JOURNAL OF RESEARCH IN SCIENCE TEACHING VOL. 39, NO. 10, PP. 928–951 (2002)

Situational Knowledge in Physics: The Case of Electrodynamics

Elwin R. Savelsbergh[1,2,*] Ton de Jong,[2] Monica G.M. Ferguson-Hessler[1]

[1]*Faculty of Technology Management, Eindhoven University of Technology,*
P.O. Box 513, NL-5600 MB, Eindhoven, The Netherlands

[2]*Faculty of Educational Science and Technology, University of Twente,*
P.O. Box 217, 7500 AE, Enschede, The Netherlands

Received 20 July 2000; Accepted 1 May 2002

Abstract: Major difficulties for a novice physics problem solver are how to interpret new problems and how to combine information given in the problem with information already known. A domain expert, by contrast, has the knowledge to take full advantage of problem features at a glance. It takes a long period of practice to acquire such situational knowledge, and it would be desirable for this to be taught more effectively. As a first step, this requires information on how situational knowledge differs across individuals of different competence levels. Related research on mental models and problem representations does not give a direct view on the knowledge subjects have of situations before being confronted with the problem. To assess situational knowledge more directly, we asked participants to respond to physics formulas (from the field of electrodynamics) by describing relevant problem situations. We compared physics problem descriptions by experts ($n = 6$) and by proficient ($n = 6$) and less proficient ($n = 6$) novices. We analyzed the situations that were described at the levels of words, sentences, and complete descriptions. Results indicate that competence is related to the structure of problem situations rather than the use of particular concepts, and that the differences in the use of multiple representations are more prominent than differences in the use of one specific kind of representation. Results also indicate that the differences between experts and novices are along different dimensions than the differences between more and less proficient novices. Implications for teaching are discussed. © 2002 Wiley Periodicals, Inc. J Res Sci Teach 39: 928–951, 2002

One reason why experts are better problem solvers than novices is that experts have the knowledge to make productive use of the particular features of a problem. Much of this knowledge is acquired implicitly through extensive practice problem solving. In some domains—for instance, in the domain of chicken sexing—explicit instruction about features to look for appeared

---

**Present address*: Center for Science and Mathematics Education, University of Utrecht, P.O. Box 80000, NL-3508 TA Utrecht, The Netherlands.

Correspondence to: E.R. Savelsbergh; E-mail: e.r.savelsbergh@phys.uu.nl

DOI 10.1002/tea.10057

effective (Biederman & Shiffrar, 1987). However, in physics, as in many conceptual domains, the features that matter generally are second-order ones (e.g., Chi, Feltovich, &, Glaser, 1981). Whereas the superficial features in the problem description can take many forms, one needs to infer features at a deeper level with the use of relevant background knowledge. Understanding why and how a feature is important thus involves conceptual understanding of the problem. In such domains an adequate mental representation of the problem is the first requirement for successful problem solving.

It has been widely recognized that novices' lacking representations pose an impediment to their problem-solving proficiency (Chi et al., 1981; De Jong & Ferguson-Hessler, 1991, 1996; Larkin, 1983; Newell & Simon, 1972). Unsurprisingly, novices have trouble identifying deep features and they find it difficult to connect to an expert explaining the use of deep features. This makes it difficult to teach about the way problem features have an influence. Several researchers have attempted to improve students' problem representations in varying domains. Some taught explicit strategies to analyze the problem (e.g., De Jong & Ferguson-Hessler, 1984; Huffman, 1997; Mettes, Pilot, & Roossink, 1981); others had the students discuss problem features while or after solving the problem (e.g., Taconis, Van Hout-Wolters, & Ferguson-Hessler, 1999; Thompson, Gentner, & Loewenstein, 2000; Wells, Hestenes, & Swackhammer, 1995). Although some researchers, especially in the latter group, report marked performance improvements, it is unclear why and in what respect some of these interventions are more successful than others. It seems that it does not help much to learn how problems should be analyzed in general; rather, a student needs domain-specific knowledge about how problem features matter in a given domain.

Unfortunately, despite a broad consensus about global characteristics of proper representations, there is much less clarity about the content of such representations in specific domains, and about the knowledge students need to construct better representations themselves. Moreover, most of the available research relates to the knowledge of experts or to differences between experts and novices. It is unclear whether more and less proficient novices differ along the same dimensions that are characteristic of the difference between experts and novices (Willson, 1990). Therefore, a better understanding of what makes one representation better than the other and what knowledge is involved in constructing a problem representation could help us understand the problem-solving process and could suggest ways to improve problem-solving assignments and tutoring.

When people try to solve a given problem, they use knowledge they have about previous, similar problem-solving experiences to construct an initial representation of the problem. They use their prior knowledge to select relevant information from the problem statement, and they add information from their prior knowledge (e.g., Braune & Foshay, 1983). Thus, they develop a mental representation of the problem, which determines further problem-solving actions they take. These actions in turn may result in a modified representation of the problem that changes the course of reasoning. Thus, a mental representation is not just a passive structure resulting from the analysis of the problem, but rather it changes continually throughout the problem-solving process.

We use the term *situational knowledge* to designate the prior knowledge that provides the outline and the building blocks for mental representations. Elements of situational knowledge include isolated situation features, conditions for the application of a solution procedure, and complete problem templates including situational features and an unknown. To be useful, situational knowledge should give access to other knowledge, such as solution methods or physics principles. This integrated knowledge structure has become known as a problem-type schema (Hinsley, Hayes, & Simon, 1977).

Our work focuses on identifying situational knowledge, which makes it different from work on mental representations of given problems, on external representations, or on the problem-solving process. Situational knowledge has some distinctive features, as it is stored in long-term memory, and in general it goes beyond a particular given problem. Yet, how a person mentally represents a problem and how one solves the problem will strongly depend on one's situational knowledge; and the other way round, how one proceeds to solve a problem influences what situational features will be remembered. Therefore, characteristics of problem representations and the problem-solving process derived from expert–novice studies can help assess situational knowledge.

Differences Between Experts' and Novices' Situational Knowledge

We will review differences between experts and novices and infer how these differences relate to situational knowledge. Table 1 presents an overview of differences in situational knowledge that we conjectured based on the literature. The table distinguishes between differences in the content of the knowledge and differences in the structure of the knowledge. *Content* refers to what elements of knowledge there are, whereas *structure* refers to how these elements are connected. These aspects are by no means independent, however.

There has been a long tradition to distinguish between abstract and concrete reasoning (e.g., Piaget, 1970). Concepts in many domains can be ordered hierarchically. People tend to learn the concrete concepts first and acquire the abstract concepts later (Lawson et al., 2000; Reif, 1983; Reif & Heller, 1982; Van Hiele, 1986). In general, the observable objects in a situation will be more concrete than nontangible objects and properties. It could thus be expected that in the situational knowledge of experts abstract entities are more prominent than they are in the knowledge of novices. A problem with this kind of distinction is that it is often unclear what is abstract and what is concrete, and to whom (Smith, diSessa, & Roschelle, 1993).

A better-defined difference, which occurs in several domains, is that novices tend to apply time-based mental simulations whereas experts tend to apply constraint-based reasoning [Goei, 1994; Larkin, 1983; Stenning, 1992; but see Clement (1994) for an account of how experts resort to time-based reasoning in unfamiliar situations]. The concepts that comprise a representation determine whether it supports time-based or constraint-based reasoning (diSessa, 1993; Fuson & Carroll, 1996; Slotta, Chi, & Joram, 1995). In physics, experts might represent a process in terms of quantities that remain constant (such as energy) rather than in quantities that vary (such as velocity). Also, in static situations such as those encountered in electrostatics, a given configuration (such as an arrangement of spheres) can be redescribed in terms of constraints (such as spherical symmetry) that help simplify the problem.

Table 1
Proposed characteristics of situational knowledge

	Mainly Novice		Mainly Expert
Content			
	Phenomenological/concrete	vs.	abstract entities
	Changing/time based	vs.	constraint based
	Topological and functional	vs.	geometrical relations
	Question	vs.	givens
	Numerical	vs.	qualitative specifications
Structure			
	Fragmentary	vs.	coherent
	Tree structure, single inference source	vs.	graph structure, redundant inference sources
	Situation and solution separate	vs.	integrated solution information
	Diffused	vs.	localized properties
	Single	vs.	multiple redundant representations

Larkin (1983) signaled a closely related difference. She presented protocols by a novice and an expert, each interpreting a mechanics situation with carts and weights and pulleys. The novice searches for topological and functional relations, such as what is on top, what is going which way, and what is pulling where, struggling to make sense of the situation. The expert, by contrast, immediately recognizes the situation and proceeds to look for geometrical relations (e.g., symmetries), which are useful to simplify the problem. Again, this difference could be expected to reflect people's situational knowledge in that experts have more knowledge regarding geometrical properties and their consequences.

As a more domain general difference, it has been claimed that novices have a preference for working backward, whereas experts work forward from the givens [Larkin, McDermott, Simon, & Simon, 1980; Sweller, Mawer, & Ward, 1983; but see Zajchowski & Martin (1993)]. This could be reflected in the question part of the problem being more prominent in novices' situational knowledge. However, there is also evidence to the opposite: Anderson (1983) stressed a clear focus on the goal as an important ingredient for successful problem solving, and De Jong & Ferguson-Hessler (1991) found that in a problem-retrieval task, successful novices were more focused on the goal than were less successful novices.

Finally, several authors found expertise-dependent differences in the use of quantitative and qualitative representations (Chi, Glaser, & Rees, 1982; Larkin, 1983; McDermott & Larkin, 1978; McMillan & Swadener, 1991). Experts have a rich qualitative knowledge which is used to modify and extend initially constructed representations throughout the problem-solving process. Weak problem solvers' qualitative knowledge is incomplete or incorrect. Consequently, they have to resort to weak problem-solving methods with quantitative representations. This could be reflected in novices' situational knowledge being more focused on numerical quantities than experts' knowledge.

In addition to these differences in content, several authors have suggested differences in the ways mental representations are structured. Novices tend to have fragmentary knowledge, where pieces of information are small and only weakly connected, whereas in coherent expert knowledge the relations between objects form an integrated part (diSessa, 1993; Hammer, 1994).

Novices also tend to use their knowledge in a linear fashion in which each inference follows from another statement in only one way, whereas experts tend to use a network structure in which inferences can be reached along different pathways (Larkin, 1983). A different perspective was taken by Chi et al. (1981), who found evidence that although novices do have schemata: "[Their] schemata may be characterized as containing sufficiently elaborate declarative knowledge of a potential problem, but lacking abstracted solution methods."

Not only the number of connections but also the ways in which objects and their properties are connected may differ between individuals. It has been claimed that the entities in an expert representation have localized attributes (Larkin, 1983; cf. de Kleer & Brown, 1981, 1983):

> [T]he entities in a physical representation have localized attributes, that is one doesn't learn any more about the entity by considering the context in which it appears. This is not the case for entities in naive representations. For example an attribute of a toboggan in a naive representation might well be that it goes down hills. However, if a toboggan does not go down a hill, the fault may be that something is wrong with the toboggan (it's improperly waxed, or full of splinters); but the fault may also be outside the toboggan (the snow is wet). Thus an attribute of the toboggan (it goes down hills) can be violated by changes not in the toboggan, but in its environment. I think this is never the case with physical entities. Nothing in the environment can change any of the attributes of the force. (Larkin, 1983, p. 80)

Thus, a localized property should refer to a specific element of the situation only. If it is posed like this, many powerful problem-solving approaches in physics that refer to properties of the situation as a whole (e.g., conservation of energy, or spatial symmetries) do not satisfy the requirement. Nonetheless, also in these cases a difference between novices' and experts' knowledge could be expected: namely, that the experts attribute the global property to more specific features in the situation. As an example from a different domain, consider how much younger students in elementary geometry develop an understanding of squareness. Initially the students may recognize an object as square, but they do not attribute this property to particular features of the object; rather, it would be a fuzzy property of the entire object. Later on the students acquire a more advanced understanding to explain what makes the object square: namely, its having straight angles and equal sides; and later on, to prove the squareness of an object from its properties (Van Hiele, 1986).

Specific to the domain of physics is that many physics problems can have several formally equivalent representations. In mechanics, a situation can be represented in terms of forces or in terms of energy; in geometrical optics, the bending of light can be described in terms of rays or in terms of wave fronts; and in electrodynamics, an equipotential surface can be replaced by field lines or a charge by its field flux through an appropriate surface. These representations are formally equivalent in the sense that within a range of application they are equally valid. Nevertheless, each may trigger particular conclusions that are not drawn straightforwardly from the others, so that an alternative representation can promote a different course of reasoning. Given that a problem can have multiple representations, and that different representations support different types of reasoning, taking a different perspective could help one resolve an impasse. Polya (1945) therefore adopted restating the problem as one of his problem-solving heuristics. More recently, cognitive flexibility theory (Jacobson & Spiro, 1995) has emphasized the importance of taking multiple perspectives to problem solving in ill-defined domains. In physics, unlike in a typical ill-defined domain such as ethics, these multiple perspectives are formally redundant. In mechanics, for instance, a description of a motion in terms of forces and one in terms of energy can both convey the same information. Even so, each may trigger new approaches to simplify and solve the problem. As a first requirement for such flexible reasoning, one has to know alternative representations. Second, one needs to know the connection among the representations.

Research Questions

As becomes clear from the previous sections, there are a variety of claims regarding the differences in problem representations and the problem-solving processes of novices and experts, and several of these differences could be related to differences in situational knowledge (Table 1). However, as there is little empirical research to provide a direct insight in situational knowledge, evidence for these differences is mostly inferred from reasoning processes, cued recall experiments, and so forth. Moreover, there are some open questions. First, although the differences in Table 1 are structured as contrasts between experts and novices, on several aspects—such as attention to question versus givens—the literature provides no unambiguous cues as to which is more prominent in experts' situational knowledge and which is more prominent in novices' knowledge. Second, whereas differences between experts and novices can give some hints for instruction, differences between more and less proficient novices could have more direct relevance to instruction. Differences of the latter kind have received far less attention in the literature, and it is unclear whether these can be captured along the same dimensions. Finally, several of the differences discussed in the previous paragraph were derived from mechanics, and it remains to be seen how important these aspects are across different subjects in physics. In our view, this situation calls for open comparisons among the situational knowledge of experts and more and less proficient novices. In this study, we therefore explore the knowledge that individuals at different competence levels possess relative to problems in electrodynamics. Specifically, we examine how this knowledge is structured and how it is connected to knowledge of concepts and solution procedures.

Method

The Subject Matter

Although this is not stressed in the cognitive literature we drew on in the previous section, the particularities of expert reasoning will vary depending on the domain, and within the domain depending on the subject matter (e.g., Glaser, Schauble, Raghavan, & Zeitz, 1992) Several of the more well-known studies on problem solving and reasoning in conceptual domains have been conducted in the field of mechanics (e.g., Chi & Bassok, 1989; Chi et al., 1981; diSessa, 1993; Larkin, 1983; McCloskey, 1983). We consider it important to broaden the scope of research. Therefore, we used an electrodynamics course from the first-year university physics curriculum as our content matter. The electrodynamics domain shares several essential characteristics with other physics domains. First, it is a conceptual domain, which implies that many of the concepts in the domain are not directly perceivable; they are more or less abstract objects. Second, it is a formal domain, which means there is a limited set of well-defined concepts and laws that can be applied in a limited and well-defined range of problems. The third characteristic shared by electrodynamics and other physics domains is that it supports different ways for describing a problem. Although these different representations are formally equivalent, one way may be more apt in a given situation than another. An example from mechanics is a motion that can be described in terms of forces or in terms of energies. In electrodynamics, to describe the field that a spherical charge distribution induces at a certain point in space one can explicitly consider the individual contributions of all parts of the distribution, or one can refer to the flux through a Gauss surface. The existence of explicit alternative formulations distinguishes physics from a domain such as medicine where, for instance, sweating, pallor, and hyperventilation can be the fuzzy symptoms of a shock, but there is no one-to-one equivalence.

Besides commonalities there are also distinctions between electrodynamics and other physics domains such as mechanics. One distinguishing feature could be a smaller amount of naive ideas directly related to the subject. Mechanics is more closely related to everyday experience. Therefore, one could expect peoples' naive ideas concerning mechanics and motion to be substantially more established than their ideas about electrodynamics. Another distinguishing feature is that electrodynamics requires an emphasis on symmetry from the very beginning, whereas in initial mechanics courses, for instance, the use of symmetries and especially the use of congruence is not profound, and sometimes even misleading (diSessa, 1993). This is important because simplifying problems on the basis of symmetries is a form of constraint-based reasoning.

Participants

As novices we selected 6 proficient and 6 less proficient first-year physics students who had just completed their Electricity and Magnetism course and who had attempted the final test. The course covered electrodynamics, including electrostatics and magnetostatics, all limited to systems in vacuum. According to the syllabus, the course should represent a workload of about 80 hours to an average student. The students were selected from a population of 103 first-year physics students. They were classified as proficient or less proficient on the basis of their grades from the regular electricity and magnetism test and from the mechanics test completed some weeks earlier. On the basis of this information, we selected 6 proficient students and 6 less proficient students for participation in the experiment. Proficient participants were recruited from those who had scored "amply sufficient" or better (a grade of ≥ 7 on a 10-point scale) on the electricity and magnetism test, and who, in addition, had scored "sufficient" or better (a grade of ≥ 6) on their mechanics test a month earlier. Less proficient participants were recruited from those who had scored "low" to "insufficient" (a grade of 3–4) on the electricity and magnetism examination, and who, in addition, had scored no better than "insufficient" on the mechanics test. Before the experiment started, students were asked for the university math grades they had obtained and for their final high school grades in the sciences. This set, nine grades in total, showed a high internal coherence (Cronbach $\alpha > .90$, $n = 10$). The ranking of students according to this scale was consistent with the division into more and less proficient participants based on grades for Electricity and Magnetism, and Mechanics. Students were paid 6 Euro for their participation.

Six experts were selected from two different subgroups: 3 lecturers and 3 doctoral students. The lecturers are domain experts in a narrow sense and the doctoral candidates are more typical of the kind of expertise students are likely to develop. It is common not to distinguish among these subgroups (e.g., Chi, 1981; Larkin, 1983). However, studies in medicine, for instance (Boshuizen & Schmidt, 1992) suggest that the two kinds of experts may differ qualitatively. Therefore, we will report results for both subgroups. One should be cautious, however, in interpreting the differences. Whereas comparisons between experts and novices, and between more and less proficient novices can be interpreted with some confidence, the small number of participants per expert subgroup does not permit strong conclusions.

Problem Construction

A variety of methods for assessing problem representations can be found in the literature. Royer, Cisero, & Carlo (1993) provided an extensive review of such methods. In their review, the sections pertaining to the assessment of depth of representation and the assessment of mental models best match our research question. The main techniques discussed in their section on depth of representation are reproduction tasks, problem-sorting tasks, and "asking learners to make judgments about problems or situations." The techniques for assessing mental models are all based on asking the participant to predict or infer something from a given situation description. None of these techniques directly assesses situational knowledge.

A straightforward approach to assessing situational knowledge would be to have participants think aloud while reading and interpreting the problem. This would also give insight into the process of constructing the representation. However, with such an approach, one cannot distinguish between information retrieved from long-term memory and information given in the problem description. Therefore, this approach is less appropriate as an assessment tool for the situational knowledge. De Jong & Ferguson-Hessler (1991) used a problem-recall task with very short presentation times to assess situational knowledge. With such a setup, participants can hardly memorize anything, and most information produced would come from long-term memory. With this technique, which they refer to as problem reconstruction, the stimulus problem serves only as a cue to activate situational knowledge and problem schemas involved in long-term memory. Nevertheless, participants may remember fragments, and, more important, the stimulus provides a cue to reconstructing the problem in a particular format.[1]

In the present study, we expand on prior research by using a different cue that requires participants to construct the entire problem representations. Because situational knowledge is connected to solution information, we expected that participants prompted with solution information might be able to indicate the situation when the information can be applied. Because the connection between situation and solution information has a preferred direction (Anderson, 1983), the information triggered can be incomplete. To obtain a more complete image of their situational knowledge, participants can be prompted with several forms of solution information. Now, as this task reverses the usual working order, it disrupts automaticity, and we may fruitfully use a think-aloud approach (Ericsson & Simon, 1993). Our approach differs from the approach of Ericsson and Simon, however, as our interest is in the content of the evoked descriptions rather than the process. Based on the above considerations, we made the stimulus design shown in Figure 1. In this design the stimulus consists of one physics law, expressed in a formula, plus a keyword referring to a subset of the problems for which the formula can be used. The keyword was included to make the participants go beyond literal recall of textbook examples. We selected eight laws and formulas that together covered a coherent part of the subject matter. These eight formulas were used to construct a total of 20 different assignments. In total, every formula was presented two or three times, with a different keyword each time, to obtain a more complete coverage of the participants' knowledge. The complete list of formulas and keywords is presented in Table 2.

Procedure

Each assignment was presented on a separate sheet of paper. After the first cycle of eight different formulas, the same formulas reappeared, this time joined by different keywords. Before turning the first page of the assignments, participants were instructed to start thinking aloud immediately and not focus on the correctness of what they said. They were instructed that they could ask if they did not recognize a symbol. Then they had to turn the page and start working on the first case.

$$F = \frac{1}{4\pi\varepsilon_0}\frac{q_1 \cdot q_2}{r^2}$$

Surface charge

a. What does the formula mean?

b. What does the keyword mean?

c. What kind of problem comes to mind?

d. Describe—in brief—a situation in which both formula and the keyword apply.

e. Describe—again in brief—a second situation in which both the formula and the keyword apply, and which differs from the first situation as much as possible.

f. Explain what is different and what is similar in the two cases.

Figure 1. One of the stimuli used in the experiment was Coulomb's law combined with the keywords *surface charge* (translated from Dutch).

Table 2
Assignments: Formulas and keywords

Formula	Keyword
1. $\vec{F} = \frac{1}{4\pi\varepsilon_0}\frac{q_1 \cdot q_2}{\lvert r\rvert^3}\vec{\mathrm{r}}$	a) Superposition principle b) Surface charge c) Symmetry
2. $\varepsilon_0 \oiint \vec{E} \cdot d\vec{A} = Q_{\mathrm{enclosed}}$	a) Surface charge b) Spherical symmetry
3. $dB_{\mathrm{p}} = \frac{\mu_0 I(d\vec{1} \times \vec{r})}{4\pi\lvert r\rvert^3}$	a) Surface current b) Conducting wire
4. $\oint \vec{B} \cdot d\vec{s} = \mu_0 I_{\mathrm{enclosed}}$	a) Large surface b) Cylindrical symmetry
5. $W = \frac{U^2\varepsilon_0 A}{2d}$	a) Force b) Displacement c) Constant voltage
6. $V(\vec{r}) - V(\vec{r'}) = \int_{\vec{r'}}^{\vec{r}} \vec{E} \cdot d\vec{s}$	a) Inhomogeneous field b) Spatial charge
7. $F_i\vec{\mathrm{r}} = q(v\vec{v} \times \vec{B})$	a) Electric field b) Conducting wire c) Voltage
8. $\varepsilon_{\mathrm{ind}} = \frac{d\Phi_{\mathrm{m}}}{dt}$	a) Motion b) Mutual induction c) Induced current

For each case, after reading the formula and keyword, participants were asked to answer the same series of questions. First, to ensure they took notice of both, they had to explain the meaning of the formula and keyword. Then they were asked to describe the type of problem corresponding to the stimulus. This question was included to stimulate the participants to describe any generalized representation they might have. Next, they were asked to describe two different situations that would involve the formula and keyword, and as a final question they were asked to summarize differences and similarities between the two situations. The aim of the final question was to give some insight into the relative importance the participants attributed to the different elements in their problem description. The first case (Table 2, Case 1a) was for training only. An analysis of some pilot data suggested that after this brief training participants' mastery of the task was sufficient to produce acceptable problem descriptions. The tasks of the experimenter were to keep participants talking, remind them every now and then to answer the questions for each case in the intended order, and when asked, tell them the meaning of a symbol (e.g., that Φ stands for flux).

No time limits were imposed in this experiment. All participants were thus able to finish the complete set of assignments. On the average the interviews took 48 minutes, with a standard deviation of 11.0 minutes. Times did not differ significantly across groups, $F(2, 17) = .22, p = .81$.

Data Handling

Because the properties of problem representations are related to different scale sizes (the entire problem, the sentence, or the word level), there is no single way to analyze the situation descriptions in such a way that all factors of interest are captured. Therefore, we used different approaches for the different levels of aggregation: a word-level approach, a sentence-level approach, and a situation-level approach. To do these analyses we first made transcripts of all problem descriptions, except those for the training case, by all participants.

For the analysis of word use, we made an inventory of word occurrences. For this inventory we used all participant texts. First we removed the text spoken by the experimenter. Then for each word the total number of occurrences in the protocol was determined. These frequencies were then used to determine which words were most used by which group.

For the analysis of sentences we used all problem descriptions again. First the transcripts had to be segmented. Sentences were defined as the smallest meaningful units in the text. As an operational definition we decided to mark the end of a segment at each punctuation mark and at each occurrence of the words *or* or *and*. Although this procedure could be sensitive to variations in transcription style, when the result was checked it appeared that the resulting segments made sense and only minor corrections were needed. Then we developed, through several revisions, a coding scheme to represent the functions a sentence can have in the problem description. The reliability of the coding was determined by having two raters score a small portion of the data. When the coding was found to be reliable, the entire data were coded by a single rater.

For the global judgments about problem descriptions, we chose a set of four cases that were analyzed for all participants. This was done because we felt that more than four judgments per participant would add little in terms of reliability, and also because this interpretative analysis was laborious. The cases were Gauss' law, and Ampère's law in integral form, each used twice with different keywords (Table 2, Cases 2a, 2b, and 4a and 4b, respectively). These are central laws, one related to electrostatics, the other to magnetism. The following aspects were evaluated: the level of generality versus instantiatedness; the coherence of the descriptions; whether any alternative situations were mentioned; and whether differences between situations or between solution methods were explicitly addressed. Two judges rated the selected cases on all aspects, after which differences were resolved by discussion.

Results

All participants were able to describe a physical situation in response to most of the presented formulas. However, the descriptions varied greatly among participants. In the following sections, we first present the results of the analysis at the word level, followed by the analysis of sentences, and the situation-level analysis.

Words

We expected differences in the content of the situational knowledge to be reflected at the level of single words. Therefore, we made an inventory of the 70 most frequently used physics-related words in all participant texts. The rationale behind the particular choice of the number 70 is pragmatic: We thought it necessary for the list to cover at least two thirds of the physics word occurrences to make the conclusions representative for all physics words. Even when we adopted a wide definition of physics words, the top 70 cover about 80% of the total physics word occurrences. The 70th word was used 28 times in total, which is about 1.5 times on average by each participant. All words on the list were used by at least three participants, which indicates there are no highly idiosyncratic words on the list. As a next step, we computed for each participant how often a word was used as a proportion of all top 70 words used by the participant. On the basis of these proportion scores, each word was allotted to the competence level where it occurred most frequently on average. We then tried to capture the differences among groups by categorizing the types of words, using an inductive approach. Because we were considering the words out of context,

Table 3
Classification of word type by user

		Expert (Lecturer)	Expert (Ph.D. Candidate)	More Proficient Student	Less Proficient Student
Physical	Physics quantities		Field Work Electric Potential Magnetic Magnetic field Energy	Flux	Field intensity Current
	Abstract object	Spatial charge Surface charge	Electron Charges Conductor	Point charge	
	Tangible object		Surface current	Stick Loop/circuit Capacitor	Plate Coil Wire Particle
	Interactions	Inductance Lorentz force Induced current Force		Potential difference	Tension
Spatial	Geometry	Symmetry Spherical symmetry Cylinder Sphere Plane	Round		
	Topology	Closed		Inside	Through Enclosed Between Outside
	Spatial quantities			Radius Area	Distance Place Velocity
Change	To change	To shift	To move	To change To arise	To cause To act To run
	No change	Zero Constant Homogeneous	Large	Infinite	
Other		Space Direction Charge	To be located Integral Surface	Charged Point	

categories were based on the meanings the words have to us as physicists. The result of our categorization is presented in Table 3. For convenience, we have structured our categories in three rubrics: physical, spatial, and change.

In the physical rubric, both groups of students mentioned tangible objects such as coils, wires, and particles. Experts focused on physics objects and quantities, with the doctoral candidates mentioning many general quantities such as energy and work, and the lecturers mentioning domain specific physical interactions such as induced current and Lorentz force. There are no convincing differences between more and less proficient novices in this rubric. In the spatial rubric, the experts (especially the lecturers) showed a clear focus on geometry, with words such as *spherical symmetry*, *cylinder*, and *plane*; novices focused on spatial quantities with words such as *distance*, *velocity*, *area*, and *radius*; especially less proficient novices focused on topology words such as *through*, *enclosed*, and *between*. In the change rubric, experts (especially the lecturers) focused on things that do not change, with words such as *zero*, *homogeneous*, and *constant*; novices held a more dynamic picture with words such as *to cause*, *to change*, and *to run*.

Sentences

At the sentence level, we expected to find aspects of structure: namely, which parts of the problem are central in the participant's knowledge (e.g., the unknown, givens, or method). We had asked participants to indicate differences between the situations they had described. The proportion of differences mentioned could be taken as an indicator for locality versus diffusedness of the concepts used. If a participant constructs two different problem situations using concepts with only a diffuse meaning, the person might find it hard to explain differences between situations, so there will be little information regarding differences in the protocol. We also expected to see the difference between a single inference source and multiple redundant inference sources reflected at the sentence level. A sentence in which the participant elaborates on information mentioned previously can be interpreted as reflecting the latter type of redundancy.

To assess these aspects, we had to classify sentences according to their function in the problem description. With this goal in mind, we developed a coding schema. After several iterative steps we arrived at the coding schema presented in Table 4.

We could distinguish nine different functions of sentences, and we added a miscellaneous category for incomplete fragments and fragments that had no clear function. Because not all of the labels are mutually exclusive, we had to assign priorities to resolve ambiguities. The categories as they appear in Table 4 are listed in order of priority, so that a sentence would be labeled with the topmost label applicable. The first two categories are just to exclude sentences from further analysis: the first, because these are statements by the experimenter; and the second, because they are not on the problem situation. The rest are ordered from more to less specific. Thus, *given* is less specific than *goal* or *elaboration*. We also assigned *difference*, i.e., the comparison of situations, a

Table 4
Classification schema for sentences

rem	Remark by the experimenter
expl	Explanation of the formula/keyword
epis	Episode concerning the context/history of the problem
eval	Evaluative remark regarding the problem or the stimulus
diff	Difference/similarity between two situations mentioned
meth	Solution method
goal	Goal (what is to be solved)
elab	Elaboration on information already given
given	Information regarding the problem situation
misc	Miscellaneous

higher priority than "goal" or "method." We did so because we take the mention of differences as an indicator of the locality of concepts. As a consequence, this analysis could not reveal the precise content of the differences. The content of the differences is further addressed in the section on Global Properties.

A small fragment by a proficient student shows how the coding was applied (Proficient Student ID 14 in reaction to the stimulus shown in Figure 1. Owing to the translation, the original punctuation has changed).

> *given*: the type of problem can be a charged sphere with charge at the edge only, a surface charge
> *elab*: that is because the charges exert force on each other, as a consequence they move as far away from each other as they can; this results in a surface charge
> *given*: a capacitor that is charged
> *given*: and grounded at one side
> *elab*: then we have a surface charge also, since because of the forces all charge is driven to one surface
> *misc*: difference is that the nature of the forces
> *diff*: well, yes, the similarity in fact, is that there are forces that make the charge drift to the surface
> *diff*: difference is that one has a radial force
> *diff*: whereas with the capacitor all forces have the same orientation.

The coding system was tested by two independent raters. Both raters were physicists, and one of them also had ample teaching experience with the electricity and magnetism course. The interrater reliability was satisfactory, $\kappa = .79$, $n = 196$. The results of the analysis at the sentence level are summarized in Table 5.

Unsurprisingly, there are large differences in the overall frequencies across sentence types. These differences can be understood from the different roles these sentences have in the problem description, and possibly from the priorities we had set in the coding schema. However, we are interested in not so much these overall frequencies, but the differences in sentence usage across competence groups. This is more complex to analyze because there are several groups of participants with multiple participants in each group, and multiple observations of a nominal variable

Table 5
Frequencies and relative proportions of sentence types by level of expertise

	Expert (Lecturer) ($n = 3$)		Expert (Ph.D. Candidate) ($n = 3$)		More Proficient Student ($n = 6$)		Less Proficient Student ($n = 6$)	
	M (*SD*)	%	*M* (*SD*)	%	*M* (*SD*)	%	*M* (*SD*)	%
Episode	3.7 (2.5)	1.1	3.7 (3.5)	0.7	0.0 (0.0)	0.0	1.0 (2.0)	0.3
Evaluation	12.3 (7.6)	3.6	13.0 (9.0)	2.5	3.2 (2.6)	1.3	11.8 (7.5)	4.1
Difference	29.3 (24.1)	8.6	76.7 (27.0)	14.7	72.2 (47.0)	29.7	44.7 (14.6)	15.3
Method	19.7 (7.2)	5.8	29.3 (29.7)	5.6	12.5 (16.1)	5.1	7.7 (6.1)	2.6
Goal	18.3 (12.6)	5.4	20.7 (11.6)	4.0	31.3 (12.2)	12.9	35.8 (22.0)	12.3
Elaboration	36.0 (15.4)	10.6	34.0 (14.7)	6.5	22.3 (11.5)	9.2	19.7 (10.6)	6.7
Given	85.7 (24.9)	25.2	129.3 (11.6)	24.7	74.3 (16.9)	30.6	98.2 (16.1)	33.6
Miscellaneous	135.3 (84.9)	39.8	216.0 (194.7)	41.3	27.0 (12.1)	11.1	73.0 (14.4)	25.0
Total sentences	340.3 (124.0)	100	522.7 (204.8)	100	242.8 (81.5)	100	291.8 (22.2)	100

for each participant. Moreover, the number of participants per competence group is small. The best approach to find out whether there are significant differences in such a situation is to use an exact test. For the given data structure, a permutation test is a suitable type of test (Manly, 1983; Good, 1994). A permutation test can be used to assess the overall significance of between-group differences in the patterns of observed frequencies. We followed an approach originally described by Manly (1983), with a slightly modified test statistic (that is, we used the maximum likelihood G^2 instead of Pearson's chi-square). All p values are based on outcomes after 5000 random permutations. A detailed account of the procedure can be found in Savelsbergh (2001). Using this permutation test, we could demonstrate a significant difference between competence groups, $p = .01$.

Although the permutation test could be used to demonstrate that there are significant differences among groups, it cannot show how groups differ. One cannot directly read these differences from Table 5, because the table gives accumulated frequencies of sentences, implying that participants who spoke more gain more weight in the table. Moreover, the table provides little insight about which differences are statistically most relevant. Therefore, we used a correspondence analysis to explore the relations between competence and sentence patterns.

Here, we provide a brief introduction to correspondence analysis; a more extensive treatment is given by Greenacre (1993). As a starting point for this analysis, we had a table of observed frequencies consisting of 18 rows, 1 for each participant, and 8 columns, 1 for each sentence type. For the current purpose, a row can be seen as a frequency profile for a participant; likewise, a column can be seen as a frequency profile for a type of sentence. The difference between a pair of frequency profiles (either participants or sentence types) can be quantified in terms of distances using the chi-square statistic. These distances are computed both between each pair of individual profiles and between each individual profile and the average profile. The distances are then used to draw a diagram where points are plotted for each participant and for each sentence type. In the diagram, a participant who has a more deviant (from the average) frequency profile will be plotted farther from the origin, and participants who have very different frequency profiles will be plotted far apart. The same applies for sentence types. Moreover, if a participant uses a particular sentence type relatively more often than the others, the participant will be plotted more closely to the sentence type.

Although this analysis helps us recognize the statistically most important patterns in the data, some information is lost. The accuracy of the representation is measured by a quantity called the percentage of inertia; a diagram that retains all the information would represent 100% of the inertia. One can choose a 1, 2, 3, or higher dimensional diagram to present the analysis. The higher the dimensionality, the more information can be retained, but a diagram in more than three dimensions is hard to visualize, so that most authors choose either two or three dimensions. We settled on a correspondence analysis in two dimensions; thus, the total inertia explained was 73% and the inclusion of a third dimension only added another 14%, which is about chance value. The resulting correspondence plot is presented in Figure 2.

Figure 2 gives data points both for participants and for sentence types. In addition to points for individual participants, means per competence group are also indicated to clarify how participants belong together. The plot should be read in the following way: If, for instance, a participant is located in the upper right of the plot, her frequency profile will be more similar to the profiles of other participants located in same the same corner than to those of the participants in the lower left corner. Moreover, she most likely has many sentences on differences (located in the same corner) combined with few evaluative statements (opposite corner).

From the diagram it appears that the lecturers and the doctoral candidates clearly overlap, and because their numbers are too small to permit sensible analysis, both kinds of experts are taken

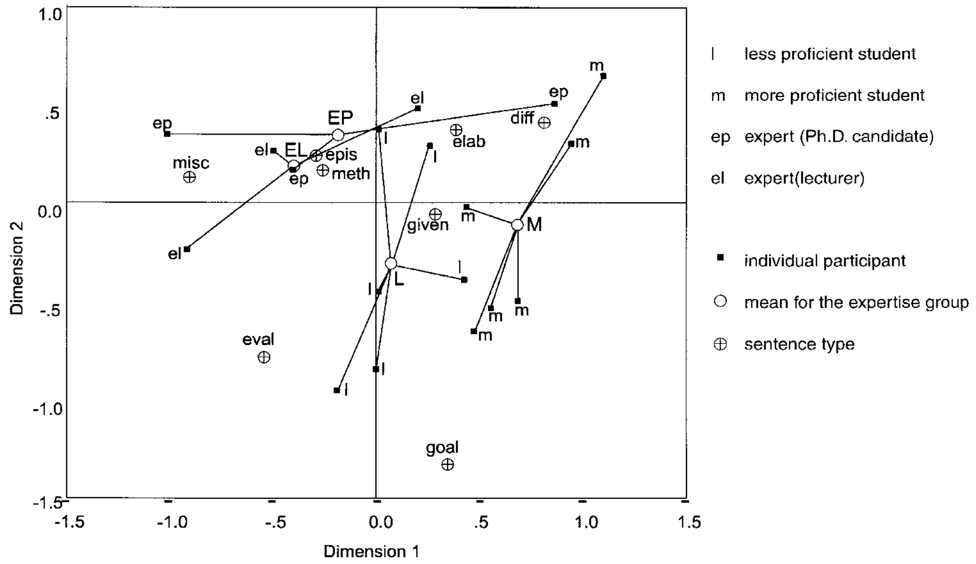

Figure 2. The relation between expertise level and the use of sentence types (correspondence analysis with canonical normalization).

together for further comparison. The remaining three groups have clearly different characteristic profiles. The graph shows that the frequent mention of elaborations combines with an emphasis on differences, and with few evaluative remarks. These properties span a dimension sloping at about 40°. More and less proficient students differ mainly along this dimension. Upon closer examination, it appeared that most evaluative remarks by novices were to express an unspecific lack of understanding. Therefore, we interpret this dimension to be associated with single inference source versus multiple redundant inference sources, and with diffusedness versus locality of properties. The graph also shows that goals are mainly mentioned by novices, experts mention few goals, and it seems that they mention more methods and episodes, although points so close to the origin should be interpreted with care. These properties span a dimension that is approximately orthogonal to the first. If we focus on the contrast of methods versus goals, this second dimension appears related to separate versus integrated solution information and to question versus givens. In addition to the differences between groups, the considerable variation within groups clearly indicates that the relation between sentence usage and competence is not 1 to 1.

Global Properties

At the level of the entire situation, we were primarily interested in qualities that could not be assessed at the lower levels. With regard to the content, we wanted to know the levels of abstraction in a representation. With regard to the structure of the problem, we wanted to assess the coherence of problem descriptions. Furthermore, we expected that the availability of a coherent and flexible problem representation that is linked to solution information would enable a person to formulate problems that differ in a relevant way and to explain these differences. Therefore, the relevance of alternatives and of differences mentioned had to be judged as well. We chose a set of four problem descriptions that were analyzed for all participants. All judgments were first made by two of the authors independently, after which differences were resolved by discussion. The authors were blind to the competence level of the participants for this rating procedure.

The first feature is the level of generality associated with the problem description. We found that there is no unambiguous standard for deciding whether a single proposition is instantiated or generic. By contrast, it is relatively clear whether a proposition is an instantiation of a foregoing one. A problem description may either start from a generic level and then become more specific, or it may be a mixture of generic and specific elements, or the entire description might be at a single level. As an example, consider the following fragment by a less proficient student in response to Gauss' law with keywords *surface charge*:

> two capacitor plates at proximate positions, you have to compute electrical intensity, something like that, I can't think of a real difference since these problems always concern capacitors, yeah, either flat plates or cylinders put together. (ID 11)

All concepts mentioned in the above fragment are at the same level of abstraction. Compare this with the following fragment by a lecturer:

> when you think of surface charge, in this case that implies conductors, *so in the inside the electrical intensity amounts to zero*, *as a consequence there is no spatial charge* . . . yes, I think of a problem that involves a cylinder, cylinders put together, and spheres put together, *leading to charges on the inside and on the outside*, *in which case components cancel out due to symmetries*, which makes the integral easier to evaluate. (ID 7)

This lecturer has a redundant description with more or less concrete elements (conductors, spheres, and cylinders), and with elaborations referring to generic properties of these elements (the elaborations are italicized). For each description, we determined whether there were multiple levels of generality. The quantitative outcome of the analysis is presented in Figure 3. To test the competence dependence of this feature, we used Spearman rank correlation because we expected a monotonous relation between competence and the use of multiple levels. It turns out that the use of

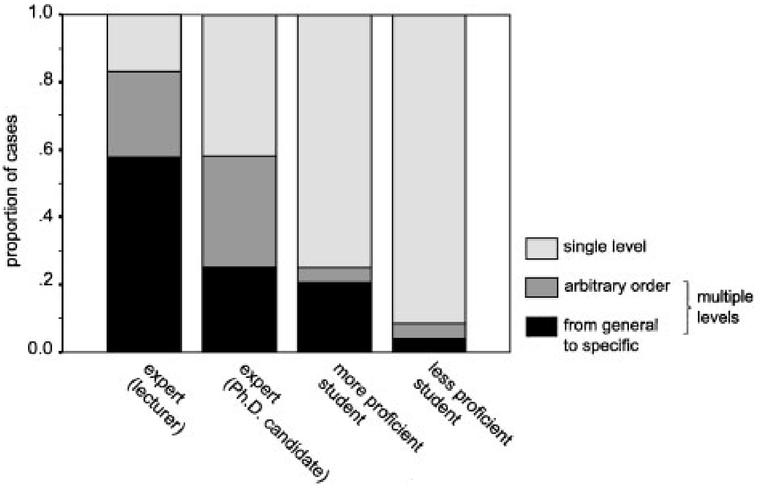

Figure 3. The presence of multiple levels in a problem description.

multiple levels clearly increases with competence, $r_S = .86$, $df = 18$, $p < .001$. Less proficient students give hardly any multilevel descriptions. They either give an abstract description or they give a concrete description but they do not connect them.

Coherence is another aspect that has to be judged for the description as a whole. The descriptions consist of a series of propositions that may be coherently linked or characterized by strange leaps. In the extreme case, it is hardly possible to give meaning to the description. (In)coherence was judged on a 5-point scale ranging from "no incoherences" to "a single incoherent step," "some incoherent steps, still easily understandable," and "several incoherent steps, hard to understand," to "completely incoherent." Examine, for instance, the following descriptions given by a less proficient student, in reaction to Gauss' law with keywords *spherical symmetry*:

> A sphere is not a field, but, I doubt whether that is relevant.... There may be a field inside the sphere, but, well if there's no charge at least. (Experimenter: do you see a problem that employs both?) in fact I do but, yes, a sphere okay, but spherical symmetry... you may, a sphere, or, uh, two spheres put together, where, if on one sphere you put this amount of charge, and on the other you put that amount, you can compute using the enclosed charge, you may compute the field, or when there is spatial charge inside the sphere, you may compute that as well. (Less Proficient Student, ID 18)

This fragment was considered to be "with several incoherent steps." It is typical of the fragmentary nature of weak novice problem descriptions. The fragment is mainly an enumeration of objects and some isolated propositions about these objects. In the first few lines the student is apparently trying to get some handle on the terms used. Then, after an intervention by the experimenter, a situation is described. The wording remains fuzzy and the relations between objects remain implicit or ambiguous. The spheres are "put together" (probably concentric) and you have to use the "enclosed charge" (probably within a surface just outside the outer sphere), and finally a field is to be computed without specification of the place where it could be computed. The final sentence states that you might compute "that," without specifying what "that" is. Nevertheless, there is a sensible interpretation to most of the description. In the next fragment, by a proficient student, there are far fewer ambiguities and the relation between a statement and the previous statement is generally clear.

> A charged sphere, for instance, the electric field clearly is a spherical symmetry, and with a Gauss box, a sphere around it for example, so you can compute the field at a given distance from the sphere, and the other way round, with a given sphere or Gauss box, and you measure a certain field intensity at that distance, you can use the formula to compute the charge that is in the Gauss box, in the Gauss sphere. (Proficient Student, ID 16)

Scores are presented in Figure 4. The image is clear: These lecturers' problem descriptions were more coherent than those of the others. The proficient students told more coherent stories than did the less proficient students. Furthermore, the difference between doctoral candidates and proficient undergraduate students was difficult to discern apart from the fact that the doctoral candidates did not tell completely incoherent stories. Statistical analysis indicates that problem description coherence increases significantly with competence, $r_S = .64$, $df = 18$, $p = .004$. We have also searched for contradictions in problem descriptions but we could find only a few, perhaps because most inconsistent descriptions were incomprehensible in the first place.

We asked participants to describe two different situations in response to a single stimulus. How well they succeeded in doing this was the next measure we examined. Participants could

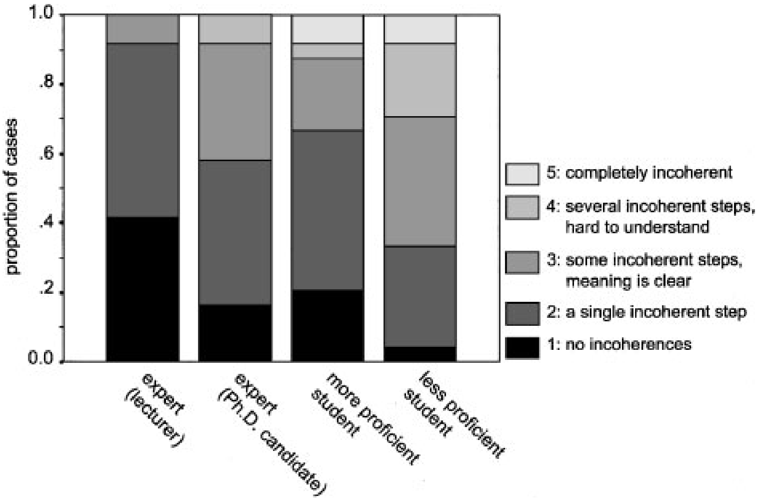

Figure 4. The incoherence of problem descriptions by level of expertise.

either describe two different situations in a way relevant or irrelevant from a physics point of view, or no alternative at all. A difference between two situations would be relevant if it involves some qualitatively different physics, or if the solution approach would be affected by the difference. A typical excerpt illustrates that alternative problem descriptions generated by the less proficient novices did not satisfy this criterion:

> two charges placed at an axis at equal distance, that cancel out completely or partially, or several charges, and then compute the force of the third charge, yes the forces that is exerted on the third charge, second situation, that you have several charges of the same, or of different, you have different kinds of charges . . . that amplify each other proportionally, or amplify. (Experimenter: what exactly do you mean by that?) both cause the same force in the same direction, difference is that in the first situation you had different, or two equal charged that cancelled out each other, in the other you had the same, or yes different charges that amplified each other. (Less Proficient Student, ID 12)

In this example, the situations described are identical for both problems. That is, only the relative magnitude of the charges involved changes, which in this case does not lead to differences in the solution approach. The student did not indicate that he was dissatisfied with the alternative he described. This would be different in protocols by more proficient participants, who in general produced more relevant alternatives. If the more proficient participants did generate an irrelevant alternative, at least they would be dissatisfied with it. We take this as evidence that in the minds of more proficient participants, solution methods are associated with problem representations. In Figure 5, the quantitative findings are presented on the presence of alternatives in problem descriptions. The increase of relevant alternatives with competence is significant, $r_S = .62$, $df = 18, p = .006$.

Apart from giving alternatives, participants were asked to state explicitly what alternatives they had introduced in their description. They could mention differences between two situations or differences between the ways the two problems were solved, which is illustrated in these fragments by one of the proficient students:

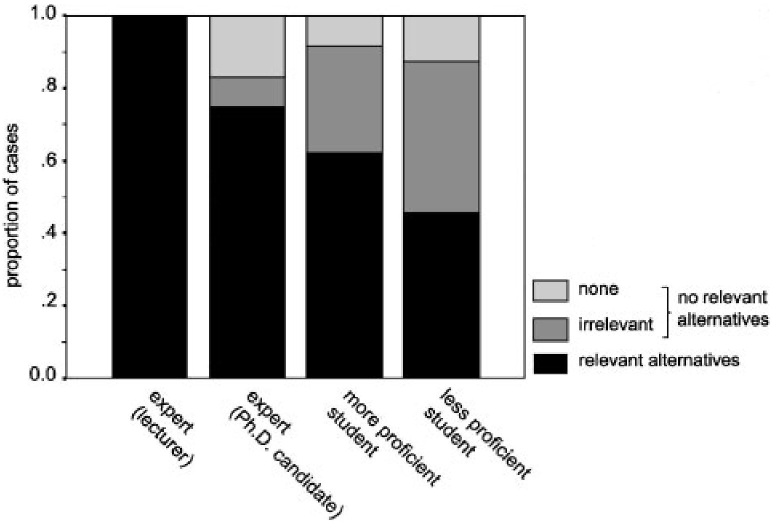

Figure 5. The presence of alternatives in the problem description.

> well, yes, the similarity in fact is that there are forces that make the charge drift to the surface, difference is that one has a radial force, whereas with the capacitor all forces have the same orientation . . . difference is that with a closed circuit, all pieces are really summed up, whereas with an infinitely long wire, we integrate to infinity, they are similar in that the same formula is valid. (Proficient Student ID 14)

The quantitative results are summarized in Table 6. Apparently, the experts (lecturers and doctoral students) mentioned more differences between the situations than the novices, but the relation with competence is not significant, $r_S = .25$, $df = 18$, $p = .32$. Differences in solution methods were not mentioned often by either experts or students, and there do not appear to be significant differences ($r_S = .26$, $df = 18$, $p = .30$).

Discussion and Conclusions

We set out to specify the competence-dependent characteristics of electrodynamics situational knowledge to find out what knowledge novices would need to construct more helpful problem representations. At three different levels of analysis (words, sentences, and complete

Table 6
Proportion of cases in which differences were mentioned explicitly by level of expertise

	Expert (Lecturer) (*n*=3)	Expert (Ph.D. Candidate) (*n*=3)	More Proficient Student (*n*=6)	Less Proficient Student (*n*=6)
	M (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
Differences between physical situations	0.50 (0.43)	0.50 (0.25)	0.42 (0.30)	0.38 (0.49)
Differences in solution method	0.25 (0.00)	0.17 (0.14)	0.25 (0.22)	0.13 (0.14)

problems), we looked for characteristics that differ among problem solvers from different levels of competence and experience (experts, and more and less proficient novices). Through the use of multiple competence levels, we could assess whether differences within the novice group are along the same dimensions as are differences between experts and novices. An advantage of having different levels of analysis is that we could triangulate our analyses, and interpret findings at one level by the use of findings at another level.

Our data show that it is characteristic of experts' situational knowledge to focus on physics concepts and quantities, on constraints, and on geometrical relations. To the expert, which variables are given and which are unknown is of secondary importance. Experts know coherent situation descriptions integrated with solution information, and they avail of multiple redundant representations for a situation.

Characteristic of novices' situational knowledge is a focus on tangible objects, topological relations between these objects, and causal relations and behavior. The unknown has a central role in their knowledge. Novices do know situation fragments but these fragments are not tightly coupled to solution methods, and they do not know alternative representations for a single situation. There was no evidence of a difference in the use of qualitative versus quantitative knowledge except that novices refer more to spatial quantities. These outcomes, except for the last, are much in line with the differences suggested by research in other domains as presented in Table 1.

The differences between more and less proficient novices are not always along the same lines as the differences between experts and novices. Although the global judgments suggest that the use of multiple representations, coherence of problem descriptions, and mention of relevant alternatives have a monotonous relation to competence, analyses of word and sentence usage yield a more differentiated image. More and less proficient students both refer to tangible objects, to causal relations and behavior, and to spatial quantities. More and less proficient novices are also similar in that they have a strong focus on the unknown. However, word usage suggests that less proficient students are more concerned with the topology of the problem. Moreover, sentence usage suggests that whereas less proficient beginners stay with an incoherent single-level problem description, proficient beginners can elaborate on the initial elements to create a richer and more coherent representation of the problem. This suggests a more networked structure in the knowledge of proficient novices. In earlier accounts of the quality of problem representations, there has been a focus on the link from surface-level features to solution approaches or physics principles (Newell & Simon, 1972; Chi et al., 1981). Our results draw attention to the many links within a representation, which may be equally important to build a flexible situational knowledge, and which deserve more attention. This finding differs from Chi et al. (1981), who concluded that novices have sufficiently elaborate knowledge about situations but that they lack solution schemata.

We cannot say much about the diffusedness of concepts except that close reading of the protocols suggests that less proficient novices can give only a diffuse meaning to the words they use. For instance, they mention words that express a relation without specifying between which objects the relation exists (e.g., "there is a force"). This might be because the less proficient beginner has a diffused concept of force, or because of the different ontological status students assign to their concepts, as students tend to treat most physical science concepts as a kind of substance (Chi, Slotta, & de Leeuw, 1994). With the current approach we can see in what context and structure a word is used, but we cannot go much beyond that.

The number of experts in our study was too small to draw strong conclusions about differences among the experts. In some respects the doctoral candidates in our sample were between the lecturers and proficient students. However, word usage suggests that whereas both kinds of experts refer to physics concepts, the lecturers refer more to domain-specific concepts, whereas the doctoral candidates refer more to general physics concepts. This supports the idea that there are qualitatively different types of expertise, and therefore that experts should not be regarded as a homogeneous group.

Educational Implications

The finding that novices have trouble constructing a basic representation is not new, but in this study part of the problem could be traced to the way novices remember earlier problem-solving situations. This study also presents evidence that at least some of the differences in situational knowledge between more and less proficient novices are along different dimensions than differences between experts and novices. These results are important in teaching. For a less proficient novice, the first step is to become a more proficient novice rather than a lecturer. Therefore, problem features that are most prominent in experts reasoning are not necessarily the first characteristics to which novices should pay attention. The less proficient students are clearly struggling with the basic layout of the situation. More proficient students are still much concerned about what happens in the situation. To the experts these steps go without saying for standard situations. To the expert it seems plausible to discuss how problem features are related to solution approaches, and an analysis of the situation serves to find hints for the solution approach. To a novice who does not understand the situation, this cannot be meaningful.

Among the differences we found between more and less proficient novices, the availability of multiple inference sources is interesting. We hold that more attention should be paid to the concrete situation in its own right, as the expert would do in an unfamiliar situation (Clement, 1994). One option to achieve this may be the line followed in the modeling approach proposed by Hestenes (1987), in which the development of a model begins with discussion of the concrete objects, followed by the introduction of variables, and relations. Another way to promote elaboration on the available information might be to use goal-free practice problems like the ones used by Sweller (1988) to reduce cognitive load.

The finding that beginners have so much incoherence in their situational knowledge suggests that the fundamental entity in a beginner's problem representation will not be the entire situation, but rather a smaller unit. At the same time it appears that beginners mention solution goals far more often than do experts. These two tendencies may be related to novices' preference for working backward. As long they cannot construct coherent problem representations, they have good reason to skip thorough analysis of the problem statement when they are trying to solve the problem (De Jong & Ferguson-Hessler, 1984). Their best bet would be to follow a kick-and-rush approach, starting from something they need to know, the entity that is being asked for. It seems plausible that a discussion of how situation features influenced the solution of the problem could be more effective after solving the problem (cf. Taconis et al., 1999).

Finally, both students and experts in our study made surprisingly little use of drawings in describing the problem situations they created. Of course, the incentive to make drawings may depend on characteristics of the experimental task. However, in many circumstances, and certainly in this domain, a diagram is an excellent means to express enhanced problem representations (Larkin & Simon, 1987). It seems worth the effort if we could help students experience drawing as a worthwhile activity. Which of these suggestions provides the best way to proceed and how this should be implemented remain issues for further research.

The authors thank Jules Pieters and Harold Bult for discussions on setting up the experiment, and Ivo van der Lans for assistance in statistical analysis of the data.

Notes

[1] In chess, in which the recall method was first applied, this might pose less of a problem because there are fewer alternative descriptions. The game of chess has a tangible level of pawns, bishops, and castles, and a level of named patterns such as gambit or a Nimzo Indian. A normal playing situation always involves a board and pieces, so it is unproblematic to use this level of representation as a stimulus. In contrast, for a physics problem there is more freedom of description and the descriptions are not privileged. Therefore, in physics the particular choice of the stimulus could influence the results considerably. To find out about the situational knowledge a participant has in mind for a certain type of problem, it seems important to pick a stimulus as open as possible.

References

Anderson, J.R. (1983). The architecture of cognition. Cambridge, MA: Harvard University Press.

Biederman, I. & Shiffrar, M.S. (1987). Chicken sexing: An expert systems and experimental analysis of a difficult perceptual learning task. Journal of Experimental Psychology: Human Learning, Memory and Cognition, 16, 640–645.

Boshuizen, H.P.A. & Schmidt, H.G. (1992). On the role of biomedical knowledge in clinical reasoning by experts, intermediates and novices. Cognitive Science, 16, 153–184.

Braune, R. & Foshay, W.R. (1983). Towards a practical model of cognitive information processing: Task analysis and schema acquisition for complex problem-solving situations. Instructional Science, 12, 121–145.

Chi, M.T.H. & Bassok, M. (1989). Learning from examples via self-explanations. In Resnick, L.B. (Ed.), Knowing, learning, and instruction: Essays in honor of Robert Glaser (pp. 251–282). Hillsdale, NJ: Erlbaum.

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

Chi, M.T.H., Glaser, R., & Rees, E. (1982). Expertise in problem solving. In Sternberg, R.J. (Ed.), Advances in the psychology of human intelligence (vol. 1) (pp. 7–77). Hillsdale, NJ: Erlbaum.

Chi, M.T.H., Slotta, J.D., & de Leeuw, N. (1994). From things to processes: A theory of conceptual change for learning science concepts. Learning and Instruction, 4, 27–43.

Clement, J. (1994). Use of physical intuition and imagistic simulation in expert problem solving. In Tirosh, D. (Ed.), Implicit and explicit knowledge. Hillsdale, NJ: Ablex.

De Jong, T. & Ferguson-Hessler, M.G.M. (1984). Strategiegebruik bij het oplossen van vakinhoudelijke problemen in een semantisch rijk domein: Electriciteit en Magnetisme [Strategy use while solving problems in a semantically rich domain: Electricity and Magnetism]. Tijdschrift voor Onderwijsresearch, 9, 3–16.

De Jong, T. & Ferguson-Hessler, M.G.M. (1991). Knowledge of problem situations in physics: A comparison of good and poor novice problem solvers. Learning and Instruction, 1, 289–302.

De Jong, T. & Ferguson-Hessler, M.G.M. (1996). Types and qualities of knowledge. Educational Psychologist, 31, 105–113.

de Kleer, J. & Brown, J.S. (1981). Mental models of physical mechanisms and their acquisition. In J.R. Anderson (Ed.), Cognitive skills and their acquisition (pp. 285–309). Hillsdale, NJ: Erlbaum.

de Kleer, J. & Brown, J.S. (1983). Assumptions and ambiguities in mechanics. In D. Gentner & A.L. Stevens (Eds.), Mental models (pp. 155–190). Hillsdale, NJ: Erlbaum.

diSessa, A.A. (1993). Toward an epistemology of physics. Cognition and Instruction, 10, 105–225.

Ericsson, K.A. & Simon, H.A. (1993). Protocol analysis: Verbal reports as data (rev. ed.). Cambridge, MA: MIT Press.

Fuson, K.C. & Carroll, W.M. (1996). Levels in conceptualising and solving addition and subtraction compare word problems. Cognition and Instruction, 14, 345–371.

Glaser, R., Schauble, L., Raghavan, K., & Zeitz, C. (1992). Scientific reasoning across different domains. In de Corte, E., Linn, M., Mandl, H., & Verschaffel, L. (Eds.), Computer-based learning environments and problem solving (pp. 345–373). Berlin, Germany: Springer-Verlag.

Goei, S.L. (1994). Mental models and problem solving in the domain of computer numerically controlled programming. Unpublished doctoral dissertation, Twente University, Enschede, The Netherlands.

Good, P. (1994). Permutation tests: a practical guide to resampling methods for testing hypotheses. New York: Springer.

Greenacre, M.J. (1993). Correspondence analysis in practice. London: Academic Press.

Hammer, D. (1994). Epistemological beliefs in introductory physics. Cognition and Instruction, 12, 151–183.

Hestenes, D. (1987). Toward a modeling theory of physics instruction. American Journal of Physics, 55, 441–454.

Hinsley, D.A., Hayes, J.R., & Simon, H.A. (1977). From words to equations: Meaning and representation in algebra word problems. In Carpenter, P.A. & Just, M.A. (Eds.), Cognitive processes in comprehension (pp. 89–106). Hillsdale, NJ: Erlbaum.

Huffman, D. (1997). Effect of explicit problem solving instruction on high school students' problem-solving performance and conceptual understanding of physics. Journal of Research in Science Teaching, 34, 551–570.

Jacobson, M.J. & Spiro, R.J. (1995). Hypertext learning environments, cognitive flexibility, and the transfer of complex knowledge: An empirical investigation. Journal of Educational Computing Research, 12, 301–333.

Larkin, J.H. (1983). The role of problem representations in physics. In Gentner, D. & Stevens, A. L. (Eds.), Mental models (pp. 75–98). Hillsdale, NJ: Erlbaum.

Larkin, J.H., McDermott, J., Simon, D.P., & Simon, H.A. (1980). Models of competence in solving physics problems. Cognitive Science, 4, 317–345.

Larkin, J.H. & Simon, H.A. (1987). Why a diagram is (sometimes) worth ten thousand words. Cognitive Science, 11, 65–100.

Lawson, A.E., Clark, B., Cramer-Meldrum, E., Falconer, K.A., Sequist, J.M., & Kwon, Y.J. (2000). Development of scientific reasoning in college biology: Do two levels of general hypothesis-testing skills exist? Journal of Research in Science Teaching, 37, 81–102.

Manly, B.F.J. (1983). Analysis of polymorphic variation in different types of habitat. Biometrics, 39, 13–27.

McCloskey, M. (1983). Naive theories of motion. In Gentner, D. & Stevens, A.L. (Eds.), Mental models (pp. 299–324). Hillsdale NJ: Erlbaum.

McDermott, J. & Larkin, J.H. (1978). Re-representing textbook physics problems: Proceedings of the 2nd national conference of the Canadian Society for Computational Studies of Intelligence. Toronto: University of Toronto Press.

McMillan, C. & Swadener, M. (1991). Novice use of qualitative versus quantitative problem solving in electrostatics. Journal of Research in Science Teaching, 28, 661–670.

Mettes, C.T., Pilot, A., & Roossink, H.J. (1981). Linking factual and procedural knowledge in solving science problems: A case study in a thermodynamics course. Instructional Science, 10, 333–361.

Newell, A. & Simon, H.A. (1972). Problem solving. Englewood Cliffs, NJ: Prentice-Hall.

Piaget, J. (1970). Piaget's theory. In Mussen, P.H. (Ed.), Carmichael's manual of child psychology, vol. 1 (3rd ed.) (pp. 703–732). New York: Wiley.

Polya, G. (1945). How to solve it. Princeton, NJ: Princeton University Press.

Reif, F. (1983). Understanding and teaching problem solving in physics. Research on Physics Education: Proceedings of the First International Workshop Lalonde les Maures (pp. 13–53). Paris: Centre National de la Recherche Scientifique.

Reif, F. & Heller, J.I. (1982). Knowledge structure and problem solving in physics. Educational Psychologist, 17, 102–107.

Royer, J.M., Cisero, C.A., & Carlo, M.S. (1993). Techniques and procedures for assessing cognitive skills. Review of Educational Research, 63, 201–243.

Savelsbergh, E.R. (2001). Testing for expertise-dependent differences in think-aloud protocols using a permutation test (tech report). Utrecht, The Netherlands: University of Utrecht.

Slotta, J.D., Chi, M.T.H., & Joram, E. (1995). Assessing student's misclassifications of physics concepts: An ontological basis for conceptual change. Cognition and Instruction, 13, 373–400.

Smith, J.P., diSessa, A.A., & Roschelle, J. (1993). Misconceptions reconceived: A constructivist analysis of knowledge in transition. Journal of the Learning Sciences, 3, 115–163.

Stenning, K. (1992). Distinguishing conceptual and empirical issues about mental models. In Rogers, A.R.Y. & Bibby, P.A. (Eds.), Models in the mind (pp. 29–48). London: Academic Press.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. Cognitive Science, 12, 257–285.

Sweller, J., Mawer, R.F., & Ward, M.R. (1983). Development of expertise and mathematical problem solving. Journal of Experimental Psychology, 112, 639–661.

Taconis, R. & Hout-Wolters van, B. (1999). Systematic comparison of solved problems as a cooperative learning task. Research in Science Education 29, 313–329.

Thompson, L., Gentner, D., & Loewenstein, J. (2000). Avoiding missed opportunities in managerial life: Analogical training more powerful than individual case training. Organization Behavior and Human Decision Processes, 82, 60–75.

Van Hiele, P.M. (1986). Structure and insight: A theory of mathematics education. Orlando: Academic Press.

Wells, M., Hestenes, D., & Swackhammer, G. (1995). A modeling method for high school physics instruction. American Journal of Physics, 63, 606–619.

Willson, V.L. (1990). Methodological limitations for the use of expert systems techniques in science education research. Journal of Research in Science Teaching, 27, 69–77.

Zajchowski, R. & Martin, J. (1993). Differences in the problem solving of stronger and weaker novices in physics: Knowledge, strategies, or knowledge structure. Journal of Research in Science Teaching, 30, 459–470.