information already present in the problem) or distracting elaborations (suggestive of an unproductive solution approach).

In this study both more and less proficient students turned out to know solution types, but they had difficulties matching problem situations to the right type. Furthermore, proficient students seem to induce a mental model of the

situation that can be used to integrate elaborations, whereas less proficient students may only benefit if they can extract a keyword that is directly linked to a proper solution approach.

Both the differences between our two tasks, and the differences between this study and some of the studies discussed in the Introduction, clearly illustrate how novices’ performance strongly depends on domain and task format. It therefore remains to be seen how these findings translate into different domains, such as mechanics. Furthermore, because failure as well as successful performance can occur for many reasons, performance data alone are not sufficient to gain insight into students’ knowledge and reasoning processes. Further research is needed to address the ways more or less proficient novices elaborate on situational features in their problem representations, and the ways this reasoning develops. This could be done, for instance, through think-aloud protocols and microgenetic approaches.

With respect to educational practice, our findings suggest little reason to have instructional strategies targeting a fundamental restructuring of knowledge. Rather, instruction should be aimed at students’ understanding of the situations and of the situational features that make a solution approach into a useful one. Promising approaches to this end are to present more context-rich and open-ended problems where a situational analysis cannot be avoided [55,56], to have the teacher act as a model demonstrating how particular elaborations can be helpful in the domain [57,58], and to have the students reflect on worked out solutions [25,59,60]. For many teachers, this could imply a shift in their selection of practice problems [61].

#### APPENDIX: FULL OVERVIEW OF PROBLEM CARDS AND EXPERT SERVINGS

See separate auxiliary material for full collection of problem cards and expert servings.

- 
- [1] E. Yerushalmi *et al.*, 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).
  - [2] J.H. Larkin *et al.*, Expert and novice performance in solving physics problems, *Science* **208**, 1335 (1980).
  - [3] R. Taconis, M. G. M. Ferguson-Hessler, and H. Broekkamp, Teaching science problem solving: An overview of experimental work, *J. Res. Sci. Teach.* **38**, 442 (2001).
  - [4] T.M. Foster, The development of students’ problem-solving skill from instruction emphasizing qualitative problem-solving, Ph.D. thesis, University of Minnesota, 2000.
  - [5] J.H. Larkin, in *Mental Models*, edited by D. Gentner and A.L. Stevens (Lawrence Erlbaum, Hillsdale, NJ, 1983), p. 75.
  - [6] M. T. H. Chi, P.J. Feltovich, and R. Glaser, Categorization and representation of physics problems by experts and novices, *Cogn. Sci.* **5**, 121 (1981).
  - [7] M.B. McVee, K. Dunsmore, and J.R. Gavelek, Schema theory revisited, *Rev. Educ. Res.* **75**, 531 (2005).

- [8] G. Cooper and J. Sweller, Effects of schema acquisition and rule automation on mathematical problem-solving transfer, *J. Educ. Psychol.* **79**, 347 (1987).
- [9] T. de Jong and M. G. M. Ferguson-Hessler, Knowledge of problem situations in physics: A comparison of good and poor novice problem solvers, *Learn. Instr.* **1**, 289 (1991).
- [10] J. H. Larkin and F. Reif, Understanding and teaching problem-solving in physics, *Int. J. Sci. Educ.* **1**, 191 (1979).
- [11] B. Rittle-Johnson and M. W. Alibali, Conceptual and procedural knowledge of mathematics: Does one lead to the other?, *J. Educ. Psychol.* **91**, 175 (1999).
- [12] P. B. Kohl and N. D. Finkelstein, Effects of representation on students solving physics problems: A fine-grained characterization, *Phys. Rev. ST Phys. Educ. Res.* **2**, 010106 (2006).
- [13] M. T. H. Chi *et al.*, Self-explanations: How students study and use examples in learning to solve problems, *Cogn. Sci.* **13**, 145 (1989).
- [14] E. R. Savelsbergh, T. de Jong, and M. G. M. Ferguson-Hessler, Situational knowledge in physics: The case of electrostatics, *J. Res. Sci. Teach.* **39**, 928 (2002).
- [15] L. Sutherland, Developing problem solving expertise: The impact of instruction in a question analysis strategy, *Learn. Instr.* **12**, 155 (2002).
- [16] R. Zajchowski and J. Martin, Differences in the problem solving of stronger and weaker novices in physics: Knowledge, strategies, or knowledge structure, *J. Res. Sci. Teach.* **30**, 459 (1993).
- [17] P. C. Peters, Even honors students have conceptual difficulties with physics, *Am. J. Phys.* **50**, 501 (1982).
- [18] M. Planinic *et al.*, Exploring alternative conceptions from Newtonian dynamics and simple DC circuits: Links between item difficulty and item confidence, *J. Res. Sci. Teach.* **43**, 150 (2006).
- [19] M. Saarelainen, A. Laaksonen, and P. E. Hirvonen, Students' initial knowledge of electric and magnetic fields—more profound explanations and reasoning models for undesired conceptions, *Eur. J. Phys.* **28**, 51 (2007).
- [20] J. Clement, Using bridging analogies and anchoring intuitions to deal with students' preconceptions in physics, *J. Res. Sci. Teach.* **30**, 1241 (1993).
- [21] A. A. diSessa, Toward an epistemology of physics, *Cogn. Instr.* **10**, 105 (1993).
- [22] J. P. Mestre, Implications of research on learning for the education of prospective science and physics teachers, *Phys. Educ.* **36**, 44 (2001).
- [23] M. Limon, On the cognitive conflict as an instructional strategy for conceptual change: A critical appraisal, *Learn. Instr.* **11**, 357 (2001).
- [24] A. Zohar and S. Aharon-Kravetsky, Exploring the effects of cognitive conflict and direct teaching for students of different academic levels, *J. Res. Sci. Teach.* **42**, 829 (2005).
- [25] D. Hestenes, Modeling games in the Newtonian world, *Am. J. Phys.* **60**, 732 (1992).
- [26] B. Rittle-Johnson, R. S. Siegler, and M. Wagner Alibali, Developing conceptual understanding and procedural skill in mathematics: An iterative process, *J. Educ. Psychol.* **93**, 346 (2001).
- [27] P. B. Kohl, D. Rosengrant, and N. D. Finkelstein, Strongly and weakly directed approaches to teaching multiple representation use in physics, *Phys. Rev. ST Phys. Educ. Res.* **3**, 010108 (2007).
- [28] H. A. Simon, *The Sciences of the Artificial* (MIT Press, Cambridge, MA, 1996).
- [29] W. Kintsch, *Comprehension: A Paradigm for Cognition* (Cambridge University Press, Cambridge, MA, 1998).
- [30] S. B. Blessing and B. H. Ross, Content effects in problem categorization and problem solving, *J. Exp. Psychol. Learn. Mem. Cogn.* **22**, 792 (1996).
- [31] P. T. Hardiman, R. Dufresne, and J. P. Mestre, The relation between problem categorization and problem solving among experts and novices, *Mem. Cogn.* **17**, 627 (1989).
- [32] L. R. Novick, Analogical transfer, problem similarity, and expertise, *J. Exp. Psychol. Learn. Mem. Cogn.* **14**, 510 (1988).
- [33] R. Braune and W. R. Foshay, Towards a practical model of cognitive information processing: Task analysis and schema acquisition for complex problem-solving situations, *Instr. Sci.* **12**, 121 (1983).
- [34] G. Polya, *How to Solve It* (Princeton University Press, Princeton, NJ, 1945).
- [35] K. VanLehn, in *Foundations of Cognitive Science*, edited by M. I. Posner (MIT Press, Cambridge, MA, 1989), p. 527.
- [36] B. Sherin, Common sense clarified: The role of intuitive knowledge in physics problem solving, *J. Res. Sci. Teach.* **43**, 535 (2006).
- [37] J. H. Larkin, in *Cognitive Skills and Their Acquisition*, edited by J. R. Anderson (Lawrence Erlbaum, Hillsdale NJ, 1981), p. 311.
- [38] D. Hammer *et al.*, in *Transfer of Learning: Research and Perspectives*, edited by J. Mestre (Information Age, Greenwich, CT, 2005), p. 89.
- [39] C. A. Kaplan and H. A. Simon, In search of insight, *Cogn. Psychol.* **22**, 374 (1990).
- [40] J. Metcalfe and D. Wiebe, Intuition in insight and non-insight problem solving, *Mem. Cogn.* **15**, 238 (1987).
- [41] A. D. de Groot, *Thought and Choice in Chess* (Mouton, The Hague, 1965).
- [42] N. Huet and C. Mariné, Clustering and expertise in a recall task: The effect of item organization criteria, *Learn. Instr.* **15**, 297 (2005).
- [43] M. T. H. Chi *et al.*, How inferences about novel domain-related concepts can be constrained by structured knowledge, *Merrill-Palmer Quarterly* **23**, 27 (1989).
- [44] R. J. Dufresne *et al.*, Constraining novices to perform expertlike problem analyses: Effects on schema acquisition, *J. Learn. Sci.* **2**, 307 (1992).
- [45] M. L. Gick and K. J. Holyoak, Schema induction and analogical transfer, *Cogn. Psychol.* **15**, 1 (1983).
- [46] A. H. Schoenfeld and D. J. Herrmann, Problem perception and knowledge structure in expert and novice mathematical problem solvers, *J. Exp. Psychol. Learn. Mem. Cogn.* **8**, 484 (1982).
- [47] J. Reitman Olson and K. J. Biolsi, in *Toward a General Theory of Expertise: Prospects and Limits*, edited by K. A. Ericsson and J. Smith (Cambridge University Press, Cambridge, England, 1991), p. 240.

- [48] J. M. Royer, C. A. Cisero, and M. S. Carlo, Techniques and procedures for assessing cognitive skills, *Rev. Educ. Res.* **63**, 201 (1993).
- [49] D. P. Birney, G. J. Fogarty, and A. Plank, Assessing schematic knowledge of introductory probability theory, *Instr. Sci.* **33**, 341 (2005).
- [50] S. Kalyuga, Rapid cognitive assessment of learners' knowledge structures, *Learn. Instr.* **16**, 1 (2006).
- [51] J. E. Gómez, A. Benarroch, and N. Marín, Evaluation of the degree of coherence found in students' conceptions concerning the particulate nature of matter, *J. Res. Sci. Teach.* **43**, 577 (2006).
- [52] T. de Jong and M. G. M. Ferguson-Hessler, Cognitive structures in good and poor novice problem solvers in physics, *J. Educ. Psychol.* **78**, 279 (1986).
- [53] P. E. Shrout and J. L. Fleiss, Intraclass correlations: Uses in assessing rater reliability, *Psychol. Bull.* **86**, 420 (1979).
- [54] M. S. Aldenderfer and R. K. Blashfield, *Cluster Analysis* (Sage, Newbury Park, CA, 1984).
- [55] D. Fortus *et al.*, Design-based science and real-world problem-solving, *Int. J. Sci. Educ.* **27**, 855 (2005).
- [56] J. Park and L. Lee, Analysing cognitive or non-cognitive factors involved in the process of physics problem-solving in an everyday context, *Int. J. Sci. Educ.* **26**, 1577 (2004).
- [57] K. Bielaczyc, P. Pirolli, and A. L. Brown, Training in self-explanation and self-regulation strategies: Investigating the effects of knowledge acquisition activities on problem solving, *Cogn. Instr.* **13**, 221 (1995).
- [58] B. L. Sherin, How students understand physics equations, *Cogn. Instr.* **19**, 479 (2001).
- [59] B. Rittle-Johnson and J. R. Star, Does comparing solution methods facilitate conceptual and procedural knowledge? An experimental study on learning to solve equations, *J. Educ. Psychol.* **99**, 561 (2007).
- [60] K. L. Malone, Correlations among knowledge structures, force concept inventory, and problem-solving behaviors, *Phys. Rev. ST Phys. Educ. Res.* **4**, 020107 (2008).
- [61] E. Yerushalmi *et al.*, Instructors' reasons for choosing problem features in a calculus-based introductory physics course, *Phys. Rev. ST Phys. Educ. Res.* **6**, 020108 (2010).