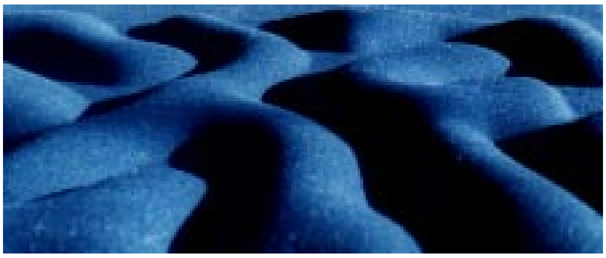
IMPROVING MENTAL REPRESENTATIONS IN PHYSICS PROBLEM-SOLVING



For further information on my work, and for the latest version of this document, visit:
http://www.mpib-berlin.mpg.de/EuB/Staff/savelsbergh/

This document also appeared in a paper version: ISBN 90-36511623
Thesis, Twente University, Enschede, The Netherlands

Cover: detail of 3-dimensional upholstery material 'feel my brain' by Marco Savelsbergh
Print: Febodruk, Enschede

IMPROVING MENTAL REPRESENTATIONS IN PHYSICS PROBLEM-SOLVING

PROEFSCHRIFT

ter verkrijging van
de graad van doctor aan de Universiteit Twente,
op gezag van de rector magnificus,
prof. dr. F.A. van Vught,
volgens besluit van het College voor Promoties
in het openbaar te verdedigen
op vrijdag 4 september 1998 te 15.00 uur.

door

Elwin Rudolf Savelsbergh
geboren op 17 mei 1968
te Amsterdam

Dit proefschrift is goedgekeurd door de promotor:

prof. dr. J.M. Pieters

Assistent-promotoren:

prof. dr. A.J.M. de Jong
dr. M.G.M. Ferguson-Hessler

Voorwoord (Preface)

Precies vijf jaar na het begin van mijn promotieonderzoek is mijn proefschrift daar. Toen ik vijf jaar geleden als net afgestudeerd natuurkundige bij de vakgroep PeTiT binnenkwam, verwachtte ik te zullen samenwerken binnen de leerpsychologische tak van de vakgroep. Al snel verdween de ene na de andere leerpsycholoog, en uiteindelijk bleef ik als enige over. Gelukkig waren er nog andere aio's, en viel er altijd wel ergens iets samen te werken, te overleggen of zomaar te bespreken. Vooral dankzij mijn beide kamergenoten en de koffiepauzegroep zijn vijf jaar omgevlogen. Anneloes, Asteria, Margie, Yvonne en anderen, bedankt.

Daarnaast waren er de Twentse collega's. Onze ProIST ontmoetingen waren ter lering ende vermaeck. Janine verdient bijzondere vermelding omdat ze mij vanaf het begin wegwijs heeft gemaakt bij de vakgroep IST en omdat we samen het karakter van onze begeleider konden analyseren.

De begeleiders verdienen natuurlijk ook bijzondere dank. Jules, als luisterende begeleider op afstand, Monica, omdat zij begrijpt hoe een fysicus denkt en tot slot Ton, omdat hij (bijna) altijd de andere kant vertegenwoordigde.

Daarnaast hebben nog anderen een bijdrage geleverd aan dit proefschrift. Ivo van der Lans heeft meegedacht over de statistische analyse van de experimenten in deel A van dit proefschrift, Cees Alderliesten, Rudy Borkus, Ger Engelbertink en Jan Kuperus hebben het experiment in deel B mede mogelijk gemaakt en Steven Ralston heeft zich gebogen over het Engels.

Tot slot bedank ik Annet die vaak niet over onderzoek wilde praten en soms ook wel, en die daar groot gelijk in heeft.

Elwin Savelsbergh

Berlijn, juli 1998

Contents

APPENDICES

1. General introduction

This research addresses the role mental representations play in physics problem solving. The mental representation of a problem has been recognised as an important determinant of problem-solving performance. A correct and adequate problem representation[1] is a basic requirement for fluent reasoning towards a solution.

To become proficient physics problem solvers, students must learn to construct proper mental problem representations. Clearly, some are more successful than others in this learning process. To support weaker students' learning, which is the ultimate goal of our research, we first need to understand what kinds of representational skills are needed and what skills students lack. This knowledge can then be used to design more effective physics teaching approaches.

In the first part of this dissertation, our focus is on describing problem representations of both physicists and physics students with different levels of competence, and on understanding the role problem representations play in problem-solving. In the second part we seek to improve problem representations physics students form in education.

In the current chapter we address the nature of physics problem-solving, compare physics problem-solving to reasoning in other domains, discuss the role of mental representations in physics problem-solving, and review what is known about the relation between problem representations and competence. The chapter ends with the main research questions and an outline of the rest of the book.

1.1 Physics problem-solving

> *The problem is this: are the physical laws symmetrical under reflection. Put it this way: suppose I build a piece of equipment, let's say a clock, with lots of wheels and things and hands and numbers, and everything, and it ticks and works and it got springs wound up inside. I look at the clock in the mirror; and it is not the thing that I want to know how it looks in the mirror, that's not the thing. The question is this: I build another clock, which is exactly the same as that clock looks in the mirror. Every time there is a screw with a right hand thread in here, I put a screw with a left hand thread in the corresponding place. When the thing has a 'two' marked on the face I mark 'two' backwards on the face over here. Each wheel, each coil, spring, how high a spring is twisted one way, twisted in the mirror image the other way. So, when I am all finished, I have two things, which have… both physical, which bear to each other the relation of an object and its mirror image, although they are both objects, that I wish to emphasise. Thus it is a physical question, not a theoretical one alone. Now, the physical question is: if the two things are started in the same condition,*

[1] *Problem representation* is short for *mental problem representation*. Occasionally, we use *mental representation* as an equivalent, to distinguish them from external representations.

> *the springs are wound with corresponding tightnesses and everything, whether the two clocks will tick and go around and look forever after as exact mirror images. I think your intuition about the laws of physics would suggest that they would* (from a lecture by Richard Feynman, 20-5-1963).[2]

Attending a lecture like Feynman's might motivate students to wonder about the mirror image of a magnetic field induced by a current loop. Likewise, exploring a physics domain with a computer simulation clearly motivates students and helps them to become acquainted with domain concepts. Mastering physics is an exacting undertaking, however, and it is only after this first encounter that students begin to learn how to solve problems in the domain. It is common that impasses and misunderstandings arise during the process, and often insight only comes after a period of 'wandering in the dark'. After the initial conceptual barriers have been overcome, it still requires considerable practice to become good at selecting the right solution step in particular circumstances, recovering from errors, and carrying out the selected solution steps. As fluent problem-solving is a core aspect of physics expertise, the development of the problem-solving ability is of major interest not only to cognitive psychologists studying physics learning, but also to physics instructors and physics students themselves.

In recent science-teaching research, the emphasis has shifted from formal problem-solving to conceptual understanding, and to understanding the nature of the scientific enterprise. This shift in attention has been motivated both by changing psychological insights and by changes in the kinds of knowledge valued in society. One of the psychological reasons is that students proficient in problem-solving may still entertain an erroneous qualitative understanding (diSessa, 1993; Hestenes, Wells, & Swackhammer, 1992; McCloskey, 1983; Minstell, 1989) and that they do not transfer school knowledge to out-of-school contexts (Lave, 1988). It has also been argued that meaningful learning requires conceptual understanding to precede formal reasoning (Van Hiele, 1957, 1986; but see Hestenes, 1986; White, 1993; see also Plötzner & Spada, 1993). Apart from the order of presentation, several educators and physicists have argued that the formal mathematical representation and mathematical deduction contribute essential parts to physics understanding (Feynman, 1965; Hestenes, 1987; a similar point for mathematics is made by Keune, 1998). We adhere to this viewpoint in that formal reasoning and conceptual understanding influence each other, and that effective instruction addresses both. Clearly, some physics phenomena are easily accessible to human sensation, and some of these can be described accurately using proportional relationships. For such phenomena qualitative conceptual understanding may suffice for most people. More subtle phenomena, such as quantum and relativity effects, but also advanced classical mechanics or electromagnetism, by contrast, cannot be understood at a purely qualitative conceptual level. Understanding relativity theory as a necessary supplement to

[2] Transcript of published recording (Feynman , 1997)

Newton's classical mechanics, for instance, requires an advanced sense of formalism.

At least as important to the changing emphases in science education is society's changing appreciation of knowledge. Policy makers and curriculum designers in many countries have come to view a qualitative understanding of physics – more broadly science –, and of the scientific enterprise, as requirements for good citizenship (compare, in The Netherlands: introduction of new examination programs in secondary education (Gravenberch, 1996); in the USA: project 2061 (Rutherford & Ahlgren, 1991)). A similar shift has occurred in other subjects in secondary education such as math education. Although such qualitative understanding of science is important, it requires additional abilities to become a scientist or an engineer. As it takes considerable practice to become fluent with formal representations, it may be advisable for secondary education to re-value its role as a preparation for university education. The physics community clearly has a stake in this issue, and it is remarkable how little physicists make themselves heard in the debate. Our research is primarily aimed towards university physics education, however, where the importance of a formal problem-solving ability remains largely unchallenged.

Before turning to requirements for becoming a successful physics problem solver, we have to define what kinds of problem-solving we are after. According to the Gestalt psychologist Duncker (1945, p. 2), 'a problem exists when a living organism has a goal but does not know how this goal has to be reached'. This definition of problem-solving clearly stretches out beyond the types of problems a physicist should be able to deal with as a professional. However, the definition also excludes some kinds of behaviours that often have been termed problem-solving too — that is, those questions where the answer is obvious to the subject, and those goals that can be achieved without conscious mental effort. Like Duncker, we exclude 'routine problems' from our definition.

The focus of our studies is on proficiency in physics problem-solving. As a domain we use electrodynamics topics that are a standard part of universities' first-year physics curricula[3]. The electrodynamics domain shares essential characteristics with other physics domains that are not commonly found in other domains. First, most problems in the domain are complex problems — that is, involving structured multi-element situations and multi-step solution procedures. Second, electrodynamics is a conceptual domain, which implies that many of the objects in the domain are not directly perceivable — they are more or less abstract concepts. Third, it is a formal domain, which means there is a limited set of well-defined concepts and laws that can be applied within a limited and well-defined range of problems. The final characteristic shared between electrodynamics and other

[3] Typical course titles are: 'Electricity & Magnetism', 'Electricity Theory' and 'Electrodynamics'. Topics include: charge distributions, symmetries, Coulomb's law, Gauss' law, field energy, dipoles and multipoles, conductors, image charges, Laplace's law, Ampere's law, Lorentz force, magnetic flux, and inductance.

physics domains is that problems can be represented in multiple ways. Though these different representations can be mapped onto each other without loss of information, one way may be more apt than another in a given situation. The existence of formally equivalent formulations distinguishes physics from a domain like medicine: Although a medical problem can be described either in terms of clinical knowledge or in terms of biomedical knowledge, there is no formal equivalence between the two descriptions.

Besides commonalities there are also distinctions between electrodynamics and other physics domains. Several of the more well known studies on problem-solving and reasoning in conceptual domains have been conducted in the field of mechanics (Chi, Feltovich, & Glaser, 1981; Chi & Bassok, 1989, Larkin, 1983; McCloskey, 1983; diSessa, 1993). Mechanics is more closely related to everyday experience than electrodynamics is. Therefore, one could expect people's naive ideas concerning mechanics and motion to be substantially more established than their ideas with regard to electrodynamics. A further distinguishing feature is that electrodynamics is much concerned with symmetry right from the start, whereas in initial mechanics courses, for instance, the use of symmetries is not profound, and sometimes even misleading (diSessa, 1993). On the other hand, in mechanics, conservation laws have a central role. Both conservation laws and symmetries can be used to simplify the problem, and in both cases this requires the problem-solver to search for *constraints*, which is quite different from the naive approach of running a *mental simulation* (compare *Dynamic behaviour* in Section 2.1).

To summarise, we assert that the psychology of electrodynamics problem-solving is similar to that of problem-solving in other physics domains. In other fields similar processes may be found, particularly in those formal conceptual domains where problems can have multiple representations.

1.2 Problem representations in physics

1.2.1 Problem representations

In their classic 'Human problem solving', Newell and Simon (1972) distinguish two main processes in problem-solving: understanding the problem and searching for the solution route. Like most studies from that period, they focus on problems that require little background knowledge. Such problems are generally easy to understand. Therefore, in these studies the focus has been on search processes, and the interpretation of the problem has been neglected as being a trivial process. Contrastingly, in knowledge-intensive domains, such as physics, understanding the problem is often not trivial. Especially with complex or ill-defined problems, a proper representation of the situation can be the key to selecting a solution method, and to interpreting the outcome.

Complex and ill-defined problems took a central place in the work by Duncker and other Gestalt psychologists, and we can neatly illustrate the importance of problem representations with the following example from Duncker's well-known experiments on the 'radiation problem':

> *Given a human being with an inoperable stomach tumour, and rays which destroy organic tissue at sufficient intensity, by what procedure can one free him of the tumour by these rays and at the same time avoid destroying the healthy tissue which surrounds it?* (Duncker, 1945, p. 1)

The following summarised fragment of a protocol by one of Duncker's experimental subjects gives an impression of the way subjects typically approached the problem:

> 1. *Send rays through the esophagus.*
> 2. *Desensitise the healthy tissue by means of a chemical injection.*
> 3. *Expose the tumour by operating.*
> 4. *One ought to decrease the intensity of the rays on their way; for example–would this work? – turn the rays on at full intensity only after the tumour has been reached. (Experimenter: False analogy; no injection is in question.)*
> 5. *One should swallow something inorganic (which would not allow the passage of the rays) to protect the healthy stomach walls. (E: it is not merely the stomach walls that are to be protected.)*
>
> ⋮
>
> 13. *Somehow divert … diffuse rays … disperse … stop! Send a broad and weak bundle of rays through a lens in such a way that the tumour lies at the focal point and thus receives intensive radiation.* (Duncker, 1945, p. 2)

Apart from the idea of using a lens, which does not work with γ-rays, the final proposal comes close to a realistic solution. Duncker interpreted the impracticable solutions proposed as arising from an incorrect representation of the situation; the proposed solutions *would* work if the situation were the way it is understood by the subject. Therefore, he stressed the role of the developing problem representation, and of reformulating the goal in particular. According to Duncker, reformulating the problem leads to a sharper, more specific problem representation that gives rise to new solution approaches and further reformulations.

1.2.2 Problem representations, mental models and schemas

Having demonstrated the importance of problem representations for problem-solving, we are ready to define precisely what is meant by the term problem representation and how problem representations compare to related constructs, such as mental models and problem schemas. 'The Penguin Dictionary of Psychology', (Reber, 1985) defines *representation* as:

> *A thing that stands for, takes the place of, symbolises or represents another thing. In studies of perception and cognition one often sees reference to the* mental representation *of a stimulus event which, depending upon theoretical orientation, may be characterised as a direct mapping of the stimulus (*see *direct*

*realism), an elaboration of the stimulus (*see *constructivism), a mental code of it (*see *idea, image) or an abstract characterisation of it (*see *proposition)…*

In the case of a problem representation, the 'stimulus' clearly is a description of the problem. As a consequence, representing a problem involves representing a physical situation. In addition, a sense of purpose is required to make it a *problem* representation. As became clear in the Duncker example earlier, a representation of a complex problem is an elaboration of the stimulus, rather than a direct mapping of it. The final two positions referred to in the above lemma touch upon the empirical issue of whether an object is represented visuo-spatially, or propositionally. We will return to this issue in later chapters.

In the sample from Duncker's radiation problem, the subject apparently had no previous content knowledge, and the representation of the problem had to be constructed from scratch. If the subject had had previous experience in the field, then episodes and abstracted situations recalled from long-term memory (LTM) could have helped to interpret the problem. We use the term representation for such situations represented in LTM as well, even though the original stimulus is long gone. So, in our view, problem representations might be found both in working memory and in long-term memory. A major difference between the two is that the current concrete problem is represented in working memory, whereas representations in long-term memory are independent of this particular situation and may be part of a larger knowledge structure.

Nowadays, mental model approaches are a dominant way of describing working memory problem representations. Two books have been seminal to the field of mental models research: one is a monograph by Johnson-Laird (1983), the other is a collection of papers edited by Gentner and Stevens (1983). In the latter volume, papers by diSessa, by Larkin, and by De Kleer and Brown are of particular interest in the present context. In both books attempts are made to define mental models. In the approach adopted by Johnson-Laird, mental models are primarily applied to syllogistic reasoning. In the papers collected by Stevens and Gentner, by contrast, the emphasis is on – physical – systems developing in time. As a consequence, Johnson-Laird's mental model theory does not comprise time, whereas in Stevens and Gentner's theory time has a central role. This may be the most profound difference between the two. It is our feeling that despite several significant differences between the two mental model notions, they have crucial aspects in common concerning the structure and semantics of mental models. The first point they agree upon is that a mental model is a structural analogue of the world. This implies that the elements of the mental model, and the relations between elements in the mental model, can be mapped in a one-to-one fashion to elements and relations in the world. A second point of correspondence between both views is how the mental model allows subjects to reason about the world. In Johnson-Laird's theory, the reasoning is about whether a proposition is compatible with the subject's model of the world. In Gentner and Stevens' theory, reasoning is about how a physical system will behave in time. In both theories reasoning is much like ‘seeing it happen in the mind’s eye’, as opposed to consciously applying rules.

Ultimately, formal physics problems are solved by applying solution procedures rather than by imagining the situation. The difference between the two processes is reflected in Young’s (1983) distinction between surrogate models, which are mental models in the sense we have discussed above, versus task/action mappings, which also are mental models, but with a different content. Task/action mappings, according to Young, are mental representations of goals states and of steps that have to be taken to reach these states. For problem situations with little structure of their own, such as the ‘Towers of Hanoi’ problem, the distinction between surrogate model and task/action mapping does not apply. For calculator use, which is the domain of Young’s research, the task/action mapping might indeed be different from the surrogate model of the device. For formal physics problems there most certainly is a difference between the representation of the situation and the representation of the solution approach. The concept of task/action mappings clearly reminds us of Newell and Simon’s (1972) notion of problem spaces, although their concept of problem space does not refer to a mental representation in the first place. As a final remark on Young’s task/action mappings, it has to be said that calculators are lasting devices, whereas a physics problem is instant, which might explain why Young does not clearly distinguish between long-term and working memory representations.

Representations in long-term memory can either be episodic memories of previous experiences (Tulving, 1983) or generalised representations. The generalised long-term memory representation is most commonly referred to as ‘schema’ (Bartlett, 1932; Rumelhart & Ortony, 1977), or, more specifically, problem schema (Hinsley, Hayes, & Simon, 1978; Chi et al., 1981). Schemas are stable mental structures that organise a set of related concepts, propositions, and procedures (Rumelhart & Ortony, 1977). In the specific case of a problem schema, knowledge elements are organised on the basis of their relevance to solving a type of problem. It is proposed that schemas comprise generic descriptions of situations, which play a role as active recognition devices (Rumelhart & Norman, 1985). The generic situation representation comes with default properties to describe the prototypical situation associated with the schema. When the schema is used in reasoning, these defaults can be overridden, however.

With schemas, like with working memory representations, we can distinguish between the physical situation and the actions to be taken to solve the problem. In ‘frames’ (Minsky, 1975), and ‘scripts’ (Schank & Abelson, 1977), the emphasis is on representing a – developing – situation. Martin (1984) distinguishes ‘plans’ as a type of schemas that are activated to perform particular tasks: ‘a plan is a set of sequenced cognitive operations that we apply to information to complete a task’. Problem schemas organise both knowledge of solution procedures and of situations (Chi et al., 1981), with the knowledge of situations in most views serving to trigger the schema in the proper circumstance (Chi et al. 1981; compare, ‘situational knowledge’ (De Jong & Ferguson-Hessler, 1986, 1996), ‘conditional knowledge’ (Alexander & Judy, 1988)).

Notwithstanding the differences, schemata and mental models are close relatives: both constructs describe the organisation of knowledge in clusters, and both types of theories give an account of how solution information is triggered when reading a problem description. Schema theories tend to emphasise the long-term structure in the mind whereas mental model theories in general are more focused on the structure that is built in the working memory during the problem-solving process. It is evident, however, that, to build a mental model, some form of long-term memory information is needed from which to construct the model. Moreover, if the mental model is supposed to trigger a solution schema in the long-term memory, it can be expected that there is a close correspondence between the elements of the mental model and the knowledge of conditions that resides in the long-term memory. In this dissertation *problem representation* will be used as a generic term to refer both to the working memory representation and to long-term memory representations. When we refer to representations in working memory or in long-term memory specifically, we use the terms *mental model* and *schema* respectively.

1.2.3 Roles of problem representations in physics problem-solving

In physics, a problem representation can have several roles. First, the problem representation guides interpretation of new information. This requires the linguistic structure of the story to be represented, so that references between parts of the problem description can be interpreted. It also requires the physical structure of the situation to be represented, so that topological relations (like *above* or *inside*) can be interpreted (Johnson-Laird, 1983; Kintsch, 1988). Second, the problem representation can be the basis for qualitative reasoning to infer further properties based on initial knowledge of situation properties, or to predict how the situation develops over time. This requires physics relations between elements (such as *attracts* or *fixed*) to be represented in such a form that the physical system can be simulated (Bredeweg, 1992; De Kleer & Brown, 1983; Johnson-Laird, 1983; Larkin, 1983). The final use of the mental representation is that it has to trigger the proper solution method for the problem. This requires a formal physics representation of the problem, which may be rather different from the information required for qualitative reasoning and mental simulation (Chi et al., 1981; Larkin, 1983; Plötzner, 1995).

It is important to recognise that in physics, many problems can be represented in multiple, formally equivalent, ways. Given the different roles the mental model plays, and accordingly the different kinds of information there are in the model, it is clear that a mental model is more than the enumeration of a minimal set of propositions that would suffice to specify the problem formally. Instead there are different representations of a single physical situation and each of the representations may trigger particular conclusions that are not drawn straightforwardly from other representations. As Sloman (1995) puts it, two representations that have the same *expressive* power still may differ in *heuristic* power.

The importance of the mental representation of a problem has now been recognised in several domains, ranging from an ill-defined domain like design (Goel & Pirolli, 1992) via medicine (Boshuizen & Schmidt, 1992) and chess (De Groot, 1946), to an allegedly well-defined domain like algebra word problems (Nathan, Kintsch, & Young, 1992) or physics (Chi et al., 1981; diSessa, 1983; 1993). It has been widely recognised that experts' mental representations are more helpful to problem-solving than beginners' representations are, and that proficient beginners' representations are more helpful than those of weak beginners (Chi et al., 1981; Chi & Bassok, 1989; De Jong & Ferguson-Hessler, 1986, 1991; Larkin, 1983).

1.3 Problem-solving competence and expertise

In recent decades, the 'expert-novice paradigm' has become a dominant methodological approach to the study of learning (Stewart & Atkin, 1982; Cooke, 1992). It compares experts' and novices' competence and knowledge in a certain field; the differences found are supposed to be informative of learning. Usually, experts are individuals who have many years of professional experience, or advanced graduate students. The definition of novices exhibits a far greater variety: in some studies novices do not have any domain knowledge prior to the experiment (Egan & Schwarz, 1979; Lewis, 1981), at the other extreme, 'novices' may have years of experience in the domain (De Groot, 1965). In most studies, however, novices are students who have taken a relevant course prior to the experiment (Chi et al., 1981; Chi & Bassok, 1989; Larkin, 1983, Ferguson-Hessler & De Jong, 1987, 1990; De Jong & Ferguson-Hessler, 1986, 1991)

The expert-novice approach has been criticised for interpreting between-group differences as developmental patterns, and for uncritically attributing differences in performance to the difference in experience, ignoring, for instance, differences in aptitude, age and formal training (Schoenfeld & Hermann, 1982; Willson, 1990). Both Willson, and Schoenfeld and Hermann conclude that a longitudinal approach could give more direct insight into developmental patterns. Still, even with a longitudinal descriptive approach, one could only prove correlations. Proving the causal role of some kind of knowledge or some process requires direct experimental manipulation. As a further problem with the expert-novice approach, several researchers found evidence that, in several disciplines, learning is not a monotonous transformation of novices into experts, but that there are qualitatively different intermediate phases (Boshuizen & Schmidt, 1992; Lesgold, Rubinson, Feltovich, Glaser, Klopfer, & Wang, 1988). Although expert knowledge remains the ultimate learning goal, information about intermediate levels might be more relevant to novice learning.

A slightly different approach that has been advocated by several researchers (Chi & Bassok, 1989; De Jong & Ferguson-Hessler, 1986, 1991; Ferguson-Hessler &

De Jong, 1987, 1990) is to compare successful and less successful learners in a domain. The amount of experience is about equal for individuals in both groups, but the successful students have become more competent problem solvers. With this approach, there is no claim that developmental patterns are described. Successful students' knowledge is more representative for weak students' learning goals than experts' knowledge, however, and instruction might aim at reducing the differences between the two groups.

1.4 Research questions and outline of the book

The main question we are addressing is how to improve physics student's mental problem representations. To answer this, other questions need to be addressed first. To begin with, the properties of mental representations need to be identified, especially the properties that are different between good and weak students or between experts and beginners. In addition, we need to understand how these differences arise in the construction of the problem representations. Only after these issues are clarified can our main problem be assessed fruitfully.

As the ultimate goal of our research is to enhance learners' problem representations, the research presented in this dissertation focuses on the relation between problem representations and competence. A person's individual cognitive style (Entwistle & Ramsden, 1983; Pask, 1976; Vermunt, 1991) may affect problem representations as well, and it may also affect what types of support are most effective (Riding & Douglas, 1993; Beishuizen, Stoutjesdijk, & Van Putten, 1994). Still, individuals clearly differ in the quality of their problem representations, and the quality of problem representations clearly affects problem-solving performance (Chi et al., 1981; Chi & Bassok, 1989; De Jong & Ferguson-Hessler, 1991; Larkin, 1983). Therefore, in most of our work, we concentrate on differences between competence groups, and only occasionally we touch upon the individual differences between members of a competence group.

The differences between experts' problem representations and beginners', and between those of proficient beginners and those of weak beginners, are broadly summarised as differences in 'depth' of representation. This is a very global term and it is unclear how these deep models are constructed. In Part A of this dissertation we address both issues. In Section 1 we try to unravel the concept of depth by giving a detailed empirical account of differences between experts', and good and weak novices' representations. In Section 3 we then proceed to identify reasoning processes that lead to the different problem representations of good and weak novices.

Part B reports on the design, the implementation, and the testing of a learning environment we designed to promote the construction of problem representations in novices, based on the findings reported in Part A.

Part A Problem-solving and representation

Al wat ik zelf begrepen heb
van en in dit ruimte-web
dat elke simpele draad beheerst
zijn de woorden ja maar eerst.

Zeg b.v. dat je 'vanzelf'
watertandt als je iets lekkers ziet.
Ja maar eerst
want zo eenvoudig is dat niet.

Eerst vormt zich, dames en heren
om van kinderen maar te zwijgen
–door draaiing van het hoofd en keren
van de ogen die hun bevelen krijgen van
hals- schouder- en oogbalspieren
die van elkaar bevelen krijgen
om elkaar te spannen of te laten vieren
en die van de hersens bevelen krijgen
die eigenlijk van de twee netvliezen
bevelen krijgen
waarop of waarbuiten je wilt kiezen
wat je het meeste interesseert–

...

Leo Vroman, 'Liefde sterk uitvergroot'

2. Competence-related differences in problem representation

Several studies have demonstrated that experts' representations are more helpful to problem solving than beginners' representations are, and that proficient beginners' representations are better than those of weak beginners (Chi et al., 1981; Chi & Bassok, 1989; De Jong & Ferguson-Hessler, 1991; Larkin, 1983). With a broad term, these competence-related differences have been summarised as differences in the 'depth' of the representation. In this section we go beyond the global concept of depth by giving a detailed account of differences between experts' and good and weak novices' problem representations. We begin by reviewing differences proposed in the literature, and then we search empirical support for the existence of these differences.

2.1 Aspects of situations knowledge

To start with, we review properties that have been ascribed to deep representations in literature. To structure our review, we group properties into four major clusters. First, two problem solvers may differ with respect to the *content* and the *structure* of their problem representations. Moreover, since proficient problem-solving requires the use of multiple representations, a further difference might be that some problem solvers use these multiple representations more *flexibly* than others. Finally, although the content and structure of a mental representation partially define how the representation is used in reasoning, the literature suggests that there may be other, qualitative, differences in the way people use their mental model of a situation. These differences are termed *dynamic behaviour*. In our review of mental representation characteristics below, we will discuss content, structure, flexibility and dynamic behaviour in this order.

Content

The first group of characteristics refers to the *content* of the representation. A first aspect of content is the abstractness of the entities referred to. In mechanics for instance, a force problem might be reformulated as an energy problem; the abstractness of the concepts is clearly different in both formulations. Several authors have suggested that domain concepts are ordered in a hierarchical way and that subjects start learning the concrete concepts, with the abstract concepts being acquired later (Reif, 1983; Reif & Heller, 1982; Van Hiele, 1986). Following this idea, one could assume that novices think of problem situations in more concrete terms than experts do.

A further aspect of content is the distinction of functional relations on the one hand and topological and geometrical relations on the other. De Kleer and Brown (1981, 1983) discuss how mental models of physical devices are constructed. According to their theory, the construction of a mental model begins with a topological model, which is a representation of the device's components and their physical organisation. Next, the functions of the components are inferred, and a causal model is constructed, which describes the functioning of the device. Finally, the causal model can be used to run a mental simulation. In physics problem-solving this is only an intermediate step, however, because the final outcome should be a formal solution. Running mental simulations has typically been associated with novice thinking (Goei, 1994; Larkin, 1983; Stenning, 1992). Experts, in contrast, make extensive use of the geometry of a situation to simplify the problem. Therefore, supposedly experts have detailed representations of the topology and geometry of a situation, whereas in novices' problem representations there is more emphasis on functional relations.

A final aspect of content is the contrast between qualitative and numerical representations. Several authors have suggested there are expertise-dependent differences in the use of quantitative and qualitative representations. McDermott and Larkin (1978; Larkin, 1983) describe how qualitative representations are associated with experts. Plötzner and Spada (1993) claim that experts' qualitative and quantitative representations are tightly related, whereas novices often do not relate these two types of representations.

Structure

As a first aspect of structure, there is a contrast between fragmentary and coherent models. In a fragmentary representation, pieces of information are small and only weakly connected, whereas in a coherent representation relations between the objects form part of the representation (diSessa, 1993; Hammer, 1994).

Second, Larkin's (1983) distinction between novice's tree-structured mental representations with single inference sources versus expert's graph-structures with multiple redundant inference sources points to a more qualitative difference in the way elements are mentally tied together.

Third, as a final aspect of structure, it has been claimed that all entities in an expert situation model need to have localised properties (Larkin, 1983; compare, de Kleer & Brown, 1981; 1983). This means that propositions referring to a part of a device should not refer to the overall device. De Kleer and Brown, who first introduced the concept, claim that mental simulations can be run more reliably if the mental model is built with localised properties. As we have discussed earlier, running qualitative simulations is typically associated with novice reasoning, however. Experts, in contrast, search to simplify problems by using global properties of situations such as symmetries and conservation principles. In electrostatics, for instance, global spatial symmetries play an essential role. Since these properties refer to the situation as a whole, it follows that localised properties in the sense defined by Larkin cannot be characteristic of expert mental representations. Also in these cases, however, there is a difference between the novice's initial representation of a property and the more advanced representation of the property. This difference can be described using the 'levels of understanding' introduced by Van Hiele (1986). According to Van Hiele, students who are in the initial phase of learning recognise a concept as a whole, while the properties that make the concept applicable remain inarticulate. It requires an advanced understanding to explain what properties make the concept apply in a situation, and it requires an even more advanced level of understanding to formally deduce whether a statement is true about a given situation. We will use the term 'non-locality' to refer to the fuzzy understanding of concepts in the field that is associated with – weak – novice problem representations, according to Van Hiele.

Flexibility

We define flexibility as the ability to use alternative representations adaptively. Thus, a first requirement for flexibility is that one has alternative representations available. In physics, unlike in an ill-defined domain like ethics, these alternative representations are formally redundant. Even so, they may trigger new approaches to simplify and solve the problem. Among the most important combinations of representations are combinations of concrete and abstract representations. Another important combination is between propositional and pictural descriptions (De Jong & Ferguson-Hessler, 1991). A final combination that is particularly interesting to physics problem-solving, is between the representation of the situation and the 'task/action representation' — a term coined by Young (1983), which refers to the representation of the steps to be taken to solve the problem.

A second requirement for using representations adaptively is that the alternative representations are consistent. If one has alternative mental representations that are inconsistent with each other, it is impossible to switch between them adaptively (compare, de Kleer & Brown, 1983).

Dynamic behaviour

As a final group of properties we consider the dynamic behaviour of the mental representation used in reasoning. Traditionally, mental models' approaches have been focused on simulation-based reasoning (compare, Johnson-Laird, 1983; Gentner & Stevens, 1983). As we have hinted at a few times already, experts use a different approach (Goei, 1994; Larkin, 1983; Stenning, 1992). Because the capacity of working memory is limited, simulation-based reasoning is only effective for relatively simple problems. In several domains there are powerful formalisms to re-describe situations in a more tractable way. In logic, for instance, experts can represent situations in terms of abstract groups, to which abstract relations apply, rather than using the representative sample approach as described by Johnson-Laird (Stenning, 1992). Stenning calls the two types of models analytic and agglomerative. In physics, experts can represent a process in terms of quantities that remain constant (constraints) rather than in the varying quantities that would be used in a time-based model (Larkin, 1983). Also, in static situations, such as those encountered in electrostatics, a given configuration can be re-described in terms of constraints (such as symmetries) that help to simplify the problem. It strongly depends on the kind of concepts a model consists of whether it supports time-based or constraint-based reasoning (Slotta, Chi, & Joram, 1995; diSessa, 1993; Fuson & Carroll, 1996). There is no one-to-one relationship, however, as one may state a problem in abstract quantities but then fail to use these to simplify the problem.

We have discussed several dimensions on which mental models may differ; Table 2.1 presents an overview of all properties discussed. In this study we explore how problem representations differ along these dimensions between persons of different expertise. Many studies have pointed to differences in problem representations, but there is a lack of methods to determine the qualities of problem representations in a detailed and direct way. Therefore, the first aim of our study was to devise a method that would enable us to gain a more direct insight into the qualities of problem representations in subjects. We use this method to address our main question: in what ways do problem representations differ between subjects, and between subjects of different competence levels in particular?

Table 2.1 Summary of characteristics of problem representations.

	mainly novice		mainly expert
content			
	phenomenological/concrete	vs.	*abstract entities*
	functional	vs.	*topological and geometrical relations*
	numerical	vs.	*qualitative specifications*
structure			
	fragmentary	vs.	*coherent*
	tree-structure, single inference source	vs.	*graph-structure, redundant inference sources*
	diffused	vs.	*localised properties*
flexibility			
	single	vs.	*multiple redundant representations*
	problem and solution are separate	vs.	*integrated solution information*
	inconsistent	vs.	*consistent representations*
dynamic behaviour			
	agglomerative	vs.	*analytic model*
	time-based	vs.	*constraint-based reasoning*

2.2 Method

2.2.1 Problem construction

A variety of methods for assessing problem representations can be found in the literature. An extensive review of these methods is found in Royer, Cisero, and Carlo (1993). 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 they discuss in relation to depth of representation are reproduction tasks, problem-sorting tasks, and asking learners to make judgements about problems or situations. The techniques for assessing mental models are all based on asking the subject to predict or infer something from a given situation description. None of these techniques directly assesses problem representations constructed by the subjects themselves. In all techniques the situation is presented to the learner, or the data only give indirect information about the problem representation, or both.

A straightforward approach to assessing problem representations would be to have subjects think aloud while reading and interpreting the problem. This would also give insight into the process of constructing the representation. The approach fails, however, because an initial problem representation can emerge instantly, even after reading the first few words of the problem description (Hinsley et al., 1978; Chi et al., 1981).

De Jong and Ferguson-Hessler (1991) attempted to assess problem representations constructed by subjects themselves from their own knowledge, by using a problem-recall task with very short display times for the stimulus problems. With such a set-up, subjects can hardly memorise anything, and most information reproduced would come from long-term memory. With this technique, which they refer to as 'problem reconstruction', the stimulus problem only serves as a cue to activate problem schemas in long-term memory. Nevertheless, subjects may remember fragments, and, more importantly, the stimulus provides a cue to reconstructing the problem in a particular format.

In the present approach matter are taken one step further by using a different cue, so that subjects have to construct the problem representations in their entirety. As one of the roles of problem representations is to give access to solution information, we expected that subjects prompted with solution information, might be able to tell in which situations this information applies. Because the connection between situation and solution information has a preferred direction (Anderson, 1983), this type of stimulus may fail to trigger all relevant situations knowledge. To obtain a more complete image of subjects' situations knowledge, subjects can be prompted with several forms of solution information subsequently. Now, as this task reverses the usual working order, automaticity will be broken, and we may fruitfully use a 'think aloud' approach to gain insight into the knowledge that is used to construct the problem representation. Our approach differs from the approach by Ericsson and Simon (1993), however, as we describe the resulting mental representation rather than the process.

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

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 the 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 2.1 One of the stimuli used in the experiment was Coulombs Law combined with the keyword 'surface charge' (original in Dutch).

Based on these considerations, we made a stimulus design shown in Figure 2.1. In this design the stimulus consists of one physics law, stated as a formula, plus a keyword referring to a sub-set of the problems for which the formula can be used. The keyword was included to make the subjects go beyond stereotypical problem situations recalled from textbook examples. The material was presented to the subject followed by some questions. First, they had to explain the meaning of the formula and the keyword, so that they would take notice of both. Then, they were asked to describe the type of problem corresponding to the stimulus. This question was included to stimulate the subjects to describe any generalised representation they might have. Next, they were asked to describe two different situations that would involve formula and keyword, and as a final question we asked them to summarise differences and similarities between the two situations. The aim of the final question was to give some insight into the relative importance the subjects attributed to the different elements in the problem description.

We selected eight laws and formulas that together covered a coherent part of the subject matter (that is, the electrodynamics course that is described in Section 1.1, see footnote 3 on page 3). These eight formulas were used to construct a total of twenty different assignments. 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. In total, every formula was presented two to three times, with a different keyword each time to obtain a more complete coverage of the subjects' knowledge.

Before turning the first page of the set of assignments, subjects were instructed to start thinking aloud immediately and not focus on the correctness of what they said. They were instructed to ask if they did not recognise a variable. Then they had to turn the page and start working on the first case. The first case was for training only. An analysis of the data suggests that subjects master the task sufficiently well after the training case to produce acceptable problem descriptions from then on. The task of the experimenter was to keep subjects talking and to remind them to answer the questions in the right order every now and then. No time limits were imposed in this experiment. All subjects were thus able to finish the complete set of assignments.

2.2.2 Subjects

In this experiment, we compared good novices to weak novices and novices to experts. The novice group consisted of proficient and weak first-year physics students who had just completed the course and who had attempted the final test. The experts were selected from two different subgroups: the ultimate expert in a narrower sense is the lecturer who teaches the course. The lecturer's expertise, however, does not reflect the kind of expertise the students are working towards all that well, and therefore, in addition three PhD candidates were included as experts.

The students were selected from a population of 103 first-year physics students. The students were classified as good or weak on the basis of their grades for the Electricity & Magnetism test and their grades for the Mechanics test. On the basis of this information we recruited six proficient students and six weak students for participation in the experiment. Proficient subjects were selected from those who had scored *amply sufficient* or better (grade '7' or higher) on the Electricity & Magnetism test, and who, in addition, had scored *sufficient* or better (grade '6' or higher) on their Mechanics test a month earlier.[4] Weak subjects were selected from those who had scored *low* to *insufficient* (grade '3' to '4') on the Electricity & Magnetism examination, and who, in addition, had scored no better than insufficient on the Mechanics test. Students were paid *f*15,- (approximately US$7.50) for their participation.

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. These, together with the grades for Mechanics, and Electricity & Magnetism, composed a set of nine grades in total, which showed a high internal coherence (Cronbach $\alpha > .90$, $n = 10$). The ranking of students according to this scale was consistent with the division into proficient and weak subjects based on grades for Electricity & Magnetism and Mechanics.

[4] In the Dutch school system, grades range from 1 to 10, 10 being the maximum score. The minimum requirement for a pass is a '6'.

2.3 Results

Since, the properties of problem representations are related to different scale-sizes (the structure of the whole representation, single propositions, and concepts used), there is no single way to analyse problem descriptions such 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 by all subjects. In the following, first the results of the analysis at the word level are presented, followed by a section on the analysis of sentences. Finally, the results of the situation-level analysis are presented.

2.3.1 Words

At the level of single words, we primarily expected the *content* of the mental representation to be reflected. The first content-related feature we discuss is abstractness. We expected that a more abstract mental representation could lead to the frequent use of words referring to abstract concepts. The second feature we focus on is the distinction between functional and topological models, which can be reflected in the types of relations mentioned (such as *carries* versus *above*). The final content-related feature is the contrast between qualitative versus quantitative representations, which could be reflected in the use of numbers, units and formulas. In addition to these content-related features we expected to find evidence of one structure-related feature, namely coherence. In a coherent mental representation of a physics problem, algebraic, logical, and spatial/temporal entities specify *attributes* on physical *objects* or specify *relations* between them[5]. The distinction between objects, attributes and relations is valuable, as it relates to the coherence of a problem situation at the word level.

For our analysis, all frequent 'physics words' had to be classified using the above aspects. Therefore, an inventory of frequent physics words was made. Prior to the analysis of word occurrences, all text spoken by the experimenter was removed from the protocols. Then, we made an inventory of the top-70 most used physics words in all texts (Appendix B). The rationale behind the particular choice of the number 70 is pragmatic: we thought it necessary that the list covered at least two-third of the physics word occurrences to make the conclusions representative for all physics words. Even when we adopt a wide definition of 'physics words', the top-70 covers about 80% of the total physics-word occurrences. The 70th word was used about 1.5 times on average by each subject. To exclude effects of individual subjects' idiosyncratic word patterns, it was verified that none of the words was used by fewer than three subjects.

[5] De Kleer & Brown (1983) present a framework in which these three elements are the building blocks of mental representations. Their view of mental representation is primarily grounded in artificial intelligence, however, and their approach is aimed at specifying mental models for robust simulations. In human mental models we find diffused properties, global constraints and redundant specifications, which cannot be represented in de Kleer and Brown's framework.

The author and another physicist independently scored the words according to the above aspects. If there was any doubt about the classification of a word on a certain aspect, the word got no mark on that aspect. Then for each aspect the words upon which the two agreed were kept, while the rest of the words got no mark. The distinction between abstract words and concrete words was the hardest one to make, for several words can refer to both concrete and abstract concepts. The word *capacitor* for instance refers to the very tangible physical object consisting of two parallel plates, as well as to the more impalpable concept of having capacity. As a consequence this word was scored neither abstract nor concrete. Even words like *place* or *position*, which are considered to be 'concrete' words, can refer to an abstract concept when they are placed in particular contexts: 'the force exerted on a point charge in a privileged position' — where the reference to symmetries is obvious. We also found some words that can represent either an object or an attribute depending on their context. *Charge* for instance is an attribute in a sentence like 'the charge of the sphere is 10 nano-Coulomb', whereas in the sentence 'the charge in the sphere tries to move to the outside of the sphere', the word charge refers to an object-like concept. The word *current* carries the same kind of ambiguity. Still, both raters agreed on the predominant meaning of many words. As there were hardly any quantitative descriptions in the entire set of problem descriptions, we excluded this aspect from further analysis, and we can only conclude that all subjects do have qualitative problem representations. The classification of all words with respect to the other aspects is presented in Appendix B.

We first turn to the contrast of abstract versus concrete representations. In Table 2.2 the average frequencies of 'abstract words' (such as *field*, *inductance*, *energy* and *flux*) and 'concrete words' (such as *plate*, *sphere*, *particle* and *charge*) are presented. The values reported, like all further values, were accumulated over all 19 stimuli. The data are somewhat complex to analyse because there are several groups of subjects, with multiple subjects in each group, and multiple observations of a nominal variable for each subject. Moreover, the number of subjects per expertise group is small. The best approach to such a situation is to use an exact test. For the given data structure, a permutation test is a suitable type of test (Good, 1994; Manly, 1983). 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 (Appendix A provides a detailed account of the procedure). All p-values are based on the outcomes after 5000 random permutations. As can be seen in Table 2.2, the weak students used proportionally fewer abstract words than the proficient students did. The overall difference between expertise groups is clearly not significant however, $p = .24$.

Table 2.2 Average frequencies and proportions of abstract and concrete words by level of expertise.

	lecturer (n=3)		PhD candidate (n=3)		good student (n=6)		weak student (n=6)	
	M (SD)	*%*	*M (SD)*	*%*	*M (SD)*	*%*	*M (SD)*	*%*
concrete	116.7 (51.5)	49.7	151.0 (42.0)	49.7	131.2 (43.9)	46.4	145.0 (21.4)	38.5
abstract	110.7 (39.1)	50.3	149.7 (48.6)	50.3	114.0 (37.7)	53.6	93.3 (36.0)	61.5

The next contrast to be discussed is that of topological versus functional words. In Table 2.3 the relative proportions of topological (such as *between*, *distance* and *outside*) and functional (such as *force*, *tension* and *cause*) relations are given. The proportions are remarkably similar between different expertise groups, and we found no evidence of any expertise-dependent differences, p = .98. So, at the word level, we found no evidence of any expertise-dependent differences in the *content* of mental representations.

Table 2.3 Average frequencies and proportions of topological and functional relations by level of expertise.

	lecturer (n=3)		PhD candidate (n=3)		good student (n=6)		weak student (n=6)	
	M (SD)	*%*	*M (SD)*	*%*	*M (SD)*	*%*	*M (SD)*	*%*
topological	18.0 (10.0)	43.9	29.3 (13.3)	48.5	28.7 (9.7)	48.8	37.3 (12.3)	50.8
functional	21.0 (2.0)	56.1	30.7 (12.0)	51.5	30.2 (9.1)	51.2	36.8 (14.9)	49.2

Beside content, we were also searching for structure reflected at the word level. Therefore, we now turn to the distinction between *objects*, *attributes* and *relations*. Average frequencies of these word types are given in Table 2.4. Using the permutation test, we found that the observed pattern of frequencies differs between expertise groups, p = .005.

Table 2.4 Average frequencies and proportions of object, attribute, and relation words by level of expertise.

	lecturer (n=3)		PhD candidate (n=3)		good student (n=6)		weak student (n=6)	
	M (SD)	*%*	*M (SD)*	*%*	*M (SD)*	*%*	*M (SD)*	*%*
object	125.7 (44.1)	49.5	193.0 (66.6)	48.5	154.5 (50.0)	53.8	125.5 (18.3)	45.8
attribute	89.3 (58.8)	35.2	145.3 (37.0)	36.5	73.7 (31.3)	25.7	74.3 (31.8)	27.1
relation	39.0 (9.2)	15.4	60.0 (25.2)	15.1	58.8 (18.1)	20.5	74.2 (12.7)	27.1

The expertise dependence in the use of object, attribute and relation words can now be further explored with correspondence analysis (Greenacre, 1984). This technique can be used to graphically represent the between-subjects' differences in frequency profiles. The analysis was done on a table of frequencies, with the subjects as rows and the types of words as columns. We computed a two dimensional solution for which the correspondence plots are presented in Figure 2.2. The first dimension represents 73% of the total inertia (that is, the association between rows and columns) in the table. As the column variable only has three values, the maximum number of dimensions in the correspondence analysis is two, and the plot in Figure 2.2 describes all the inertia.

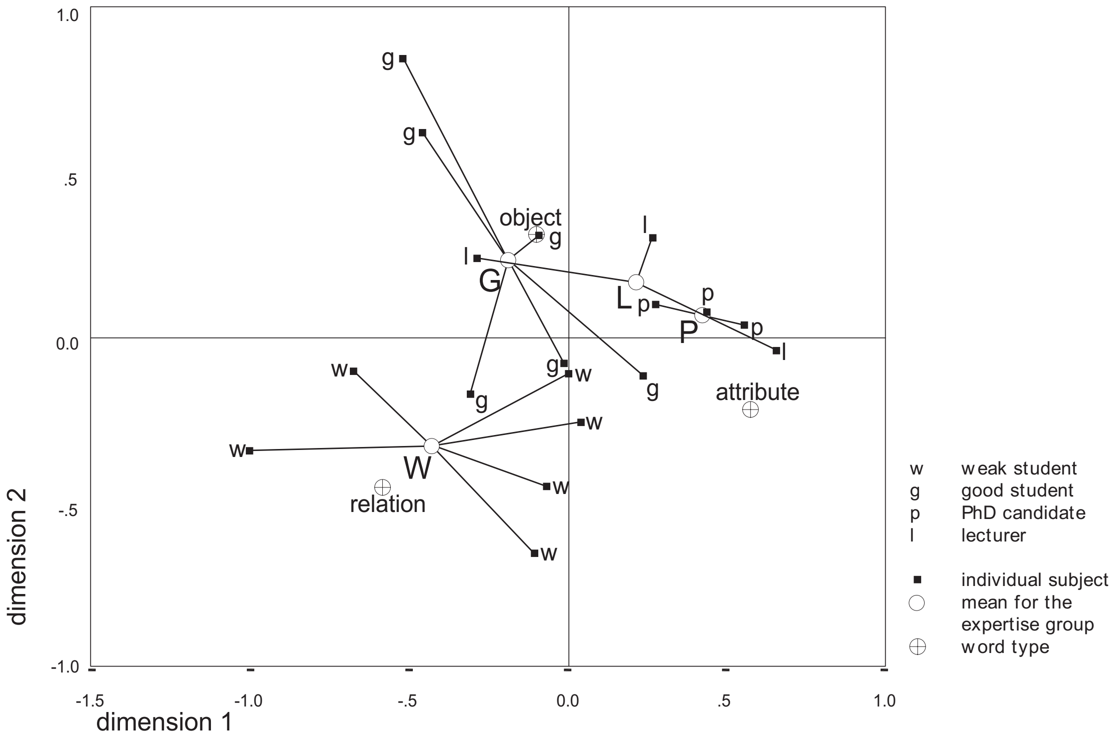


Figure 2.2 The relation between expertise level and the use of word types (correspondence analysis, canonical normalisation).

The origin of the plot represents the average frequency profile; subjects with more deviant frequency profiles are plotted farther from the origin. The word-type vectors (the vectors pointing from the origin to a 'word-type' label) point in the direction that frequent users of that type of word are found. In addition to values for individual subjects, the means per expertise group are indicated. The plot clearly shows that weak students mainly differ from good students through their greater use of relation words. Upon closer examination of the data, it appears that weak students mention relations without mentioning the objects that the relation is between more often than good students do. As an example of a relation with no object specified, consider the following fragments, both in response to Biot-Savart's law, with 'electric wire' as a keyword:

> *... in one case you have a straight wire, in that case a turned motion in the magnetic field is caused, and if you make the wire coiled, yes, it does go round, but inside the coil it is almost parallel ...* (weak student, id19)

> *... compute the magnetic field caused by the circulating current, compute the magnetic field caused by an infinitely long straight wire ...* (good student, id17)

In both fragments the relation words are underlined. In the first fragment it remains unclear what has a 'turned motion' relative to what, what is causing what, and what is parallel to what. In the second fragment relation words are used more properly. So, even though weak novices mention relation words frequently, this contributes little to the structure of their problem descriptions.

A further interesting difference that appears from the plot in Figure 2.2 is that the expert groups mainly differ from the novices by the fact more attributes are mentioned. This might be because the expert's problem descriptions are more precise, as in the following example, again in response to Biot-Savart's law with the keyword electric wire:

> *an <u>infinitely long straight</u> electric wire that has an <u>infinitesimal diameter</u>* (lecturer, id24)

To summarise our findings at the word level: word frequencies gave little evidence for competence-related differences in *content*. It appeared, however, that – weak – novices use words differently, and more in particular, they use words imprecisely. This was confirmed when we looked at the *structure*-related distinction between object, attribute and relation words. Here we found that weak novices frequently use relation words ambiguously, and that the precise specification of attributes is characteristic for experts.

2.3.2 Sentences

The second level of analysis is the level of single sentences. At the sentence level we expected to find aspects of *structure* that could not be assessed at the level of words. An aspect that should be reflected clearly at the sentence level is *redundancy*. A sentence where the subject elaborates on information mentioned before can be defined as redundant. We also expected the *locality* versus *diffusedness* of the concepts used to be reflected. If a subject constructs two different problem-situations using concepts with only a 'diffuse' meaning, the person might find it hard to explain differences between two situations. So, there is little information regarding differences in the protocol.

To assess the above aspects, we had to classify sentences according to their function in the problem description. With this goal in mind we developed a coding schema through induction. 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[6] and at each occurrence of the words 'or', or 'and'. The segmentation procedure was executed automatically. Though this procedure could be sensitive to variations in transcription style, it appeared that when the result was checked manually afterwards, only minor corrections were needed. After several iterative steps we arrived at the coding schema presented in Table 2.5.

---

[6] Dutch punctuation differs slightly from English punctuation; in Dutch commas are used almost exclusively to separate sub-sentences, rather than single words.

Table 2.5 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

We distinguished nine different functions of sentences, and we added a miscellaneous category for incomplete fragments and fragments that had no clear function. The categories as they appear in Table 2.5 are listed in order of priority, so a sentence was labelled with the topmost label applicable. As a consequence of the inductive approach we used, not all codes are directly related to the properties listed in Table 2.1. Here, we will focus on those codes that have a direct relation to these properties. To begin with, the codes *remark* and *explanation* were assigned only to exclude sentences from further analysis. *Episode* and *evaluation* cannot straightforwardly be placed in the framework we are using, nevertheless they may give relevant information about knowledge and learning (see Tulving, 1983; Kolodner, 1984, and Ferguson-Hessler & De Jong 1990 respectively). The ability to mention *differences* and similarities between situations relates to the 'diffusedness versus locality' aspect of structure, because it would be impossible to describe differences and similarities on the basis of concepts with a diffuse meaning alone. When solution *methods* are discussed within a situation description, we take this as evidence that the two are mentally integrated, which is an aspect of flexibility. Awareness of *goals* is considered important to problem-solving (Anderson, 1983), but the literature suggests no specific hypotheses about competence-related differences in the role goals play in problem representations. An *elaborated* description of a problem situation indicates that the subject has a graph-structured representation with multiple inference sources, which again is an aspect of structure. All other statements that introduced information about the problem situation were coded *given*. Finally, the remaining sentences were coded *miscellaneous*.

Examples of all categories are given in Appendix C. A small fragment by a proficient student shows how the coding was applied (good student id14 in reaction to the stimulus shown in Figure 2.1. Due 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 & Magnetism course. The inter-rater reliability was satisfactory, $\kappa = .79$, $n = 196$. The results of the analysis at the sentence-level are summarised in Table 2.6.

Table 2.6 Average frequencies and proportions of sentence types by level of expertise.

	lecturer (n=3)		PhD candidate (n=3)		good student (n=6)		weak student (n=6)	
	M (SD)	*%*	*M (SD)*	*%*	*M (SD)*	*%*	*M (SD)*	*%*
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
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
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
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
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
episode	3.7 (2.5)	1.1	5.5 (2.1)	0.7	0.0 (0.0)	0.0	3.0 (2.8)	0.3
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 count	340.3 (124.0)	100	522.7 (204.8)	100	242.8 (81.5)	100	291.8 (22.2)	100

Using the permutation test discussed earlier, we could demonstrate a difference between expertise groups, $p = .01$. To further explore the relations between expertise and sentence patterns, we made a correspondence analysis with the subjects defining the rows and the sentence types defining the columns. The inertia explained was 55% for the first dimension and 17% for the second dimension. We settled upon a correspondence analysis in two dimensions, because the total inertia explained with two dimensions 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 2.3.

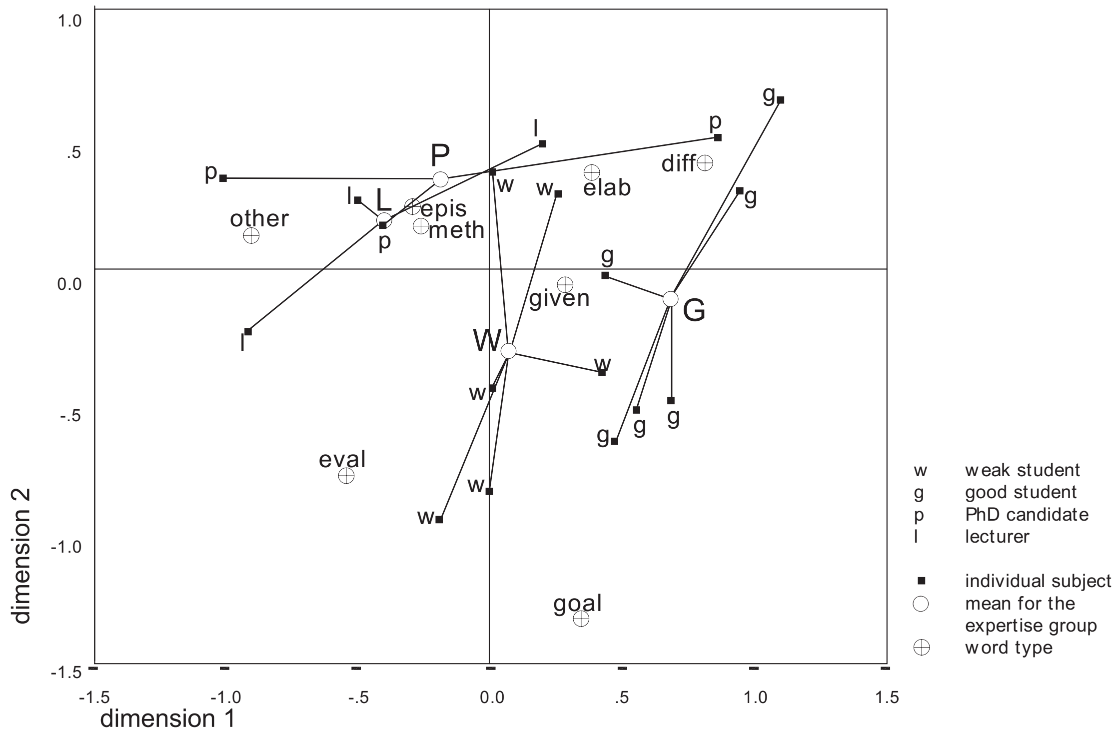


Figure 2.3 The relation between expertise level and the use of sentence types (correspondence analysis, canonical normalisation).

Figure 2.3 shows how the different expertise groups can be separated on the basis of their different use of particular sentence types. The expert-novice contrast is mainly along the vertical dimension, while the good novices are separated from the weak novices along the horizontal. The lecturers and the PhD candidates clearly overlap, and because their numbers are too small to permit any sensible analysis, both kinds of experts are taken together for further comparison. The graph shows that the frequent mentioning of *elaborations* combines with an emphasis on *differences*, and with few *evaluative* remarks. A subject's position along the dimension spanned by these properties reflects structure-related aspects. Good students differ from weak students through the greater number of elaborations and differences, and fewer evaluative remarks. Upon closer examination it appeared that most evaluative remarks by weak novices just stated a general lack of understanding, which is consistent with our interpretation. The graph also show that *goals* are almost exclusively mentioned by novices, whereas *methods* are more frequent among experts than among novices. These two properties span a dimension that is approximately orthogonal to the first. This second dimension could be interpreted as *where to go* versus *how to go*; it is related to the final aspect of flexibility, namely *problem and solution method are separate* versus *integrated*. The outcomes show that the 'expert versus novice' contrast is qualitatively different from the 'good versus weak' beginner contrast.

2.3.3 Entire situations

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 *flexibility*, we would like to know whether the subject has multiple representations of a situation, and in particular whether these representations are at different levels of abstraction. As regards the structure of the problem, we would like 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 too. We choose a set of four problem descriptions that were analysed for all subjects. All judgements were first made by the author and another physicist independently, after which differences were resolved by discussion. The entire rating procedure was carried out blindly with respect to subject's expertise level.

The first feature is the level of generality associated with the problem description. 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. We found that there is no unambiguous standard for deciding whether a single description is instantiated or generic. Besides it is doubtful whether such a measure would be very useful. Whether an utterance is considered to be instantiated depends on the context of the utterance, and in particular on the stimulus that the utterance was a response to. On the other hand, it is relatively clear whether a proposition is an instantiation of a foregoing one. As an example, consider the following fragments, both in response to Gauss' law with keyword 'surface charge'; the first by a lecturer, the second by a weak student:

> *when you think of surface charge, in this case that implies conductors, <u>so in the inside the electrical intensity amounts to zero</u>, <u>as a consequence there is no spatial charge</u> …yes, I think of a problem that includes a cylinder, cylinders put together, and spheres put together, leading to charges on the inside and on the outside, <u>in which case components cancel out due to symmetries</u>, <u>which makes the integral easier to evaluate</u>…*(lecturer, id7)

> *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…*(weak student, id11)

In the first fragment a redundant description is given with the redundant elaborations underlined. The first elaborations are more abstract properties of the conductor mentioned. The last two elaborations are abstractions of the concrete situation as well, but, in addition, they also refer to the solution method. In the second fragment, in contrast, all concepts mentioned are at the same level of abstraction. The quantitative outcome of the analysis is presented in Figure 2.4. It turns out that the use of multiple levels clearly increases with competence, $r_S = .86$, $df = 18$, $p < .001$. Weak students hardly use multiple levels. This implies, they either give an abstract description or they give a concrete description, but they do not connect the different levels.

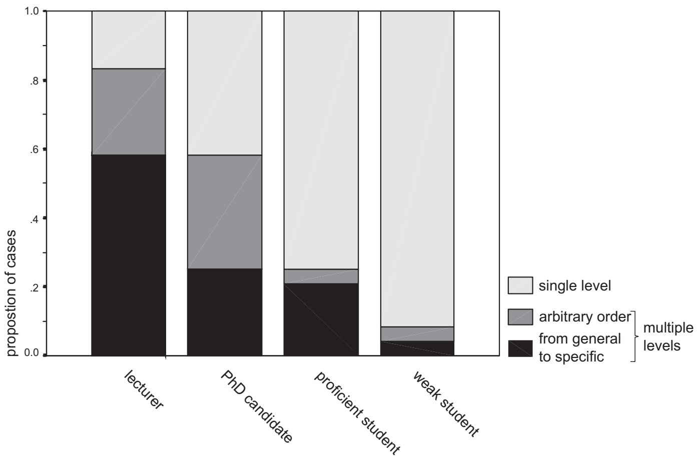


Figure 2.4 Presence of multiple levels in a problem description.

The next feature that has only significance for the entire situation is coherence. A description can be fluent or it may contain all kinds of strange leaps. In the extreme case it is hardly possible to give any meaning to the description. (In)coherence was judged on a five-point scale ranging from 'no incoherences' via 'a single incoherent step', 'some incoherent steps, still easily understandable', and 'several incoherent steps, hard to understand', to 'completely incoherent'. Compare, for instance, the following two descriptions, both in reaction to Gauss' law with keyword '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…*(weak student, id18)

> *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...* (good student, id16)

The first fragment, by a weak student, was considered to be 'with several incoherent steps'. The fragment, as a typical example, illustrates the fragmentary nature of weak novice problem descriptions: The fragment is mainly an enumeration of isolated elements and some propositions. In the first few lines the student is apparently trying to get some grip 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, for instance, 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. Though in most cases there is a sensible interpretation to the student's statements, many of the statements and especially the relations between them are unclear. In the second fragment, by a proficient student, there are far fewer ambiguities, and the relation between a statement and the previous statement is generally clear.

Scores are presented in Figure 2.5. The average coherence of a subject's problem descriptions depends significantly on competence, $r_S = .64$, $df = 18$, $p = .004$. The image is clear: lecturers' problem descriptions are more coherent than those of the others. Proficient students tell more coherent stories than weak students do; and the difference between PhD candidates and good undergraduate students is not too clear, apart from the fact that PhD candidates do not tell completely incoherent stories. We have also searched for inconsistent problem descriptions, but we could find only few, perhaps because most inconsistent descriptions were incomprehensible in the first place.

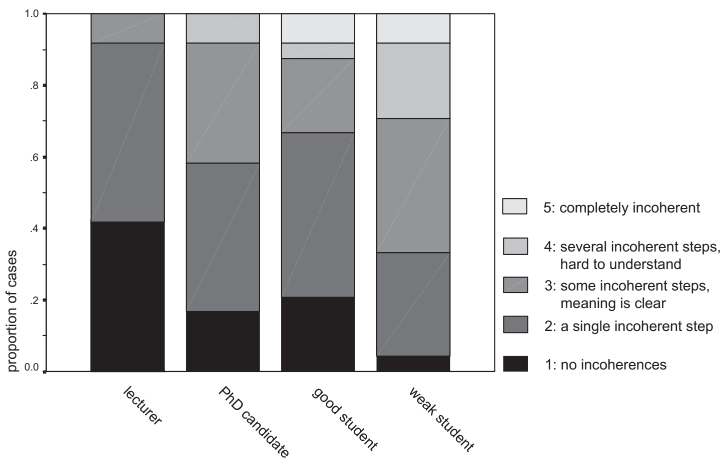


Figure 2.5 The incoherence of problem descriptions by level of expertise.

We had asked subjects to describe two different situations in response to a single stimulus. The next measure is how well they succeeded in doing this. Subjects could either describe two situations that were different in a way relevant from a physics point of view, irrelevant from a physics point of view, or they did not give an alternative at all. A typical excerpt from a weak novice problem description shows that alternative problem descriptions generated by weak novices do not involve any qualitatively different physics:

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 (weak student, id12)

In this example the situations described are identical for both problems, only the relative magnitude of the charges involved changes, which in this case does not lead to any differences in the solution approach. The student did not give any indication that he was dissatisfied with the alternative he described. This would be different in protocols by more proficient subjects, who, in general, produced more relevant alternatives and in case they had generated an irrelevant alternative they would be dissatisfied with it at least. We take this as evidence that, in the minds of more proficient subjects, solution methods are associated with problem representations. Figure 2.6 presents the quantitative findings on the presence of alternatives in problem descriptions. The increase of relevant alternatives with competence is significant, $r_S = .62$, $df = 18$, $p = .006$.

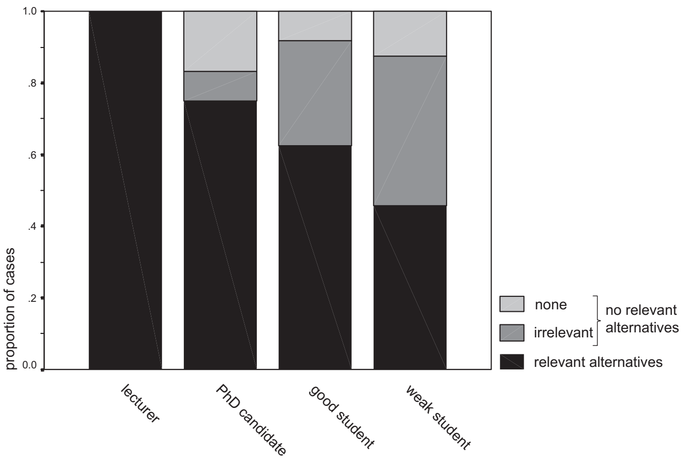


Figure 2.6 The presence of alternatives in the problem description by level of expertise.

Apart from giving alternatives, subjects were asked to make explicit 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 two examples by one of the proficient students:

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* (good student id14).

The quantitative results are summarised in Table 2.7.

Table 2.7 Average proportion of cases where differences were mentioned explicitly, by level of expertise.

	lecturer (n=3) *M (SD)*	PhD candidate (n=3) *M (SD)*	good student (n=6) *M (SD)*	weak student (n=6) *M (SD)*
difference in physical situation	0.50 (0.43)	0.50 (0.25)	0.42 (0.30)	0.38 (0.49)
difference in solution method	0.25 (0.00)	0.17 (0.14)	0.25 (0.22)	0.13 (0.14)

Apparently, the experts of both kinds mention somewhat more differences between the situations, but the relation with competence is not significant, r_S = .25, df = 18, p = .32. Differences in solutions methods were not mentioned often by either experts or students, and there do not appear any meaningful differences (r_S = .26, df = 18, p = .30).

2.4 Discussion and conclusions

This chapter set out to specify competence-related characteristics of physics problem representations. At three different levels of analysis we sought for characteristics that differ between problem solvers of different competence. An important gain from the analysis at different levels is that we could triangulate our analyses, and interpret findings at one level by the use of findings at another level. Our findings primarily apply to problems in formal conceptual domains — more in particular physics problems. As a further restriction, we limit our discussion to problems that can be solved using a known theory, such as the find-problems, proof-problems and prediction-problems found in standard physics textbooks. It is only for these types of problems that issues we have discussed in this chapter, such as expertise, depth of representation, and complexity of the problem, can be specified in a precise way.

The experimental method was based on the assumption that verbal descriptions of problems, retain essential characteristics of the underlying mental representations. There may be some problems with this assumption: firstly, subjects may censor their utterances; secondly, propositions may not be the format in which problems are represented mentally; and finally the verbal description of a problem representation could be disrupted by a lack of verbal ability. To solve the first problem we made subjects verbalise their thoughts concurrently. This method is known to be effective for think-aloud protocols (Ericsson & Simon, 1993). The second problem regarding the verbal nature of a problem representation is more fundamental. However, even through the stimulus material was designed to evoke different types of problem representations, there was a consensus among subjects about the format of problem descriptions. Problems were described verbally, and sometimes a drawing was added to clarify the geometry. There also was an agreement on the level of specification: numbers and variable names were generally left out. This is not too surprising as the theory of the domain is in a propositional format, and thus verbal reasoning can be assumed to play an important role. That could also be an answer to the third problem: if verbal reasoning is so important, poor verbal ability might very well result in poor mental representations.

We had proposed that expertise-dependent characteristics of problem representations can be ordered in four major groups: content, structure, flexibility and dynamic behaviour. We searched for content-related properties at the level of single words. At this level we found no evidence for advanced problem-solvers using different concepts in their mental models.

In the structure of problem representations, however, we found clear differences. We searched for structure-related differences at all three levels of analysis. At the level of words we found that weak beginners tend to mention more 'relation words'. Upon closer inspection it appears that the weak beginners mention these relation words without specifying between which objects the relation exists (for example, 'there is a force'). So, the weak novice's use of relation words contributes little to the overall coherence of the problem description. At the level of sentences, we found evidence that weak novices have few redundant elaborations in their problem description, and that they can only give a diffuse meaning to the words they use. At the level of the entire problem description we demonstrated a clear increase of coherence with increasing proficiency.

The next group of properties is related to the flexibility of the representation. The differences in the presence of multiple representations are clearly among the more convincing in this study. This suggests that the coupling between representations – and between concrete and abstract representations in particular – might be more important than the sole use of a particular type of representation. At the sentence level there was some evidence that expert problem-solvers have more solution information as part of their problem representations. The analysis of the alternatives that were mentioned by the subjects clearly shows that there is an increase in the amount of relevant alternatives mentioned with increasing expertise. We take this as evidence that in the minds of more proficient subjects solution methods are associated with problem representations. We had hoped to find further evidence on this point from the ways subjects explain the differences between the situations they mentioned. The information on the differences is hard to interpret however; the average scores on this aspect were low and several subjects were slightly irritated by being demanded to summarise differences they felt to be obvious. Still, the findings regarding solutions information indicate that in more advanced problem-solvers the representation of the problem is more tightly interwoven with solution representations than it is in less proficient problem-solvers.

The final group of properties is related to the use of the model in reasoning. Since our material consisted of static descriptions, and, in addition, most of the problems were about electrostatics or magnetostatics, we could not hope to find extensive information on the dynamics of reasoning. Nevertheless, we found qualitative evidence of time-based reasoning in problem descriptions. We can distinguish between problems that involve changing situations and situations that relate to one static situation only. In the case of changing situations, both experts and novices gave script-like descriptions as in the following fragment, which was a response to the formula for Lorentz force plus the keyword 'electric wire':

> *for example a, yes a bar with a rope attached, and a mass that drops down, and it pulls the bar forward like that, and then the bar rolls forward like that, and then here you have, square, so in fact you have a loop, and so the flux changes, and then you have, and then you can compute the what the intensity of the magnetic field must be, to make it stand still, so that a Lorentz force* (PhD candidate, id11)

In the descriptions of problems involving a static situation, we also found some descriptions that suggested a time course, but these were most frequently found in novice descriptions, as in the following example, which was a response to Coulomb's law with the keyword 'surface charge':

> *kind of problem can be a charged sphere, where the charge is found only at the edge, so surface charge, and that is because the charges exert a force on each other, which makes them try to get away from each other as far as possible, so that surface charge results* (proficient student, id14)

From the data a further difference emerges that does not fit in our framework straightforwardly. From the analysis at the sentence level, it appears that beginners mention 'solution goals' far more of than experts do. This tendency may be related to the difference between backward and forward thinking in problem-solving. Beginners tend to start their reasoning from that which was 'asked for', looking for a formula that has the desired entity as its outcome, whereas experts solve simple problems in a more forward fashion, starting from an analysis of the given situation (Larkin, McDermott, Simon, & Simon, 1980; Sweller, Mawer, & Ward, 1983; but see Priest & Lindsay, 1992). If – weak – beginners really are unable to construct a coherent representation of a problem, this sheds new light on the finding that they tend to skip the thorough analysis of the problem statement when they are trying to solve the problem (de Jong, 1986); since they are unable to construct a coherent representation, their best bet is to follow a kick-and-rush strategy, starting from something they need to know anyway: the entity that is being asked for.

From the present study it appeared that weak beginners mostly have rather incoherent single-level problem representations. Apparently, successful and more advanced problem-solvers elaborate on the initial elements to create a richer and more coherent representation of the problem. The finding that – weak – beginners have so many incoherences in their problem representation is suggestive of the hypothesis that the fundamental entity in a beginner's problem representation is not the entire problem situation, but rather a smaller unit. These differences led us to study the process of integrating initially unstructured propositions into an enriched problem representation. In the next chapter, we clarify this process, again with a focus on differences between proficient and weak students.

3. The importance of an enhanced problem representation

It is a well known finding that novices, when prompted to categorise a set of problem descriptions, do not sort the problems according to solution principles. Instead they sort the cards according to rather superficial similarities between the problem descriptions (Chi et al., 1981). This finding has been interpreted to imply that novices have not structured their knowledge in problem schemas. A problem schema is supposed to be an LTM structure centred on a solution method, which comprises knowledge of concepts and procedures, and generalised problem situation representations (see Section 1.2.2). These generic problem representations are the activation patterns for the schema, thus providing a link from a concrete situation to the appropriate solution method. Given this definition, the empirical findings could equally well imply that, although novices do know solution methods and generic problem types, they fail to map the problem at hand to one of the generic problem types they know. This could be caused either by the generic problem representations being too specific to be easily matched, or by the failure to elaborate on a given concrete problem description to rendering it a richer 'physics representation' that could activate a schema. It is our objective to discriminate between the above mechanisms and to judge their relative importance.

There are several studies, both empirical and theoretical, that provide us with evidence that the way problems are represented mentally, influences the physics problem-solving ability (Elio & Scharf, 1990; Larkin, 1983; Plötzner, 1995). In the previous chapter we identified characteristics of problem representations that differ between problem solvers of different competence levels. In that study we made a comparison between different kinds of experts, and good and weak novices who had to construct and describe problem situations. Among the major results from that study were the greater coherence and consistency of problem descriptions given by experts, compared to those of novices, and of problem descriptions given by good novices compared to those of by weak novices. Coherence, in this context, means that the relations between elements in the situation description are not just implied, but are made explicit. Consistency refers to the absence of contradictions in the description. A related outcome of the previous study was the difference in the use of elaborations: good novices, in contrast to weak novices, elaborated on a situation after giving an initial description (compare, Figure 2.3). The use of elaborations was even more prominent among experts. An important role of the elaborations in the problem descriptions analysed, was to express the relation between the concrete situation at hand and the abstract underlying physics structure of the problem. In order to clarify in what ways coherence, consistency, and elaborateness of the problem representation may influence the problem-solving process, we will propose a structure model for the elements of the problem-solving process in the following section.

The current study focuses on the differences between good and weak novices. From an educational point of view, the differences between good and weak novices are particularly interesting, because they may provide insight into what makes some students less successful than others, and that in turn may contribute to improving education. From the perspective of experimental psychology the difference between good and weak novices is interesting because both groups received about the same amount of training, and studied the same information, so that differences between the two groups can be attributed to differences in the way they processed the information.

3.1 The role of elaborations in physics problem-solving

3.1.1 The problem-solving process

In this section we discuss the psychological process of solving a physics problem. Our focus will be on 'true' problems — that is, problems which are not trivial from the viewpoint of the problem solver. This type of problem cannot be solved on the basis of direct recall of specific explicit features encountered in previous problem-solving experiences, but rather it is necessary to recognise the underlying physics-structure of the situation. Though even in these cases the final goal is well-defined in physics problems, problem-solving in physics shares some characteristics with problem-solving in ill-defined domains such as design (Goel & Pirolli, 1992), which are not prominent in traditional fields of problem-solving research like arithmetic, logic reasoning, and puzzles. In early work within the information processing paradigm, these problem types were the main focus of research. In this early research two major processes were distinguished in problem-solving: understanding and search. Understanding was considered the less interesting of the two and consequently most attention was paid to search processes (VanLehn, 1989; compare, Newell & Simon, 1972). For problems like the ones we use, understanding the problem is less trivial, and, moreover, the distinction between understanding and search is less clear cut than with well-defined, knowledge-lean, problems.

Among the features that distinguish physics problem-solving from 'simple' problem-solving, the most remarkable may be the amount of 'restructuring' that the problem requires (since there are many ways of restating and enhancing the information from the original problem description in ways that make more or less sense from a physics point of view). As a consequence, the problem space becomes an intricate structure where a route that leads to a dead end, and thus forces the problem solver to go back, still may result in a modified representation of the problem (Greeno & Berger, 1987; Goel & Pirolli, 1992). A second – related – feature common to many ill-defined problems and physics problems is the lack of a parameter that would indicate how close one is to a solution. Such a parameter would be a prerequisite to the fruitful application of weak general problem-solving strategies such as hill-climbing[7].

The process of restructuring the problem implies that, between reading the words of the problem description and finding a solution to the problem, the words and propositions from the initial problem representation somehow become connected, with the addition of knowledge from LTM, to form a more or less structured mental representation of the problem. A structured representation implies a rich (that is, redundant) mental encoding of the situation. Such a rich encoding provides more cues to select the proper solution representation, and, moreover, is more robust (that is, less vulnerable to inconsistencies). We will refer to these added assertions by the name of *elaborations*. Following VanLehn (1989, p. 539), we define elaborations as follows: *'An elaboration is an assertion that is added to the state without removing any of the old assertions or decreasing their potential relevance'.* VanLehn further comments that for problem types where elaboration plays an important role, the distinction between understanding and search becomes blurred. The process of elaboration on the initial problem representation is also known by the name of 'deep processing': Craik and Lockhart (1972), for instance, in their classic paper on 'levels of processing', argue that:

> *[The] conception of a series or hierarchy of processing stages is often referred to as "depth of processing" where greater "depth" implies a greater degree of semantic or cognitive analysis. After the stimulus has been recognised, it may undergo further processing by enrichment or elaboration. For example, after a word is recognised it may trigger associations, images or stories on the basis of the subjects past experience with the word. Such "elaboration coding" […] is not restricted to verbal material.* (Craik & Lockhart, 1972, p. 675)

Quite confusingly, the term 'depth' has thus been used both to refer to the amount of processing and, as we have seen in Chapter 1, to an inherent property of the problem representation itself. That the two uses refer to different meanings is illustrated by the following: on the one hand, in the case of a child on a merry-go-round, the observation that there is a friction force between behind and bench is just as deep (in the Craik and Lockhart sense) as the observation that the child becomes dizzy. The goal, and the prescribed context of physics theory make one observation 'deeper' than the other (in the second sense). On the other hand, in a problem where one is asked to compute the centripetal force for a point mass orbiting at a given radius from the origin and with given angular velocity, there is

---

[7] The accepted definition of a well-defined problem takes into account whether the initial state, the final state, and the operators are well-defined (Bunge , 1967 p. 137; Landa, 1969 (cited in Mettes and Pilot, 1980, p. 46); Simon, 1973; VanLehn, 1989). To the psychology of problem-solving these parameters might not be the most important ones, however. Chess problems are well-defined according to this definition, but, because of the large number of moves in between the initial and the final state, and because of the large number of operators (moves) that can be applied, knowing the desired final state and all legal moves is of little help. There may be more psychological importance to whether the sub-goals are well-defined; in chess – and in physics – generally they are not.

nothing deep (in the Craik and Lockhart sense) about the 'deep structure' of the problem. To avoid confusion, inherent properties of problem representations may be more aptly referred to in terms of naïve versus physics representations, or in terms of the specific properties we have discussed in Chapter 1.

If the problem description is textual[8], the initial mental encoding will closely follow the structure of the text. In the course of restructuring the problem representation, the mental representation will take a structure that becomes less dependent on the textual description and that is more dominated by the structure of the situation being described. This process has been recognised in model of language comprehension too. Kintsch, for instance, in his construction integration model proposes the following phases in the initial comprehension process:

> *a) forming the concepts and propositions directly corresponding to linguistic input; b) elaborating each of these elements by selecting a small number of its most closely associated neighbours from the general knowledge net; c) inferring certain additional propositions; and d) assigning connection strengths to all pairs of elements that have been created.* (Kintsch, 1988, p. 166)

Though the importance of elaborations in problem-solving is clearest for complex semantically rich problems like physics problems, elaborations may play an important role in other domains too. Logical reasoning, for instance, is a domain where much emphasis has been placed on situation-model based reasoning (Johnson-Laird, 1983). There is some evidence however, that, also in this domain, where the understanding of the problem has traditionally been taken for granted, restructuring the problem and elaborating on the original problem statement may play an important role in solving the problem. Polk and Newell (1995) present a reanalysis of reasoning protocols obtained with categorical syllogisms that were presented verbally. They propose that the reasoning process in these cases could equally well be explained in terms of verbal reasoning (that is, repeated verbal re-encoding) instead of manipulation of a situation model. They suggest that mental models do play a role in problems that demand the use of external knowledge, and in problems presented in a visual format, but that in verbally presented problems that require no external knowledge – such as logic problems – a verbal representation of the problem may suffice.

In our framework the process of elaboration has the role of transforming the initial mental representation of a problem, which is a sparse set of isolated propositions, into a much richer structure. In many studies it has been proposed that this evolution is controlled by a separate cognitive process, that requires specific control knowledge (such as metacognition, strategic knowledge, etc.). However, assuming such a general control mechanism introduces serious conceptual problems (Baddeley, 1986; Dennett & Kinsbourne, 1992) and, moreover, several studies have shown that content knowledge itself can organise

[8] If the initial problem description is in a pictorial format, the theoretical description of the process becomes more complicated because pictures can be recognised as a whole.

the reasoning process (Anderson, 1983; Rumelhart, McClelland, & the PDP research group, 1986). Since our interest in this study is in the role domain knowledge plays in reasoning, we will choose the second approach, stressing the associative structures in LTM that give rise to a self-evolving mental model — which is not to say that conscious control and directed attention have no role.

Next to reasoning about the problem situation, reasoning about the solution method is also an aspect of problem-solving. Both the situation and the solution method can be mentally represented. We call these structures the situation representation and the solution representation, respectively. Both the representation of the situation and the representation of the solution method may become active structures that guide further reasoning. The situation representation supports mental simulation, whereas the remembered solution method is a guide to formal problem-solving. We assume that there is a close connection between the two in proficient problem solvers who have their knowledge organised in problem schemas that involve both generalised situation models and solution approaches.

The process that we have described here is mainly data-driven at first. In the course of reasoning about the problem there is a shift towards schema-driven reasoning. This shift either occurs gradually, or as a moment of 'insight' that comes in an all-or-nothing fashion, comparable to the recognition of a Gestalt[9]. In the physics problem-solving context, the Gestalt would represent a meaningful type of problem. As soon as a 'schema-structure' is matched sufficiently well, the instantiated schema may direct further reasoning. This transition is a critical event in the problem-solving process.[10]

[9] A notable difference between the types of problems that were used in traditional Gestalt psychology (like Duncker's (1945) radiation problem, or Maier's (1931) two-string problem), and the physics problems that we used, is that in traditional Gestalt problems, there is a very direct relation between elaborations and problem-solving actions: elaborations always related to 'how to use the properties of an object' — like for instance the scissors that can be used as a mass to make a pendulum swing. In physics problem-solving, by contrast, understanding the situation is a goal in itself and it can be useful to enhance the mental model of the situation even without directly finding a formal solution method.

[10] With the development of expertise, the problem schema structure will become increasingly powerful. As a consequence, the schema is matched more easily and more rapidly, so that less data-driven reasoning is required for common problems. This transition is quite comparable to the knowledge encapsulation process that has been observed in physicians (Boshuizen & Schmidt, 1992)

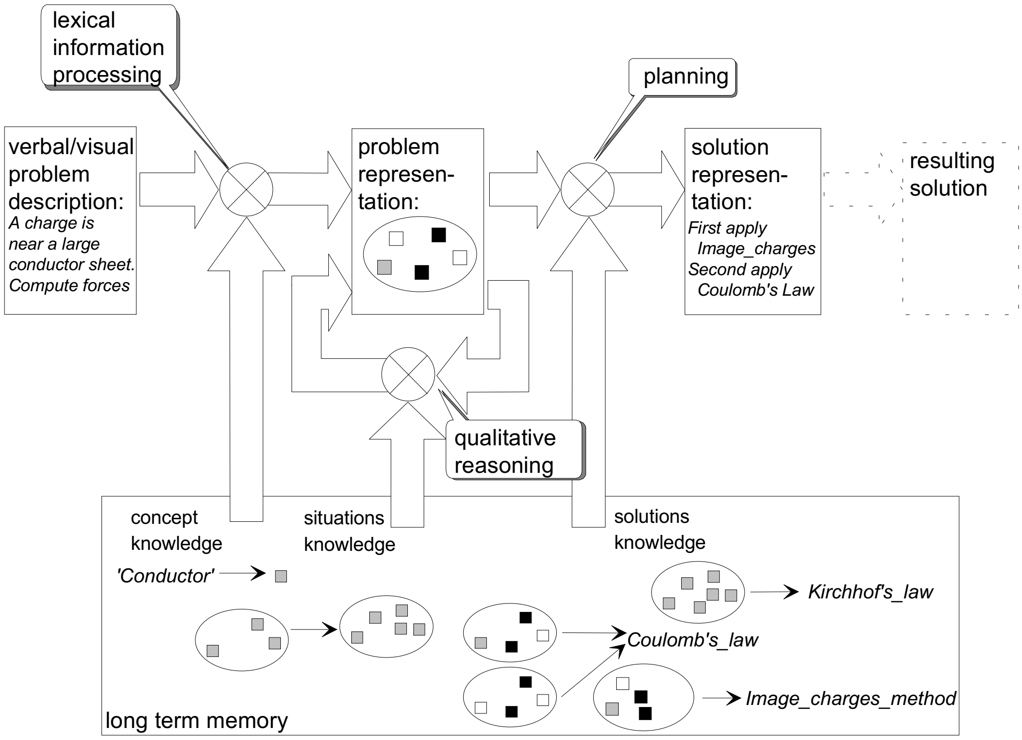


Figure 3.1 Model of the states, processes and information flows in the problem-solving process. In the bottom rectangle, the three types of LTM associations are depicted.

These considerations have led to a structure model of the problem-solving process that is depicted graphically in Figure 3.1. The aim of the model is not to give a chronological account of the problem-solving process, but rather to delineate the states, processes, and information flows that play a role.[11] In the model three subprocesses can be distinguished: the processing of lexical information, the qualitative reasoning process that elaborates on the problem representation, and the planning process that results in a solution representation. Beyond these steps, the implementation of the solution follows. The process leading to the implementation of the solution is beyond the scope of our research, and has been omitted from the model. Each of the subprocesses in Figure 3.1 corresponds to a particular type of knowledge structure. The first process, lexical information processing, demands knowledge of how words map to concepts. In Figure 3.1 this type of knowledge is represented by: 'conductor' → □. The second process, qualitative reasoning, elaborates on the problem representation, by adding matching elaborations to the current model of the problem. These inferences may be represented in LTM in the following format: 'IF property A [AND property B]

[11] In actual problem-solving, the processes outlined in the model will often alternate, rather than being executed in a simple linear fashion. This is not only true for beginners in a domain, but also for experts when they are working on complex problems (Chi, Glaser, & Rees, 1982; De Jong, 1986; Hayes & Simon, 1974/1979; Larkin, 1983).

APPLIES THEN ALSO property C APPLIES'.[12] In Figure 3.1 this type of knowledge is represented by: →. Finally, the third process makes an appeal to knowledge of solution methods. Again, knowledge of this type comprises a certain pattern of situational properties that has to be matched sufficiently well to activate the solution information. In this case however, the corresponding production would be of the format: 'IF property A [AND property B] APPLIES THEN APPLY method C'.[13] In the figure, this type of knowledge is represented by: →Kirchhof's_law. In the following we will focus our attention to what happens prior to the instantiation of the full schema.

3.1.2 Elaborations

As we have argued in the previous section, elaborations play an important role in problem-solving. In this study, our specific interest is in the role elaborations play in qualitative reasoning, especially during the initial phases of the structuring and restructuring of information from the problem description to form a coherent mental representation of the problem. Reasoning in this initial phase is in part automatic, and reasoning episodes may have a duration as short as a few hundred milliseconds (Shastri & Ajjanagadde, 1993). Consequently the depth of such reasoning is quite limited. This type of reasoning is known under such names as on-line inference (Noordman, Vonk, & Kempff, 1992), or reflexive reasoning (Shastri & Ajjanagadde, 1993). Due to its automaticity, the process that leads to the initial interpretation of the situation cannot be captured in think-aloud protocols (see also: Ericsson & Simon, 1993; de Groot, 1965). In most cases, at least with novice problem solvers, reflexive reasoning does not suffice to trigger a problem schema, so that a more deliberate qualitative reasoning effort is required. This effort leads to more complex elaborations being made, and deeper conclusions being drawn.

In line with our model, which stresses the associative nature of reasoning, we propose that there is no reason why elaborations should be activated in an all-or-nothing fashion. Rather, we adopt a mechanism similar to the one described in the ACT* model by Anderson (1983) stating that each node in a knowledge network has a certain activation strength. According to such a model activation spreads over the knowledge network in LTM, and the activation of a subsequent

[12] Alternatively, the mental model of the problem may be thought of as a pattern rather than a verbal statement. The LTM representation of the inferences would then be triggered by a matched activation pattern rather than a formal condition. From that viewpoint the rule-format that is presented here, may be accepted as an approximation to the 'real' structure (cf. Smolensky , 1988).

[13] It is important to recognise that a primary difference between reasoning about a problem situation, and reasoning about a solution method is that both processes have a different focus. In the former process, the physical system is the subject, whereas in the latter process, the reasoning process concerns solution procedures. Equivalently, in the former process, the question being posed is after a property of the system, whereas in the second process the question being posed is how to *determine* a property of the system. The two questions are closely related but not equivalent.

elaboration would be inversely proportional to the number of elaborations accessible from the initial knowledge element.

In order to gain insight into the types of elaborations that can be made, and the role they play, we analysed a number of physics problems. For each problem a hypothetical 'reasoning trace' was constructed, consisting of all the – atomic – reasoning steps that would be necessarily made to be able to plan the solution procedure for the problem. Underlying any conclusion to be drawn we posited a generalised rule: the 'inference rule'. These inference rules range from very simple, broadly applicable rules, to complex, highly dedicated rules. The conclusion that a body is a conductor, given that it consists of copper, is an example of a simple elaboration that can be inferred using the following rule:

> IF medium(O)=copper APPLIES THEN ALSO medium(O)=conductor APPLIES

This elaboration can be said to be at the lowest level of complexity, because it is a one-to-one mapping of a statement from the problem description to a new statement. A somewhat more complex elaboration can be inferred when two objects are concentric, and one has a smaller radius than the other. The conclusion that the first object is inside the second object may be drawn according to:

> IF concentric(O_1,O_2) AND shape.radius(O_1)<shape.radius(O_2) APPLIES THEN ALSO position(O_1,O_2)=in APPLIES[14]

This elaboration is at a higher level of complexity, because it combines two pieces of information to infer something new. Likewise, we could go on with higher levels of complexity. In our analysis we found several examples of statements that integrate five bits of information from the original problem statement to reach the next conclusion. Typically, these conclusions could be reached via intermediate reasoning steps that combined only two or three pieces of information at a time. In some cases, however, no intermediate steps could be defined, and in these cases a truly complex elaboration results. We may suppose that these highly complex elaborations correspond to impasses in the problem-solving process. A typical – hypothetical – reasoning trace tends to start with simple elaborations, and proceeds with more demanding elaborations being made when reasoning progresses. Two hypothetical 'reasoning traces' are presented in Appendix D[15].

---

[14] 'shape.radius(O_1)' Denotes the 'radius' aspect of the 'shape' attribute of object O_1. One may argue that the two objects both should have a regular shape in order for the above rule to be valid; adapting the rule accordingly would render it an even more complex rule.

[15] This type of 'reasoning' bears strong resemblance to traversing the problem space by executing productions using a breadth-first strategy. There is a difference in that there is no 'operation' that changes the state of the system like there is in most puzzle problems. A second important difference is that in the present type of reasoning the original problem description is not replaced with a modified one, but instead the problem description becomes enhanced with some new information.

The elaborations we have discussed tend to combine multiple pieces of information to generate something new. As is apparent from the foregoing, there are more and less complex elaborations. It may be conjectured that complex elaborations can only be made on the basis of the simultaneous awareness of all relevant problem features. Thus, many problem features have to be represented in working memory simultaneously. As working memory has a limited capacity for unstructured information (De Groot, 1965), this requires an integrated representation of the problem situation. On the other hand, as a result of adding elaborations relations between elements in the problem situation are established, leading to a more integrated problem representation. Therefore, elaborations may be seen as transforming a collection of initially isolated propositions into a structured mental model.

Now that we have shown that elaboration can be important for problem representation restructuring, the next issue to address is under what conditions particular elaborations are made. There are several factors that determine when elaborations are made, and what elaborations are made. Some factors lie in the individual, whereas other factors are determined by the input material, such as a problem description. An important factor determined by the problem-solver might be the goal the person has in mind. Some evidence that this factor not only influences conscious reasoning efforts but also the more automatic reflexive reasoning, comes from a study by Noordman et al. (1992), who addressed the influence of the goal a reader has in mind while reading a text. They showed that the reader's goal determines what inferences are made on-line while reading.

McNamara, Kintsch, Butler Songer, and Kintsch (1996) and Reder, Charney, and Morgan (1986) demonstrated how properties of the input material can influence the reasoning process. In both studies, the effect text coherence has on the reasoning process was investigated. McNamara et al. (1986) compared learning from more and less coherent study texts by high and low prior knowledge subjects. Two types of learning outcome were distinguished: recall of the text base and the quality of the situation model subjects had constructed. They found that the less coherent text impaired recall of the text in both high and low knowledge subjects. In contrast, the acquisition of a situation model was promoted by presenting a less coherent text for high knowledge subjects but not for low knowledge subjects. So, from this study we also have evidence for another problem-solver factor, namely prior knowledge. The findings are interpreted as evidence that high knowledge subjects are able to infer the missing relations themselves, and that the process of inferring helped them to engage in actively constructing a situation model. Reder et al. (1986) manipulated the elaborateness of texts too. They made different versions of a manual for a computer task. They had two types of elaborated texts: one version had 'conceptual elaborations' the other version had procedural elaborations (=syntax elaboration). The version with syntax elaborations led to superior performance on the task, whereas the conceptual elaborations had no effect. Though McNamara et al. (1996) did not distinguish between conceptual and procedural elaborations, the examples they give make clear that their manipulation affected the amount of conceptual elaborations in the study text. Thus, these studies illustrate that the effects of adding conceptual elaborations to a text may vary depending on the type of task and the performance measure that is used.

3.1.3 Research questions and related experiments

Now that we have discussed the problem-solving process, and the role elaborations may play in this process, we are ready to address our major questions: first, what makes beginners perform worse than experts when it comes to recognising the type of solution a problem requires, and second, what makes proficient beginners perform better than weak beginners? We started this chapter by describing two possible reasons why novices fail to recognise the proper solution type for a particular problem. Either they do not know any solution types at all, or the core problem to weak problem solvers is not that they have no schemata, or that they do not see solution principles as important, but rather that they fail to translate between the problem at hand and the problem types they know. Our first hypothesis is that the latter is true for the kind of novice we are studying (that is, students who have attended an initial university level course and who have attempted to pass the test).

A powerful method for investigating subjects' schemata is the use of categorisation tasks. In this type of task the subjects have to sort a pile of problems according to the similarities between their solutions. This task cannot be completed successfully using a mechanical trial and error approach. In contrast, it is necessary to recognise the global structure of the solution without actually executing all the steps, which is exactly the kind of reasoning we are interested in. Therefore, in the present experiment, we used the method of card sorting.

If it is true that even weak beginners basically know the types of solutions prevalent in the field, we should expect them to name their problem categories accordingly when they are trying to sort problems according to solution methods. Therefore, our first hypothesis predicts that a typical – weak – beginner sorting would not be systematically different from the experts outcome, but rather the – weak – beginners should be expected to come up with something like a blurred image of the experts sorting.

This kind of experiment, with similar subjects, has been done by Chi et al. (1981) too. They found major qualitative differences between expert sortings and novice sortings. However, in their sorting experiments they had only few subjects, so that they could not carry out a statistical analysis. Moreover, as pointed out by Taconis (1995), their study has some methodological limitations, both due to the instruction they gave to their subjects and due to the problem set they used.

With respect to the instructional format, it should be noted that Chi et al. (1981) did not give any indication of the criteria with which the problems were to be sorted. Differences in instructional format may explain the mixed findings in other studies too. Gruber and Ziegler (1995) for instance found that chess novices, when categorising chess positions, labelled their clusters qualitatively differently from the experts' labels. In the Gruber and Ziegler experiment the instruction gave no clue as to what kind of categorisation was desired. Taconis (1995) had high school students sort physics problems. He tested the effect of different kinds of instruction. He compared a version that gave no criteria, a version mentioning that 'problems you can solve in the same way' were to be put together, a version where examples were given too, and a version with all pile labels provided. Results clearly varied with the type of instruction.

With respect to the second factor, the composition of the problem set, it is important to note that Chi et al. used a problem set with 'misleading' cover stories in their experiment. In their article they reported on two problem sorting tasks. The first task (with 8 expert and 8 novice subjects) led to the conclusion that the novices gave different names to their clusters. In the second version of the task (where they reported on 4 subjects of different expertise levels) they deliberately chose their problems in such a way that the surface properties of the problem suggested a different ordering than the ordering based on solution procedures. In this experiment they found that the novice subject responded to surface properties, the expert responded to 'deep' properties, an advanced intermediate sorted according to solution principles but erred sometimes, and a less advanced intermediate subject did some hybrid sorting distinguishing piles both on the basis of solution principles and surface properties. In physics, however, unlike in algebraic word problems, the 'cover story' of the problem is not normally independent of the deep structure. Problem statements in traditional university level physics text books generally go no further than the relevant physics context. The problems used in our experiment were formulated accordingly. In many recent text books (see for instance, Halliday, Resnick, & Walker, 1993; Young & Freedman, 1996) practice problems are embedded in a 'human interest' cover story, which is also the way in which problems were presented in the experiments by Chi et al. (1981). Such human interest cover stories can be quite unrelated to the deep structure of the problem.[16] It is conceivable that this type of cover stories conveys a kind of information that is particularly salient to beginners, and that it is this type of information that caused Chi et al.'s subjects to induce the superficial sorting criteria.

If our first hypothesis proves to be true – that is, if novices appear to know some rudimentary form of a problem schema – their frequent failures to recognise the proper schema can have two different causes. Either it is because their conception of the generic problem type is too narrow or it is because they fail to elaborate on the current problem properly. If the elaborations play the role we assume they do,

[16] As an example consider the following problem where most of the cover story can be neglected: ***What Shakespeare Didn't Tell Us***. *Romeo is tossing pebbles at Juliet's window to wake her. Unfortunately she is a sound sleeper. He finally throws too large a pebble too fast. Just before crashing through the glass, the pebble is moving horizontally, having travelled horizontally a distance* x *and vertically a distance* y *as a projectile. Find the magnitude and direction of the pebble's velocity as it leaves Romeo's hand.* (Young & Freedman, 1996)

it may be supposed that, depending on the subject's level of expertise, some elaborations are made automatically upon first sight of a problem. Other elaborations are simply too complex and will never be made, and even if they were made (or told) they would not trigger any relevant thought, because they are not connected to anything known. In between these two extremes, there must be an intermediate level where elaborations no longer are unproblematic. It may be supposed that presenting someone with an elaboration of this level of complexity may provide a scaffold, that enables the problem solver to solve a more complex problem. If the level of the given elaboration is too high, it will not connect to anything, and thus it will not help. If the level of the given elaboration is too low, however, the given information may even hamper the active reasoning process, and thus the formation of a situation model (McNamara, 1996). Whether an elaboration is too easy, too difficult, or has the proper level, depends on the proficiency of the problem solver. If both proficient and weak beginners do engage in this kind of elaborative reasoning, we should thus expect that when elaborations are given at a very simple level, weak beginners profit more than proficient beginners do, which is our second hypothesis. In order to test this hypothesis we made another version of our problem-sorting task with an extra elaboration given with each problem. We adjusted the elaborations we gave to be as simple as possible, in other words, to be very close to the original problem statement.

3.2 Method

3.2.1 Material

As a subject matter we chose the field of electricity and magnetism. The particularities of this domain compared to other domains were discussed in Section 1.1. In order to measure the effect of elaborations, two different sets of problems were required, to let each student work on an elaborated and a non-elaborated version of the task. We constructed two sets of problems from the domain of electricity and magnetism. One set comprised 20 problems from the field of electricity, the other set consisted of 20 problems on magnetism. Our intention was to design both sets in such a way that an expert would distinguish four different clusters of problems in the set. We tried to make the problems within a cluster sufficiently varied, both in their wording and in their physics content. The constraint we used was that we did not want to include catch problems of types the students did not regularly encounter in their practice problems. The design procedure started from a larger set of problems that was presented to several experts[17]. Problems that were not classified in a more or less consistent way were removed from the set, as well as problems that were too hard for undergraduates according to the experts. The composition of the remaining two sets of 20 problems each is summarised in Table 3.1. Then we used the

[17] An expert was defined as someone who had recently been involved in teaching the subject of electrodynamics at the undergraduate level.

judgements of four independent experts to determine which problems within a set were very similar (that is, belonging to the same cluster according to at least 3 of the experts) and which problems were very dissimilar (that is, none of the experts placed them together). The results are summarised in Appendix E. These results were the standard against which the students' sortings were judged.

Table 3.1 The 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	Ampere's law	6
image charges	5	dipole approximation	5
dipole approximation	5	induction/flux	5
Coulomb's law/superposition	4	Biot-Savart's law	4

For each of the problems an elaboration was constructed that had a low level of complexity. Hence, there were two versions of each set, one with and one without elaboration. All students were to sort one of the sets with elaborations and the other set without elaborations. Examples of the problem cards are given in Table 3.2.

Table 3.2 Two examples of problems used in the experiment. For both problems there was an elaborated version and a non-elaborated version. The non-elaborated version only gave the normal printed text, the elaborated version also gave the italicised text.

electricity, Gauss' law problem	magnetism, induction/flux problem
E17 Compute the field of a planar charge distribution that extends to infinity *The field of an infinitely large planar charge distribution has a field component perpendicular to the plane only*	M2 A coil with radius r, length l and N turns, is being rotated at angular velocity ω in a homogeneous magnetic field. The axis of rotation is perpendicular to the field. Compute the voltage induced in the coil. *As a consequence of the rotation the magnetic flux through the coil changes*

3.2.2 Procedure

From previous research it appeared that the instruction that is given to the subjects strongly influences the outcome of the experiment (Taconis, 1995). In the present experiment, the goal was to investigate how well the subjects were able to form meaningful categories based on solution methods and how well they were able to see to what category each problem belongs, instead of testing whether the subjects would perceive this type of ordering as the most natural one. Therefore, the instruction was quite explicit about the ordering principle that was intended in the experiment. A straightforward way of laying out the intended structure would be to give an example. However, an example from an adjacent physics domain (say mechanics), was felt inappropriate because it would give too much information. With these considerations in mind it was decided to base the instructions on an example from an entirely different domain: cooking. The instructional text explained how a cook might answer when he would be asked what dishes where similar. Examples mentioned were cream of chicken soup and cheese sauce that were categorised together because both recipes involved preparing a roux, and an apple turnover and a savoury pie that both involve making pastry and baking in the oven. The instruction went on saying that the student was supposed to read all problem cards in the present set, prior to doing any sorting. When the sorting was done the student would copy the numbers of all problems in a category on a results form, together with the most appropriate name for that category. Subjects who asked for a proper number of clusters were told that neither one big cluster containing all problems, nor 20 separate problems were considered desirable.

Before starting, the subjects were told that the first set would take approximately 50 minutes to complete. Then there would be a coffee break of about 15 minutes after which the second set would be sorted in about 40 minutes. Subjects who could not finish the first task within 50 minutes were permitted to work on during the coffee-break. On the second task, subjects could work on until they were finished, but only a few people spent more than 40 minutes.

3.2.3 Subjects

Subjects were first-year physics students who had completed an initial course on electrodynamics in vacuum. Based on power considerations, it was estimated that we needed about 80 subjects to establish the interaction between the effect of elaboration and the student level. Subjects were recruited from two different universities (hereafter 'University A' and 'University B'), because the population within each university was too small to provide us with the subjects we needed.[18] First-year students selected from the faculty's phone directory were approached by telephone until the desired number of 80 participants was reached. Subjects were paid *ƒ*20,– (US$10 approximately) for their participation.

Students were classified as proficient or weak students on the basis of past test results. Both high school final examination grades and the scores obtained on several university physics tests were available. In Section 2.2.2 we already analysed the coherence among several grades. The current sample comprised far more subjects, so it rendered more accurate information about relations between grades. After comparison with the current set of data, we dropped high school biology from the scale because correlations with the other grades were rather low.

Due to the use of subjects from two universities, the university physics test results were not directly comparable for members of the different subgroups. The problem was resolved by first converting all scores into z-scores; then the means

[18] There are some differences between the universities, and between their respective populations, that should be noted here. At University A, the course on electrodynamics is placed in the entrance semester, at University B, electrodynamics is taught in the second semester. This difference may influence the attitude the students have towards physics learning at the time they studied the subject. A second difference is that at University A, first, both magnetism and electricity are treated for in-vacuum systems. This course is followed by a second course that includes the presence of dielectric materials. At University B, in contrast, electricity, with and without dielectric is covered in a first course and in the sequel to this course, magnetism gets full coverage. A third difference between the two institutes is that University A is a polytechnics institute, whereas University B is a general university. There is no evidence of any qualitative differences between the two sub-populations' competencies, however, and we therefore treat them as a single population.

and variances on the high school final examination scores were determined for each subgroup. It appeared that the variance was about equal for both subgroups, but that the mean of the high school final examination scores for students of university B was significantly higher. With the help of this information, the university test scores for both universities could be matched. Due to the construction procedure of the scales from sub-scales, we were left with reliability coefficients for the sub-scales and for the combination of scales. As a reliability coefficient Cronbach α was used. The following values were obtained: high school final examination: $\alpha = .86$; test scores university A: $\alpha = .88$; test scores university B: $\alpha = .93$; the reliability of the resulting combined scale: $\alpha = .88$. The latter value is considered to be quite acceptable. On the basis of the resulting scale, a median split was carried out, resulting in high and low performance groups of equal size that were used for further analysis

The experiment could only be carried out when all the relevant material had been covered in the Electricity & Magnetism course. To make sure that the students had really spent some time on the topic so that they would at least understand all the words in the problems descriptions, the experiment could not be carried out before the final test had been taken. For University A this implied that we could start experimenting after the first semester had been completed. Due to practical reasons there was a lapse of three months between the first opportunity to take the final test and the start of the experiment. In fact, during this period, some of the subjects had taken part in the second opportunity to do the final test already. These circumstances may have caused some extra 'noise' in the data. At University B coverage of the relevant topics extended over two courses that spread across the entire second semester. At this university the experiment was started immediately after the final test of the second course.

Effects of order were averaged out by using the design shown in Table 3.3. The distribution of proficient and weak students over the cells was controlled for by asking the students, prior to assigning them to a condition, whether they had passed their first Electricity & Magnetism test or not. Since only about half of the sample had passed the test, this criterion led to equally sized successful and weak groups.[19] The distribution of subjects from both sub-populations over experimental conditions was also controlled.

[19] This – rough – criterion was only used to control the distribution of subjects over experimental conditions. In the analyses we used the criterion that was described in the previous section, which takes into account performance on several more previous tests.

Table 3.3 The experimental set-up

	first with elaboration	first without elaboration
proficient student		
first electricity	$n = 10$	$n = 10$
first magnetism	$n = 10$	$n = 10$
weak student		
first electricity	$n = 10$	$n = 10$
first magnetism	$n = 10$	$n = 10$

3.3 Results

3.3.1 The search for a 'weak-beginner criterion'

The first question we will try to answer is whether the better and weaker students apply different sorting criteria from the criteria applied by the experts, or whether they try to apply the same criteria and are just less able to do so. To answer this question, we considered both the clusters that emerged and the types of pile names students used. We performed the same analysis both for the electrostatics problem set and for the magnetism problem set.

Table 3.4 Average numbers of piles for all groups.

		electricity	magnetism
experimenters' proposed categorisation		4	4
other experts* $n = 3$	*M (SD)*	6.67 (0.58)	6.33 (2.08)
proficient students $n = 36$	*M (SD)*	5.97 (1.50)	5.64 (1.51)
weak students $n = 35$	*M (SD)*	5.94 (1.73)	5.69 (1.49)

*) The problem sets presented to the experts were slightly larger than those presented to the students

To compare the sortings for proficient students and weak students, the student level scale had to be dichotomised. We chose to draw the cut-off line at the median to get two groups of equal size. Out of the 80 subjects who participated in the problem sorting experiment, 9 subjects had either omitted a card from one of their problem sorting forms, had mentioned a card twice, or could not be classified as proficient or weak because of missing information. Therefore, the number of valid observations for all the analyses in this study will is $N = 71$. After applying the median split procedure, we were left with a group of 36 proficient students and a group of 35 weak students.[20] In Table 3.4 we give, as a first descriptor of the data, the mean numbers of clusters we found for good and weak novices and for experts for both problem sets.

[20] The general performance level of students at University B was somewhat higher than that of students at University A. Therefore, in the good students group we find 16 students from University A and 20 students from University B. The weak students group consists of 21 students from University A and 14 students for University B.

We used a hierarchical cluster analysis to study the clusters of problems that emerged in the proficient and weak students' sortings respectively. For the purpose of this analysis we made no distinction between the elaborated and non-elaborated versions, because that would leave us with groups too small to permit any sensible analysis. Moreover, we had no reason to expect that the elaborated and non-elaborated versions would be sorted in fundamentally different ways.

The hierarchical cluster analysis procedure demanded that the data were structured in a particular way: for all the subjects in the experiment a correspondence matrix was computed for each problem sorting. Such a correspondence matrix consists of 20 rows, one for each problem card. There also are 20 columns, one for each problem card. A position takes the value '1' if two problems are placed in the same pile, or if the position is on the diagonal of the matrix. Otherwise the position takes the value '0'. The correspondence matrices for all subjects were concatenated to a single file. From this file Euclidean distances between problems were computed for both the below and above median groups. Then a cluster analysis was performed using the 'average linking' algorithm.

The results were summarised in dendrograms that are presented in Appendix F. As can be seen from these dendrograms, the expert clusters can be clearly recognised in both the weak and the proficient students sortings. The main difference between proficient and weak student clusterings is that the distances within a cluster are larger for the weak student sortings, which implies that the problems within the cluster are less tightly bound together. In general where the student clustering deviates from the 'norm', there is at least one of the experts who 'deviated' in the same way.

For the electricity problems, the proficient students' clustering was entirely according to the norm except for one problem card (E14). The 'misplaced' card was also the card that had the weakest association to its cluster. The resulting combinations with three problems in the cluster Coulomb's law/superposition were made by none of the experts. In the weak students' clustering three cards were placed in clusters different from the 'norm' (E5, E17, and E18). These three cards that were placed differently rendered only one combination that none of the experts made (E9 with E17). For the magnetism problems, the clusters for proficient and weak students were identical. Both the proficient and the weak students placed two cards (M7 and M11) in piles different from the 'norm'. Four of the resulting combinations of M11 and problems in the cluster 'Ampere' were made by none of the experts. Two of the cards that posed most problems to the novices are presented in Table 3.5.

Table 3.5 Two frequently 'misplaced' problems

E14 Given are two parallel thin metal cylinders, carrying opposite charge Q and -Q. The length of both cylinders amounts 1 [m] and the distance between both cylinders amounts 1 [mm]. Compute the field in a position 1 meter away from the wires in the perpendicular bisectric plane to the cylinders (intended cluster: dipole approximation)	M11 Given is a densely wound flat coil that has N turns. The inner diameter of the coil is r the outer diameter of the coil is 2r. All turns of the coil are in the same plane. Compute the field at the centre of the coil (intended cluster: Biot-Savart's law)

Though the cluster analysis did not suggest a fundamentally different 'weak beginner problem-sorting criterion', we went on analysing the names subjects gave to the piles they made, searching for the 'weak beginner criterion'. Because this is a more laborious type of analysis, we only did this analysis for a subset of the data. We took the labels invented by 10 subjects who were randomly selected from the 20 most proficient subjects, and compared them to the labels that were given by 10 subjects who were selected from the 20 least proficient subjects. Since all the subjects had made two problem sortings, we had two collections of labels, one for the electricity problems and another for the magnetism problems. In total the selected subjects had 107 named clusters for the electricity problems, and 105 clusters for the magnetism problems. To enable a comparison between labels, all labels with an equivalent meaning had to be clustered together first. This was done independently by two physics experts, after which differences were resolved by discussion. During the process the experts did not know whether a label was given by a proficient or by a weak student. The experts agreed that there were 20 different meanings in the electricity labels, with 7 unclassified labels. For the magnetism problems, there were 14 categories of meaning, with 13 labels unclassified. The resulting content categories were then classified according to the type of information they expressed. The resulting scheme is presented in Table 3.6.

Table 3.6 Taxonomy of category labels (items printed in italics represent a *class* of content labels rather than actual labels)

type of label	electricity label category	magnetism label category
strong physics procedure	Gauss' law Gauss' law in differential form Coulomb's law image charges dipole approximation	Ampere's law Biot-Savart's law dipole approximation
general physics procedure	superposition & *algebraic meth.* potential & *algebraic method* charge distribution & *method*	superposition & *algebraic meth.*
physical relation	force in field superposition conductor & induced charge	Lorentz force superposition induction
physics quantity & geometry	field & *geometry* potential & *geometry*	flux & *geometry* field & *geometry*
physics quantity	field potential charge density/distribution	flux field
geometry	capacitor	*geometry*
algebraic procedure	integration/*algebraic method*	integration/*algebraic method*
other content-related labels	*other content-related label*	*other content-related label*
not content-related	*not content-related*	*not content-related*
'I don't know'	I don't know	I don't know

The proportions of labels in the different categories are presented in Figure 3.2 and Figure 3.3 for electricity and magnetism problems respectively. Apparently, for both proficient and weak students the proportion of 'strong physics procedures' labels is highest. The main differences between proficient and weak students are that proficient students tend to mention more 'strong physics procedures' and the weak students tend to mention more 'not content-related' labels. The latter effect is mainly accounted for by a single individual in the weak students group, who used 'not content-related' labels only.

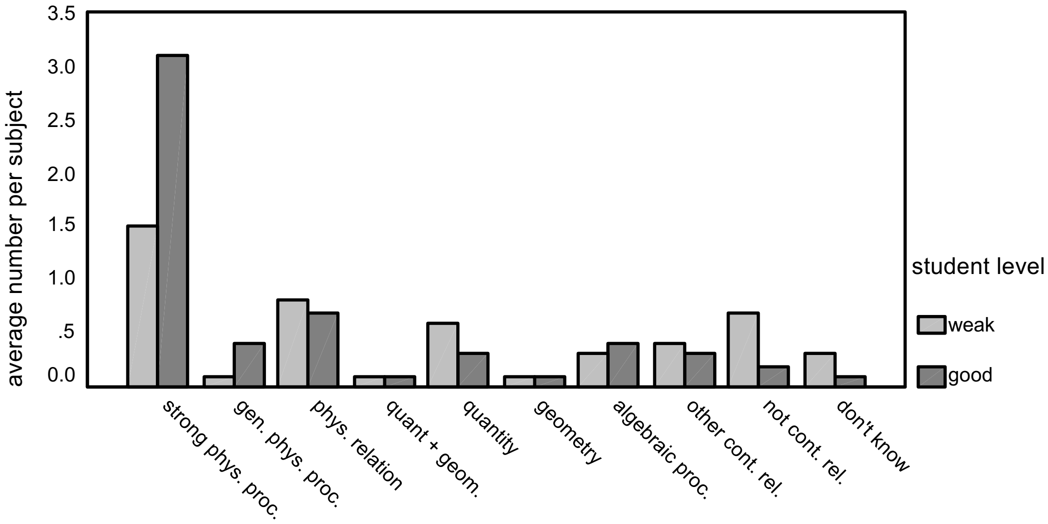


Figure 3.2 Distribution of labels for electricity problems

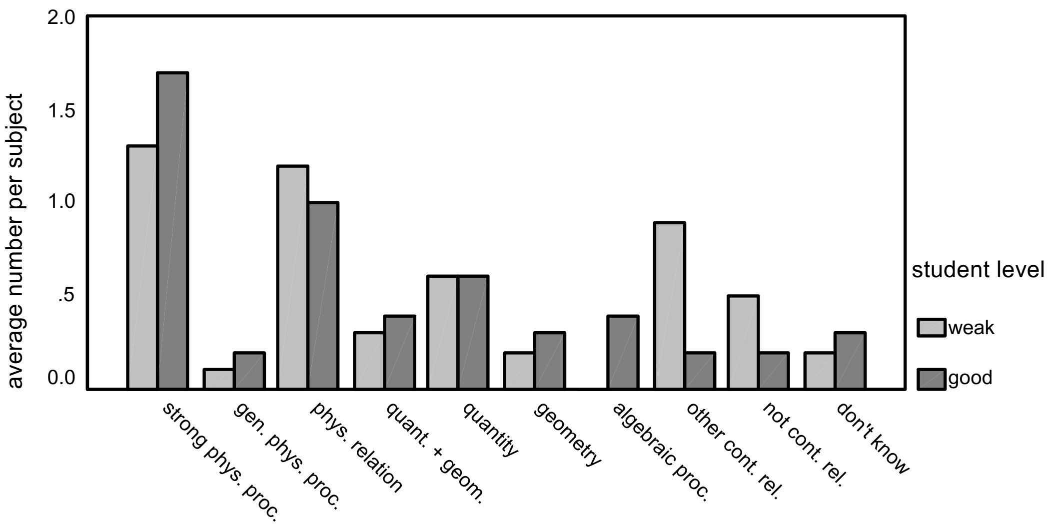


Figure 3.3 Distribution of labels for magnetism problems

3.3.2 Similarity to expert sorting

3.3.2.1 The problem sorting performance-scale and descriptives

Now that we have demonstrated that the cluster solution for the good and weak beginner groups as a whole are similar to each other and to the experts' solution, we want to assess the quality of individual sortings to test whether the elaborations had an effect. We therefore constructed a formula that expresses the similarity between sortings in a single numeral. From the analysis of the expert sortings, it appeared that some pairs of problems occur frequently in expert sortings and that some combinations never occur. A combination was judged to occur frequently among experts if the experimenters and at least two out of the three other experts had made the combination. The similarity measure was based on these two groups of pairs; all other combinations were neglected in computing the expert-likeness. Thus for each subject we had two scores per sorting:

$N_{exp\text{-}like}$: number of combinations the subject made, that are made by – almost – all experts too,
$N_{not\ exp\text{-}like}$: number of combinations the subject made, that no expert makes.

For each set of cards we had two normalisation parameters:

$N_{exp\text{-}always}$: number of combinations that are made by – almost – all experts,
$N_{exp\text{-}never}$: number of combinations that no expert makes.

The resulting 'expert-likeness' score 'E' for subject 'i' was computed from the following formula:

$$E_i = \frac{N_{i,\ exp\text{-}like}}{N_{exp\text{-}always}} - \frac{N_{i,\ not\ exp\text{-}like}}{N_{exp\text{-}never}}$$

The denominators in both fractions are different for the set of electricity problems and the set of magnetism problems (values can be found in Appendix E). The numerators are measures of a particular students' problem sorting. The score according to this formula is a rather well-behaved similarity measure; a 'perfect' sorting would yield the maximum score: '1', all problems thrown together in a single pile would give '0' as an outcome, which would also be the result if each problem is sorted in a separate, one-card, pile.

A remarkable difference between the electricity and the magnetism problem sets is that, although both sets have a comparable – large – number of not-expert-like combinations, the numbers of problems that are put together by all experts is rather different, being 32 for electricity and only 18 for magnetism problems. As a consequence the score for magnetism problems is less robust than the score for electricity problems.

To test the expert-likeness score E's sensitivity to random variations, and to compare the performance of the subjects the average score for a 'blind' sorting, we generated a set of 1000 random sortings. In these sortings the number of piles per sorting was distributed binominally with the average number of clusters set to 5.9, which closely corresponds to the average number of piles in the real data (Table 3.4). Both the scores for real novices and the artificially generated scores are described in Table 3.7. From these data we can conclude that, though the scores for magnetism problems indeed show a greater variance than the scores for the electricity problems do, the average random scores are clearly small compared to the real scores for both the electricity and the magnetism cards.

Table 3.7 Means and standard deviations of the expert-likeness score E per experimental group for real data and for artificially generated random data.

		electricity		magnetism	
		without elaboration	with elaboration	without elaboration	with elaboration
weak students	*M (SD)*	.31 (.14)	.23 (.14)	.31 (.20)	.48 (.21)
	n	16	19	19	16
proficient students	*M (SD)*	.36 (.22)	.58 (.21)	.46 (.25)	.57 (.25)
	n	19	17	17	19
random data	*M (SD)*	.0012 (.08)		.0019 (.10)	
	N	1000		1000	

3.3.2.2 Effect of student level and elaboration

Problem-sorting performance was supposed to depend on the level of the student, on whether or not the problem was presented in an elaborated format, and on the interaction between both factors. In order to visualise the effects of elaborations, we needed to compare scores for electricity problems to those for magnetism problems. Therefore, the scores on all electricity problems were converted to z-scores and the same was done for magnetism problems. As a consequence the systematic difference in scores between magnetism and electricity problem sortings (compare, Table 3.7) was eliminated. The data are presented in the following graph:

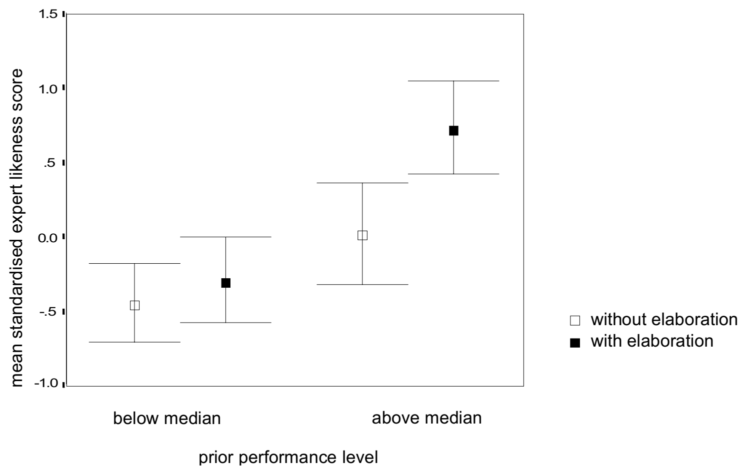


Figure 3.4 Expert-likeness score by student level. The values along the vertical axis are standardised scores. The vertical bars indicate the 95% confidence interval for the mean.

In the experiment we had two blocking factors, namely the experimental task version and the order of the tasks. The main effects of these factors are between subjects. The factor we are interested in (with/without elaboration) is varied within subjects. As we expected no interactions between the blocking factors and the experimental factor, the blocking factors did not need to be included as factors in the analysis.

A repeated measures analysis confirmed that there was a main effect for student level, $F(1, 69) = 19.0$, $p < .001$, and also a main effect for the presence of elaborations in the material, $F(1, 69) = 12.0$, $p = .001$. Moreover, the interaction between student level and the presence of elaboration was confirmed, $F(1, 69) = 5.1$, $p = .028$. A further analysis within the proficient and less-proficient subgroups shows that for proficient students there is an effect of the presence of elaborations, $F(1, 69) = 16.5$, $p < .001$, whereas for the less proficient students we could not demonstrate any effect at all, $F(1, 69) = 0.72$, $p = .398$.

The effect sizes, in terms of explained variance, were moderate: main effect of student level: $\eta^2 = .216$, main effect of the presence of elaborations: $\eta^2 = .148$, interaction between student level and the presence of elaborations: $\eta^2 = .068$. Although these effect-sizes are too small for the intervention to be directly used in instruction, they are large enough to have implications for psychological theory.

3.4 Discussion and conclusion

Our hypothesis was that all our subjects would have some knowledge of problem types and their solutions, and that their major problem was to choose the proper problem type for a given situation. In line with this hypothesis, we found that the result of the cluster analysis for the beginners groups basically resembles the expert sorting, and that beginners often give their piles proper names as well. We found no evidence of a systematically deviant 'beginner sorting', but rather random failures to see the proper problem type. Thus the conclusion by Chi et al. (1981) that novices sort according to superficial criteria and name their categories accordingly, should be qualified to apply in particular conditions only. As noted earlier by Taconis (1995) two factors may explain for the different results: the instructional format and the composition of the problem set. Most notably, sorting the problem set used by Chi et al. requires students to have generalised their understanding of problem types beyond what is needed for problems commonly encountered in the standard curriculum, which would be a remarkable feat.

We had predicted that the availability of simple elaborations would be more helpful to weak beginners than they would be to good beginners, but we found the opposite. As we have pointed out, the elaborations that we provided would be helpful in the process of analysing, restructuring and solving the problem. Our tentative conclusion is that the weak students had no model of the situation at hand that could serve to integrate the given elaborations into a coherent whole. A text base or discourse representation would be of little help in doing so, whereas a situation model would. This is because the given elaborations were redundant from a physics point of view, but not from a discourse point of view, as they appealed to external physics knowledge. In the present experiment it was taken for granted that all subjects mastered the knowledge underlying the elaborations. This is a reasonable assumption since the elaborations presented were very simple. To rule out the possibility that the weak students simply missed the conceptual physics knowledge that would be necessary to comprehend the elaborations presented, this assumption could be subject to direct measurement in a further experiment.

There are two possible mechanisms that would explain how proficient students do profit from the elaborations we gave: either they had already inferred a corresponding elaboration themselves, or they had not. In the first case, the supportive effect of giving elaborations should be explained in terms of focusing attention to, or increasing activation of a particular piece of information. In the latter case, the elaboration really added something to the problem representation. The two mechanisms could be distinguished experimentally by first presenting a 'minimum' problem description, and then presenting an elaboration. The time a subject uses to interpret an elaboration he already inferred would be shorter than the time required to interpret an elaboration that was not inferred yet (Noordman et al., 1992).

Because the type of elaborations that we gave could be only useful insofar as they contributed to inducing a situation model, we interpret our results to imply that proficient students are inducing a situation model from their text-base representation, whereas weak students are unable to induce such a model. This conclusion is in line with De Jong and Ferguson-Hessler (1991), who made subjects reconstruct problem statements that were presented only briefly. They found no difference between proficient and weak students when it came to reconstruction of information in the same format that it was presented in, whereas if the reconstruction was to be in a different mode (verbal vs. visual), results of proficient students proved to be better. Since the latter task would require a mental model of the situation, whereas the first task could be done relying on a text-base representation, they concluded that weak students had problems constructing a mental model of the situation.

From the present findings, we conjecture that weak students, being unable to induce a situation model from text, may be better off if, instead of having to induce a situation model from scratch, they are provided with a representation that matches the inherent structure of the situation model much closer, namely a situation representation in the figural mode. If a weak student has a proper type of drawing available in addition to the textual description of the problem, he may infer relations between the two deductively, which would be less demanding to his constructive abilities. This hypothesis could be tested in a further experiment that basically would be a replication of the current experiment, apart from providing diagrams of the problem as an elaboration, instead of the textual elaborations that we employed in the current experiment.

This study suggests that, for weak students, the pay-off of a more extensive analysis of the problem statement would be small. This finding sheds new light on the disappointing results and especially the lack of transfer of general problem-solving skills programs (Mansfield, Busse, & Krepelka, 1978; Feuerstein, 1980; Mettes, Pilot, & Roossink 1981; De Jong, 1986). The present study suggests that well-chosen content-related support may be more successful in advancing the reasoning process during problem-solving. To conclude, we propose that verbal reasoning deserves further attention as a potentially important phase in the construction of a situation model in – physics – problem-solving.

Part B Supporting problem representations in learning

Still I didn't want to leave any student completely behind, as perhaps I did. I think one way we could help the students more would be by putting more hard work in developing a set of problems which would elucidate some of the ideas in the lectures. Problems give a good opportunity to fill out the material of the lectures and make more realistic, more complete, and more settled in the minds that have been exposed.

Richard Feynman, preface to the Lectures

4. A learning environment for physics problem-solving

In the previous chapters we demonstrated that mental representations and reasoning processes in weak problem solvers are qualitatively different from those in proficient problem solvers. Among the most salient differences we found were the lack of coherence in the weak problem-solvers' problem representations, the sparseness of their problem representations, and their lack of associations between problem features and solution methods compared to proficient problem solvers. We argued that these features of weak problem solvers' problem representations form a serious impediment to successful problem-solving. Therefore, here we present an instructional format we devised to promote the students' understanding of problems, and to support the formation of associations between problem representations and solution information. In the next section we discuss the design of the learning environment. In the following sections we present an experiment we conducted to test the effectiveness of the module, and a discussion of design improvements and design guidelines that were suggested by the outcomes of the experiment.

4.1 Design of the learning environment

We begin this section by reviewing the relevant learning processes. Next, we describe the course as it has been taught for many years, with its goals, its approach and its shortcomings. Then, we discuss the possible improvements to the course, and we review the merits of available tools. Furthermore, we explain our choice for a particular type of tool, and, finally, we discuss the implementation of an adapted course module that is based on this tool.

4.1.1 Learning processes

Learning how to solve a particular class of physics problems fluently, is a complex and time-consuming process. As a primer, a student may listen to a lecture, or interact with a computer simulation to become acquainted with the domain concepts. It is only after this first encounter, however, that the student begins to learn how to solve problems. The continued learning process first requires the learner to combine information from different sources, such as the textbook, previous problem-solving experiences, and mathematics and physics pre-knowledge. Second, it requires the learner to go beyond the literal information presented in order to create understanding, to see implicit regularities, and to learn to routinely apply domain theories. It is common that impasses and misunderstandings arise during this process, and insight often comes only after a period of 'wandering in the dark'. After the initial conceptual barriers have been overcome, it still takes considerable practice to fluently select the right solution step in particular circumstances, recover from errors, and carry out the selected solution steps.

This sequence of stages in the learning process parallels the different phases in a child's development as defined by Piaget (1970). Van Hiele (1957, 1986) was the first to recognise that one passes through similar phases each time one starts learning a new subject. Van Hiele distinguishes between a visual level, a descriptive level, and a theoretical level of understanding, and he continues with more abstract levels beyond these (Van Hiele, 1986, p. 53). According to Van Hiele's theory, the levels follow in a strict order, and specific materials and kinds of problems are associated with each level. Each learning phase, through which a following level of understanding is attained, requires a specific focus and specific activities. To begin with, attainment of the visual level requires extensive playing with the relevant concrete objects. In this stage of learning, the students have no means to formulate goals yet, and consequently they are not yet involved in problem-solving. To attain the second level the student has to practice at expressing the system's properties in words. For the third level, students practice with proof problems. Sometimes students learn tricks to solve problems beyond the level they have attained. According to Van Hiele, this impedes further learning, because a transition from one level to the next comes only after the crisis that occurs when the lower level approach no longer suffices.

Van Hiele's theory is primarily grounded in high school geometry, which might explain the emphasis on playing with concrete objects. In physics at the university level, the role of concrete objects and representations is different from that in geometry. Often the concrete representations in physics are crude approximations meant to illustrate theoretical ideas. Moreover, unlike high-school students, university students already have some background in reasoning with mathematical concepts. Therefore, the strict temporal ordering proposed by Van Hiele cannot be directly transplanted to university physics. Nevertheless, in university physics we can also distinguish the levels described above, and we follow Van Hiele in that understanding at a higher level needs to be based on understanding at the lower levels.

The theory by Van Hiele outlined above clearly illustrates the constructivist view of learning. Constructivist approaches emphasise that knowledge is constructed by students themselves rather than being transmitted from teacher to student. Many constructivist theories, such as Van Hiele's, but also Piaget's, Bruner's and Vygotsky's, take a global developmental and/or cultural perspective. The construction of knowledge can also be analysed at a more detailed cognitive level in terms of information processing, however.

From an information processing point of view, there are two relevant approaches to the learning process described earlier: one is the broad class of production-rule theories; the other is the schema-theoretic approach. First, we look at production-rule systems. We take Anderson's ACT* theory as a well-known instance of the production-rule approach. ACT* theory distinguishes three major phases in learning: acquisition of declarative knowledge, proceduralisation of the knowledge and, finally, tuning of the skills acquired (Anderson, 1983). These steps are worked out in detail in ACT* and, based on this detailed theory, some quite specific tutoring principles have been proposed (Corbett & Anderson, 1992). The most notable recommendations are to emphasise the analysis of goals and sub-goals, to provide instruction in the problem-solving context, to provide immediate feedback about failures, to provide ample opportunity for practice and, finally, to minimise the working memory load.

The other view comes from schema theory, which focuses on the formation of networks and structures (Rumelhart & Norman, 1981). Schema theories stress the organisation of knowledge in clusters, with each cluster centred on a type of problem that requires a particular approach. The learning process in this view is characterised as failure-driven and explanation-based. Although the approach is less clearly associated with a particular person or named theory, the approach also leads to some clear recommendations for instruction. The most relevant ones are to explicate the expert's reasoning process; to present a realistic problem-solving context where the learner has to construct a situation model; and, finally, to put the learner in control of the learning process, which includes the diagnosis of failures (Glaser & Bassok, 1989).

Though both approaches account for the importance of content-related abilities, the schema-theoretical view quite naturally seems to account for impasses, sudden insights, and the importance of situation representations. ACT*, by contrast, depicts learning as a smooth process, resulting in habits. Consequently, the learning processes that occur in complex problem-solving domains are better characterised by a schema-theoretical description than by rule-based approaches such as Anderson's (1983). Therefore, in cases where both theories lead to contradictory advice, we best follow the advice from schema theory.

4.1.2 The standard electrostatics curriculum

The domain of our study is an electrostatics course module, taken from the standard curriculum for first-year physics students. The module is taken as part of a longer course on electrodynamics. Topics covered in this module include charge distributions, symmetries, Coulomb's law, Gauss' law, dipoles, multipoles, conductors, computation of potentials with given boundary conditions, dielectrics and polarisation.

The course has three major components: lectures, work groups and homework. Lectures last two hours, and a typical audience is about one hundred students. In the lectures, the theory is presented and examples of typical problems are worked out. During work groups, small groups and individual students are assigned a set of problems to solve. A tutor is available for every twenty students, who assists if problems arise. Students are expected to solve additional problems at home, and to study the book 'Introduction to Electrodynamics' (Griffiths, 1987). The projected total workload for the course is 80 hours for the average student.

The main aim of the course is to give students a thorough understanding of fundamental concepts and approaches. Here, the groundwork is laid both for more advanced theoretical courses and for application-oriented technical courses. The concepts taught are discussed in a simplified way, and the methods presented are only practical for some idealised problems. However, because later courses build on the material, the student needs to become fluent with the basic concepts, relations between them, applicability of the methods and assumptions underlying them.

It is a well-known problem that students, even though they have learned the methods, fail to see how and when these methods can be applied in new situations. More specifically, in this course, students fail to see how they can use the geometrical properties of a situation to simplify the problem. This might be explained from our earlier work where we demonstrated that novices in this field do not integrate solution information in their mental problem representations (see Chapter 1).

A further problem is that weak students frequently seem to misunderstand problem descriptions, and often fail to make proper drawings of situations (Feiner-Valkier, 1997). This might be explained by the students inability to form a coherent understanding of the problems (see Part A), and by the weak students failing to switch between propositional and pictural representations (compare, De Jong & Ferguson-Hessler, 1991).

4.1.3 Goals and requirements for improved instruction

Based on our view of the learning process and on shortcomings of the current approach, which we have outlined in the previous sections, we assume that students could benefit from a training procedure promoting the construction of problem representations. Our primary goal is to enrich students' mental models of problem situations with elements helping them to construct an integrated model of the situation, and to connect this situation representation to solution information.

The constructivist viewpoint suggests that an effective training procedure is best accomplished by helping students to construct elaborate problem representations themselves, rather than directly providing them with extensive problem descriptions. Both schema theory and production rule theories suggest that good problem representations can only be learned in the context of a real problem-solving activity. Therefore, our approach is to support the formation of problem representations during practice problem-solving. In addition it may be helpful to already start with the integration of theory and situations while studying theory and examples.

Previous studies suggest that proficient and weak students need qualitatively different kinds of support. Strong evidence comes from our own research where we found that proficient students, in contrast to weak students, perform better on a problem sorting task with expanded problem descriptions than with minimal descriptions (see Chapter 3). The difference may indicate that weak students are still wrestling to attain the first Van Hiele level, whereas proficient students are already working towards the second or third level of understanding. This gives rise to the assumption that whereas weak students should primarily be supported in constructing coherent problem representations, proficient students may benefit more from support in constructing flexible problem representations. The different needs of different students can be met either by a system based on an advanced learner model that offers customised instructions, or by an open system that lets students decide what kind of support they need themselves.

Students may be helped in constructing coherent problem representations by stressing the relation between the propositional problem description and the visual representation of the problem, because the visual format reveals relations between propositions (Larkin & Simon, 1987; but see Mayer & Sims, 1994). Flexible coordination of situation representations and solution procedures may be stimulated by making solution procedures recognisable as functional blocks and by demonstrating how modifications in the situation affect the use of procedures.

If students are to be trained in constructing problem representations themselves, they should be allowed to elaborate on their knowledge of the situation freely. So they must be free to modify the problem or pose their own problem. Therefore, the teacher should be reticent with judgements, in order not to frustrate the exploratory process. In other words, the students should be 'owners' of the problem. These requirements suggest a learning environment that allows the learner much freedom. Consequently, such a learning environment will call upon the regulative abilities of the student. Students who lack these abilities may easily loose their way. Moreover, a strong appeal to the student's regulative abilities may interfere with content learning, because of cognitive overload. To reduce these threats, some guidance should be provided and care should be taken to minimise cognitive load extraneous to the task. This can be done in several ways. Firstly, all required information can be integrated into a single structure (Chandler & Sweller, 1991, 1992). Secondly, worked examples can be presented instead of problem-solving tasks (Sweller & Cooper, 1985; Zhu & Simon, 1987; Paas & Van Merriënboer, 1994). Finally, 'goal-free problem-solving' (in other words, computing anything you can for a given situation) also helps to reduce cognitive load (Sweller, 1988). The latter two recommendations are clearly at odds with ACT* and schema theory recommendations and findings, as in both the importance of practice problem-solving and the importance of goal analysis are emphasised (Glaser & Bassok, 1989; Corbett & Anderson, 1992). The differences might be explained by differences in learning tasks and desired types of learning outcomes. Still, worked examples could be used as an introduction to problem-solving tasks.

The requirements we have discussed thus far, and especially the need for visualising solutions, suggest the relevance of a computer-mediated learning tool. This would impose an extra source of cognitive load, namely the interaction with the computer itself. Since students greatly differ in the amount of computer experience they have, this poses more problems for some than for others. Therefore, it is important to offer appropriate user support. A further factor that may affect learning in a computer-mediated environment is computer anxiety. On the one hand, it has been claimed that there is no relation between computer anxiety and learning results when outcomes are corrected for computer experience (Szajna & Mackay, 1995). On the other hand there is clear evidence that anxiety affects information processing (compare, Wilder & Shapiro, 1989; Baron, Inman, Kao, & Logan, 1992). So, it may be worthwhile to attempt reducing computer anxiety, if only to improve the students' motivation.

4.1.4 Instructional tools and interventions

In the previous section we identified two major problems: weak students' lack of internal structure in their problem representations, and weak students' failure to switch between different representations. These problems lead us to search for an instructional tool that would help students to actively construct problem representations themselves, rather than provide them with ready-made situation models. We searched for earlier work on science instruction that we could build upon. Our main interest was to find interventions that could contribute to students constructing adequate problem representations. Because most studies combine several interventions, studies cannot be ordered in straightforward clusters, and there is no conclusive evidence about the effectiveness of most types of intervention. We will briefly discuss the types of intervention found, and some evidence on their effectiveness.

In older work there was an emphasis on acquiring general problem-solving abilities – also: strategic knowledge –, such as analysing, planning and evaluating (compare, De Jong, 1986; Mettes, Pilot, & Roossink, 1981; Reif, Larkin & Brackett, 1976; Van Weeren, De Mul, Peters, Kramer-Pals, & Roossink, 1982). Of these abilities, analysis is most relevant to constructing of a problem representation. However, attempts to improve general problem-solving skills have mostly had disappointing results (compare, De Jong, 1986). Insofar as a gain occurs, it fails to transfer to new tasks and improvements may well be ascribed to a side effect of the training — that is, because problem-solving skill is trained in the context of a domain, students learn what problem features matter in that domain.

Several studies have focused explicitly on teaching how problem features matter in a domain. Taconis (1995) reports on an approach in which students work in small groups to analyse differences between problems and between their solutions. Based on these characteristics students infer how problem features affect problem solutions. Halloun and Hestenes (1987) used 'paradigm problems' to evoke a similar kind of classroom discussion. Afterward, these paradigm problems could also be used as prototypes for further problem-solving. Mestre, Dufresne, Gerace, Hardiman, and Touger (1993) used a computer program to pose qualitative questions about problems, thus forcing the students to focus on problem features. A further way to have students think about problem features is to use problem sorting (Mestre et al., 1993; Bunce, Gabel, & Samuel, 1991) or concept-mapping (Pankratius, 1990) as a tool for instruction. The latter two interventions have nothing inherent to reward the use of relevant problem features over irrelevant ones, however, and therefore, they can only be useful in combination with goal-task-related instruction. Results of the interventions discussed in this paragraph have been diverse, with some being quite successful.

Characteristic of the above approaches is that they attempt to make students identify relevant features, rather than discussing them explicitly. Alternatively, the relation between problem features and problem-solving steps can be discussed explicitly with example problems (Sweller & Cooper, 1985; VanderStoep & Seifert, 1997; Ward & Sweller, 1990; Zhu & Simon, 1987). With such an approach, students do not construct situations representations themselves, however. Reported gains are high but there is little evidence of any transfer beyond the simple goal task that has been learned.

Diagrammatic representation is regarded as a powerful tool for enhancing problem representations, and several of the aforementioned authors have instructed students to construct formal diagrams. Students often resist using diagrams when they are left to their own devices, however, because they have not learnt how to construct and interpret such diagrams (Van Heuvelen, 1991a). In some studies content-related diagramming skills have been explicitly taught (Larkin, 1983; Halloun & Hestenes, 1987; Van Heuvelen, 1991b). Interventions of this type have been quite successful.

To summarise, we found three major types of instructional interventions to enhance problem representations: firstly, interventions aimed at structuring the reasoning process (either by direct instruction or by using appropriate tasks); secondly, interventions to teach the relevant problem features and their implications explicitly; and, finally, interventions to teach more powerful representational formats, such as formal diagrams. Of these approaches, directing student's attention to relevant problem features through appropriate tasks, and teaching powerful representational techniques best match our learning goals. However, we sought to give students more direct feedback on their problem representations than there was in the experiments we have discussed, and to offer them more powerful tools with which to construct their problem representations. Therefore, our focus was on computer-based tools.

In Appendix G, we present a representative sample of software packages and we briefly discuss the role these packages could play as instructional tools. The many computer programs that are made for demonstrating one particular situation – like most of those from the American Institute of Physics – have been excluded from that review. We restrict our discussion to software that can be used for the study of a broader range of physical situations. The programs that we examined fall into four broad classes: visual interface numerical simulation environments, simulation packages with a combined visual and formula interface, numerical packages with formula input and visualisation facilities, and computer algebra packages. We will briefly discuss the four below.

Visual interface numerical simulation environments

In visual interface numerical simulations (such as XYZet, Interactive Physics, and SimQuest), the student can interact with a numerical simulation of a physics phenomenon. The user can construct a situation by clicking and dragging elements, vary quantities by setting sliders, etcetera. These packages are appropriate for constructing situation models. They do not address the propositional representation, including the use of formulas, however. Moreover, they cannot be used in formal problem-solving. In summary these packages may be better suited to use for instruction at the first Van Hiele level, that is, to gain an intuitive understanding of the concepts used.

Simulation packages with combined visual and formula interface

Numerical simulation packages with combined visual and formula interface (such as Modellus and Labview) support the learner in switching between visual and propositional representations. The user specifies a physical system using formulas, and then specifies a visualisation. These packages support the construction of situation representations, as they force the user to specify the situation precisely. Moreover this type of environment is unrestrictive. Yet, it does not support formal problem-solving.

Numerical packages with formula input and visualisation facilities

Numerical problem-solving tools with visualisation facilities (such as MathCad and MatLab), can be used to specify and visualise situations and solutions. They can also be used to practise problem-solving, but the problem-solving methods they can be used for are restricted to numerical methods. An advantage of numerical methods is that they can be used to solve more complex problems. A disadvantage of these methods is that they are farther from the theoretical framework of the domain, and, since the outcomes are expressed as numbers rather than formulas, the outcomes do not provide useful insights.

Computer algebra packages

Computer algebra packages (such as Derive, Maple and Mathematica) are similar to those in the previous group, but they can also be used to solve symbolic equations.

Comparison of the above packages leads to the conclusion that computer algebra packages offer the best functionality for supporting the learning processes we are considering. A computer algebra system (CAS) in itself is no more than a high level programming language for symbolic and numerical computation. In general, a CAS has both imperative and procedural programming facilities. For our intended use, three properties of CASs are of importance: firstly, CASs demand precise specification of problems, in a highly constrained formal specification language; secondly, CASs take over algebraic calculations; and finally, most CASs have visualisation facilities. The required precise specification of the problem and the assistance in algebraic calculations can be used to direct students' attention to the properties of the problem situation: Once a first case has been worked out, situation properties can easily be manipulated, and the solution of a first case can be reused in a following case, provided that the situations match. In addition, the algebraic support may help students to focus on the main line of the solution rather than algebraic calculation details. Finally, visualisation facilities may be used to construct graphical representations of situations and solutions that would otherwise remain abstract. The CAS could thus be regarded as a tool for constructing powerful representations. Alternatively, the CAS's 'programming language' can be seen as a powerful representational system itself, because problems are easily solved once they are properly formulated.

4.1.5 Implementation of the learning environment

We decided to implement our learning environment using a computer algebra system as a problem-solving and visualisation tool. Among the more well-known examples of these systems are Mathematica and Maple. Because the intended participants in the study already had some experience working with Mathematica, we decided to use that program.

Mathematica (version 3.0), like Maple, has an advanced hypertext interface that can present pictures and conventionally formatted mathematical formulae alongside the user-typed input. So, a course module written in Mathematica can be set up as an 'interactive book', with all the necessary information – including theories – presented on-screen. Compared to a regular hypertext, the major differences are that students can modify the content, that the package supports problem-solving, and that the package can be used as a visualisation tool. As discussed in the previous section, these features of CASs could be used to support the learning process. To stimulate the use of computation and visualisation facilities, we set up assignments requiring the use of these features.

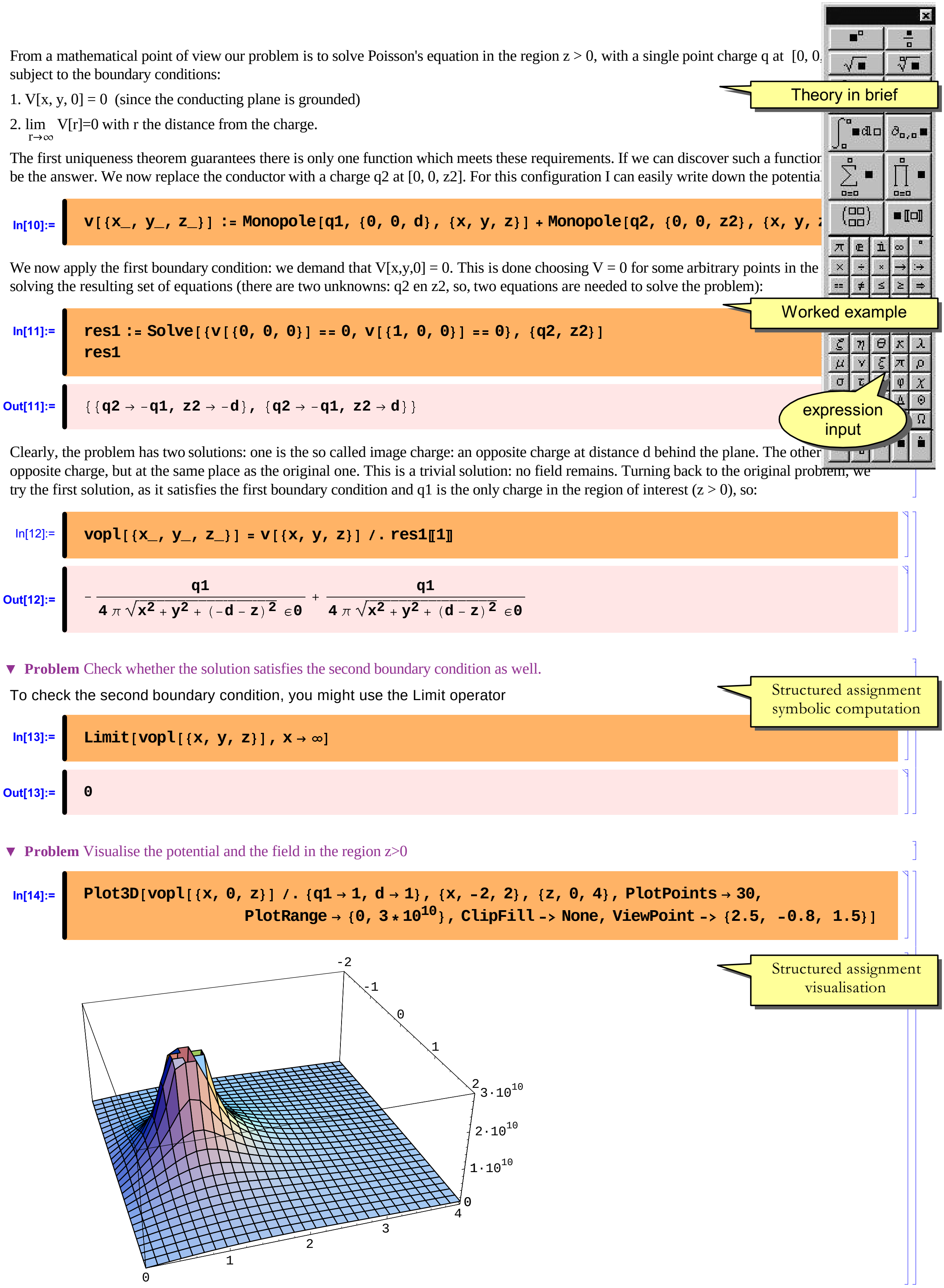
A CAS can be used in many ways. We chose to provide an integrated learning environment that would present the theory in brief, then worked examples, after which various types of assignments would follow. Theory would be presented only briefly, because a book is more convenient for extensive reading. The first assignments on a topic would be highly structured, requiring the learner to modify something in a worked example or to complete an incomplete solution. Later assignments would be more open, requiring the learner to construct the entire solution. This set-up was intended to minimise extraneous cognitive load.

The worked examples help to reduce the effort that goes into mastering the programming language. The structured assignments are intended to focus on variations of situation properties within a class of problems, and on the consequences the properties have for the solution of the problem. Thus, structured assignments may help the students to identify relevant properties of a situation (=elaboration). The visualisation assignments may help to connect a concrete (visual) physical representation to the abstract formalism of both the situation and the solution (=visual elaboration). Finally, practice problems require the problem solver to elaborate on the problem statement (visual and propositional) and, moreover, they provide a training opportunity. The difference from normal practice problems is the support offered by the CAS.

The experimental course was intended to represent 8 to 10 hours of workload on the average student. The general subject of the course was 'special techniques for calculating potentials'. There were four sections:

- general introduction and instruction Mathematica
- introduction E-field and potential
- image charges
- dipole and multipole expansion.

In Figure 4.1 we present a brief example of a topic in the experimental course (a more extensive sample can be found in Appendix H). The example in Figure 4.1 begins with a summary of the relevant theory, followed by a worked example. In this worked example the text in lines labelled 'In' is already present when a student starts working with the material. When the student executes the commands in these lines, by pressing $\boxed{\text{SHIFT}}$-$\boxed{\text{ENTER}}$, the computer generated output labelled 'Out' appears. The student can also modify the input lines and then execute the commands again to examine the effects of a different situation. Below the worked example follow two examples of – parts of – structured assignments. The input shown in these examples (in[13] and in[14]) has to be thought up and typed in by the students themselves. To facilitate entering symbols and expressions, a floating 'palette' was provided in a separate window. The students could pick an element from the palette by clicking on it.



From a mathematical point of view our problem is to solve Poisson's equation in the region z > 0, with a single point charge q at [0, 0, subject to the boundary conditions:

1. V[x, y, 0] = 0 (since the conducting plane is grounded)
2. $\lim_{r\to\infty}$ V[r]=0 with r the distance from the charge.

The first uniqueness theorem guarantees there is only one function which meets these requirements. If we can discover such a functior be the answer. We now replace the conductor with a charge q2 at [0, 0, z2]. For this configuration I can easily write down the potentia

```
In[10]:= v[{x_, y_, z_}] := Monopole[q1, {0, 0, d}, {x, y, z}] + Monopole[q2, {0, 0, z2}, {x, y, z
```

We now apply the first boundary condition: we demand that V[x,y,0] = 0. This is done choosing V = 0 for some arbitrary points in the solving the resulting set of equations (there are two unknowns: q2 en z2, so, two equations are needed to solve the problem):

```
In[11]:= res1 := Solve[{v[{0, 0, 0}] == 0, v[{1, 0, 0}] == 0}, {q2, z2}]
         res1

Out[11]= {{q2 → -q1, z2 → -d}, {q2 → -q1, z2 → d}}
```

Clearly, the problem has two solutions: one is the so called image charge: an opposite charge at distance d behind the plane. The other opposite charge, but at the same place as the original one. This is a trivial solution: no field remains. Turning back to the original problem, we try the first solution, as it satisfies the first boundary condition and q1 is the only charge in the region of interest (z > 0), so:

```
In[12]:= vopl[{x_, y_, z_}] = v[{x, y, z}] /. res1[[1]]
```

Out[12]= $-\frac{q1}{4\pi\sqrt{x^2+y^2+(-d-z)^2}\,\epsilon0}+\frac{q1}{4\pi\sqrt{x^2+y^2+(d-z)^2}\,\epsilon0}$

▼ **Problem** Check whether the solution satisfies the second boundary condition as well.

To check the second boundary condition, you might use the Limit operator

```
In[13]:= Limit[vopl[{x, y, z}], x → ∞]

Out[13]= 0
```

▼ **Problem** Visualise the potential and the field in the region z>0

```
In[14]:= Plot3D[vopl[{x, 0, z}] /. {q1 → 1, d → 1}, {x, -2, 2}, {z, 0, 4}, PlotPoints → 30,
                PlotRange → {0, 3*10^10}, ClipFill -> None, ViewPoint -> {2.5, -0.8, 1.5}]
```


Figure 4.1 Brief example from the section on image charges in the experimental course. (translated).

4.2 Training for enhanced problem representations: a classroom evaluation

The experimental course we had designed was expected to improve students' understanding of physical situations and to strengthen the relation they see between solution methods and situation features. To test these hypotheses, we designed and carried out an experiment to compare learning outcomes of the experimental course with those of the normal way of practising. We compared two conditions: a computer group where the experimental course module was used and a 'traditional' group. In the traditional group the students used paper and pencil to solve problems similar to the ones presented as open assignments in the computer course. In the present section we discuss the method and the outcomes of the experiment.

4.2.1 Method

4.2.1.1 Measures of the learning outcome

As Reif has argued, physics education lacks the precise standards and measurements methods that made much of the progress in physics itself possible. Setting such standards would not only allow the comparative assessment of different teaching approaches, but it could also help to operationalise learning goals and to decide their relative importance (Reif, 1996).[21] For this experiment we constructed a test to assess the specific learning goals we had. In addition to that test, students also took their regular final examination. We will discuss both, beginning with the last.

---

[21] It has been demonstrated in several studies (diSessa , 1993; McCloskey, 1983; Minstrell, 1989) that severe conceptual misunderstandings may persist, even when students do well on their regular final exams. There have been some attempts to construct tests to assess conceptual understanding, such as the Force Concept Inventory (FCI: Hestenes , Wells, & Swackhammer, 1992) and the Mechanics Baseline Test (MBT: Hestenes & Wells, 1992) both for mechanics. Some of these tests have been used successfully to demonstrate differences between instructional approaches, but there is only limited information about the psychometric properties of the tests. When physicists specify learning goals they tend to be quite explicit about the content that is to be learned, but rather vague about the kind of knowledge that is to be acquired. Even in the aforementioned FCI and MBT, the concept 'qualitative understanding' had no explicit operationalisation — although the authors of these tests do distinguish between the two tests; the MBT being 'the next step above the [FCI] in mechanics understanding', because it requires 'formal knowledge about mechanics'. Educational psychologists, in contrast, have developed adapted tests to measure particular types of knowledge such as conceptual knowledge, structural knowledge, or intuitive knowledge (Swaak & De Jong 1996). Such instruments have not been made for university physics knowledge, however, and, moreover, the relation between most of these specific knowledge types and – university – physics learning goals is unclear.

Examination score and average prior-performance

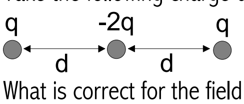
As a first measure of the learning outcome we have the scores on the regular final examination for the course. For this examination students had to do three assignments consisting of 14 sub-problems in total. In a typical assignment, the first sub-problem requires a basic understanding of the situation, the next sub-problems apply to a standard solution procedure, and in the final sub-problem some more creative and subtle manipulations are required. The scores on the sub-problems were used to compute a reliability parameter for the total examination score. These scores were also available for students not participating in the experiment.

To match the control group and the experimental group for prior performance and to correct for differences in average performance level, we collected the students' scores on four previous tests. These tests were: the VWO (=grammar school level) final examination scores for physics and for maths for the sciences, and the university final tests for introductory mechanics and introductory relativity theory. In earlier research we demonstrated that these scores provide a reliable indicator of prior performance (see Section 2.2.2). The prior performance score was taken to be the mean of these four test scores. Missing values were replaced by the mean score for that test. The prior scores were also available for students who did not participate in the experiment, so that we compare the performance of the students in our sample to the overall population performance.

Pre and post-test

Because we were specifically interested in students' understanding of problems and in their choices of solution methods, we developed a pool of test items that were supposed to assess just these abilities. All items were tested by several expert teachers. The expert teachers also judged the quality of the distractor answer alternatives, and they assigned a difficulty-rating to each item. As a final result we had a set of multiple-choice items of two different types: elaboration questions, and solution type questions. In the first type, a property of the situation was asked for. In the second type, the appropriate solution method was asked for. Examples of both types are given in Table 4.1.

Table 4.1 Examples of the two types of items that were used in the pre-test and post-test.

elaboration item	solution-type item
Take the following charge distribution  What is correct for the field outside the charge distribution? ♦ the field equals zero nowhere ♦ the field equals zero in exactly one point ♦ the field equals zero in exactly two points ♦ the field equals zero on an entire surface in space	The electron charge distribution of a hydrogen atom can be described by the following formula $\rho(r) = c.e^{-r/a}$. If you want to describe the electric field caused by this charge distribution, your best option is: ♦ to use Coulomb's law ♦ to integrate over the spatial charge contributions explicitly ♦ a multipole approximation ♦ to use Gauss' law

From the pool of test items we constructed a pre-test and a post-test. The item types, the subjects addressed and the difficulty of the items (as judged by the expert teachers) were adjusted to have equivalent tests. After final corrections, a pre-test of 21 items and a post-test of 24 items resulted.

4.2.1.2 Evaluation and observation

Apart from direct performance outcomes, we were interested in the students' opinions about the experimental course, and in the learning process that took place in the experimental group. To assess the students' opinions we constructed a questionnaire. Because of limited resources, our only way to keep track of the learning process in the experimental group was by informal notes kept by the experimenter who taught the experimental course.

We made two versions of the questionnaire, one for the experimental group and one for the control group. Both versions of the questionnaire addressed the following major issues:

- number of assignments completed
- attractiveness of the instruction (3 items)
- difficulty of the subject (2 items)
- navigation and control (4 items)
- help and collaboration (5 items).

In the questionnaire for the experimental group, there were additional items to assess the use and appreciation of features of Mathematica, such as graphing and computing integrals, and instructional elements, such as worked examples and practice assignments. Except for the number-of-assignments-completed items, all items had to be answered on five-point Likert scales. Statistical analysis and further interpretation were done at the level of single items

4.2.1.3 Procedure

Participants first had 30 minutes for the pre-test. Then, spread over three days they had 7 hours available for the learning task, which is somewhat shorter than the projected time required to complete the experimental course module, so that even the quicker students had to use all time available. Students in both groups worked under the supervision of a tutor, who was available to answer any questions. Students in both groups were free to co-operate or to work individually. The third, and final, session ended with a post-test of 30 minutes.

4.2.1.4 Subjects

The experiment was conducted at the Faculty of Physics at Utrecht University. There are approximately 90 first-year physics students. All these students were contacted individually and invited to participate. In total 42 students agreed to participate. These students were assigned to the two experimental groups by the experimenters. On the basis of available prior performance scores, two groups were formed of equal ability. Because of the limited number of computers available, the computer group had to be split up into three equally sized sub-groups taking the course at different times. To maintain equal group sizes, the control group was also split in three sub-groups. In the first session 33 of the 42 students attended (17 in the computer group and 16 in the traditional group). During the two subsequent sessions, all remaining 33 students kept attending. Students in both groups were paid *ƒ*50 (US$ 25 approximately) after they had completed all three sessions.

4.2.2 Results

4.2.2.1 Learning outcome

We had chosen two measures for learning outcome: the examination result and the adapted test, which was supposed to assess more specifically the quality of the student's situation representations and the linking between situation representation and solution approach. From now on we will refer to the tests as regular test and adapted test respectively. Next, we will discuss the outcomes of both measures.

In total 56 subjects took the regular test. Except for one student from the computer group, students who had participated in the experiment also took the regular test. To assess the reliability of the test score, we computed Cronbach α, taking the 14 sub-problems in the test as items. With all 56 subjects included we found a reliability of $\alpha = .90$. Descriptives of the scores for groups of students are given in Table 4.2.

Table 4.2 Descriptives of the prior performance scale, the regular test scores, adapted pre-test and the adapted post-test.

		regular tests		adapted tests	
condition	*n*	prior results[a]	electrostatics[a]	pre-test[b]	post-test[b]
computer	17				
M (SD)		7.8 (0.8)	6.1 (2.2)[c]	0.53 (0.12)	0.49 (0.15)
traditional	16				
M (SD)		7.4 (1.0)	5.8 (2.0)	0.52 (0.15)	0.52 (0.14)
non-participating	24[d]				
M (SD)		6.5 (1.6)	4.1 (2.8)		

[a] scores represent grades according to Dutch school system: 1 is very poor, 10 is excellent.
[b] scores represent proportion correct answers.
[c] $n = 16$ because one of the students did not take the final test.
[d] only students who did took the regular final test were included.

When we test the significance of the between-group difference, we have to include the prior performance level as a covariate, to correct for prior differences between the groups. The four-item prior-performance scale is a quite reliable estimator of the general prior performance level, $\alpha = .87$, $n = 57$. After the inclusion of this covariate in the analysis, no significant effect of the experimental treatment remains, $F(1,29) = 0.26$, $p = .77$.

We also compared the results of participants in the experiment to those of the students who had not participated. Participants clearly differ from the non-participating students: both on prior performance and on the regular test their scores were significantly better ($F(1,55) = 8.17$, $p = .006$ and $F(1,54) = 10.81$, $p = .002$ respectively), which indicates that mainly proficient students volunteered. Although the participants in the experiment were not drawn randomly from the population, we can get some idea of the main effect of attending an extra 7 hours of instruction by comparing the final test scores between participants, and non-participants. The prior performance level has to be included as a covariate again, to compensate for the general difference in achievement level. Surprisingly, the gain for the students who took the extra instruction, relative to the results of non-participating students, is not significant, $F(1,53) = 0.50$, $p = .48$.

Apart from the main effect of the treatment a second, and equally important, issue is whether the treatment reduces or increases the difference between weak and proficient students. Here, again, we found no significant interaction between experimental condition and general performance level, $F(1,28) = 0.70$, $p = .41$. The trend in the data, however, suggests that the experimental instruction is more favourable to weak students than to proficient students (Figure 4.2).

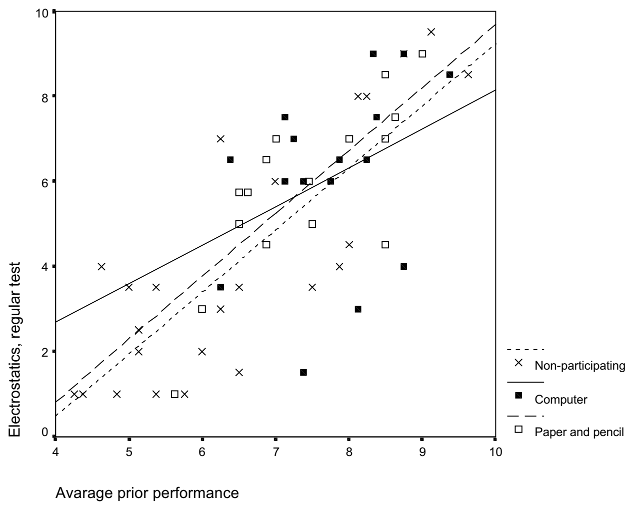


Figure 4.2 The scores on the regular final test, for both groups in the experiment and for the non-participating students, plotted as a function of their average prior-performance levels.

We now turn to the scores on the adapted test. Both a pre-test and a post-test were administered. The test reliability was estimated using Cronbach's α. For the pre-test, we found a quite low reliability, $\alpha = .42$, $n = 21$, and for the post-test we also found a poor reliability, $\alpha = .55$, $n = 24$. These values suggest that the items were not parallel tests of the same ability. The total number of participants was too low, however, to permit any proper further analysis of the tests such as item response models or multidimensional scaling techniques, which otherwise could have been used to find out whether all items respond to the same ability. Although the pre and post-test were intended to measure the same ability, the scores on both tests could not be compared directly, as we had no empirical data on the relative difficulty of both tests. Therefore, we compared the scores of both groups on the post-test, and we included the pre-test score as a covariate. We tested both the main effect of experimental condition and the interaction between the experimental condition and the covariate, but neither gave a significant result ($F(1,29) = 1.92$, $p = .18$ for the main effect and $F(1,29) = 1.51$, $p = .23$ for the interaction). Here, the main effect displays a trend in favour of the traditional group, and the trend in the interaction suggests that proficient students may be better off with the computer learning environment.

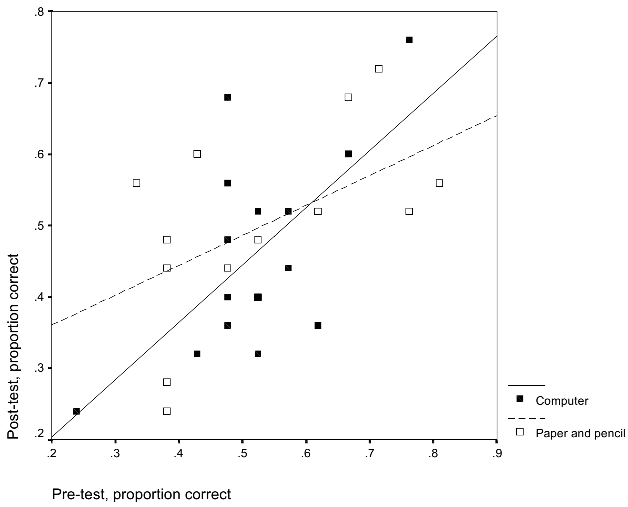


Figure 4.3 The scores on the adapted post-test, for both groups in the experiment, as a function of the scores on the pre-test.

4.2.2.2 Students' evaluation

To gain insight into the mechanisms leading to the learning outcomes, we also collected some evaluative information regarding the process. The evaluation consisted of three parts: a set of questions posed to students in both groups, a set of questions specifically addressing features of the computer algebra environment and, finally, remarks collected from the evaluation forms.

First, we shall discuss the items that can be compared across groups. The first issue addressed was the number of assignments completed: students in both groups had to solve the same sets of assignments. The question was about how many of these assignments the student had completed. The worked examples and completion exercises that students in the computer group had to study prior to starting with open assignments were not included in the comparison. The average number of assignments completed was significantly lower for the students in the computer group, compared to students in the traditional group, $F(1,31) = 29.1$, $p < .001$ ($M = 3.6$, $SD = 2.1$, and $M = 7.0$, $SD = 1.5$, respectively). This indicates that students in the experimental group spent much of their time on worked examples and completion exercises.

We found no evidence that students found one condition more attractive, $F(1,31) = 0.20$, $p = .89$. In the further items that could be compared across both groups we found a number of differences between both groups. The students in the Mathematica group judged the physics content in the experimental instruction to be simpler than the students in the control group did, $F(1,31) = 4.48$, $p = .042$. Because the difference was not reflected in the students' judgements about the difficulty of the general course, we must conclude that the difference is in the content of the course module we used in the experiment. This is easily understood because students who worked with the experimental course spent so much of their time on the introductory assignments. Given this difference, it is not surprising that the students in the control group score higher on the question whether they had learned much about physics content.

We had also posed some questions regarding regulation, navigation and confusion. Students in the computer group clearly had more trouble finding their way than students in the traditional group had, as is demonstrated by the scores displayed in Table 4.3, where the computer group had higher scores on all four items.

Table 4.3 Evaluation items related to navigation and confusion (the wording was slightly different for both groups: adaptations for the traditional group are italicised).

While I was working with the system *(solving the exercises)*, I often forgot what I did before, $F(1,31) = 3.39$, $p = .075$.
The system *(I)* usually could solve the problem, once I had entered *(formulated)* it correctly, $F(1,31) = 7.31$, $p = .011$.
While I was working with the system *(solving the exercises)*, I often lost an overview of things that appeared on the screen *(I had already written)*, $F(1,31) = 10.343$, $p = .003$.
While I was working with the system *(solving the exercises)*, I often had to look up things in earlier work, $F(1,31) = 9.40$, $p = .004$.

We found marginally significant differences in the amount of help required and the type of help required. The computer students required slightly more help, $F(1,31) = 3.28$, $p = .08$, and the help they needed was less focused on physics understanding, $F(1,31) = 3.07$, $p = .09$. Scores for the students in the Mathematica group, indicated that they needed more help on Mathematica, than they did on conceptual physics problems, $F(1,16) = 5.54$, $p = .03$.

The next part of the evaluation concerns specific features of the Mathematica learning environment. Several features were assessed on each of the following aspects: clarity, attractiveness, difficulty and instructiveness. Two of the items addressed inherent features of Mathematica itself, namely visualisation, and computation of gradients. For these features we also asked how frequently they had been used. Three other items addressed the elements that we had built into the learning environment: worked examples, completion tasks and open assignments. The results of the evaluation are summarised in Table 4.4 and Table 4.5.

Table 4.4 Evaluation outcomes for two features of Mathematica (n = 17). Scores are on five-point Likert scales.

		amount of use	clarity	attractiveness	difficulty	instructiveness
visualisation	*M (SD)*	4.06 (0.97)	4.53 (0.62)[a]	4.06 (0.97)	2.71 (1.05)	3.94 (0.75)
gradient	*M (SD)*	3.71 (1.21)	3.87 (0.96)[a]	3.65 (0.79)	2.29 (1.05)	2.41 (1.12)

[a] n = 16 because of one missing value.

Table 4.5 Evaluation outcomes for three elements of the learning environment (n = 17). Scores are on five-point Likert scales.

		clarity	attractiveness	difficulty	instructiveness
worked examples	*M (SD)*	4.53 (0.80)	3.88 (0.70)	2.06 (0.56)	3.53 (1.33)
completion assignments	*M (SD)*	3.88 (0.86)	3.59 (0.51)	3.12 (1.11)	3.53 (1.12)
open assignments	*M (SD)*	3.12 (1.11)	3.53 (0.72)	3.88 (0.78)	3.82 (0.72)

It immediately becomes clear that, overall, visualisation is judged more favourably than gradients. We tested the significance of the differences in individual aspects by using a repeated-measures ANOVA. We found a marginally significant difference for clarity, $F(1,14) = 3.65$, $p = .076$, and a clear difference in instructiveness, $F(1,15) = 26.3$, $p < 0.001$. For the comparison of worked examples, completion assignments and open assignments, we found significant differences in clarity, $F(2,32) = 14.3$, $p < .001$, and difficulty, $F(2,32) = 27.1$, $p < .001$. All pair-wise comparisons gave significant results too, so we conclude that worked examples were clearest and simplest, and that the open assignments were the least clear and the most difficult.

As the final part of the evaluation, we summarise remarks collected from the evaluation forms. We will focus here on the remarks collected in the Mathematica group. Some students expressed their general feeling about the course:

> *Working with Mathematica is a good supplement to problem-solving on paper* (ID8).
>
> *After all it was fairly instructive, [….] A problem is, however, that while sitting behind the computer I can't concentrate on physics problems too well* (ID15).
>
> *It was fun to participate, but I did not find it very instructive … The paper-and-pencil work groups are far!! more boring* (ID18).

Two students commented that the learning environment evoked a passive attitude:

> *Solving the exercises tends to come down to using the "copy" and 'paste" options of Mathematica. This did not contribute to the understanding of what really happened* (ID6).
>
> *[…] because of these worked examples you knew exactly what you had to do, so little initiative was required, whereas initiative should be important (you must be able to do it yourself, not to copy)* (ID31).

There were some positive remarks specifically about the visualisation facilities:

> *The benefit of the method is that the pictures give a good insight into what's going on. This may be helpful when you later come across a similar exercise. Pictures are easier to remember than formulas are* (ID7).

> *Working with Mathematica is a good supplement to problem-solving on paper. The pictures give an insight into what you are working on* (ID8).

> *After all it was fairly instructive, especially the visualisation…* (ID15).

Finally, some students commented on problems they had with Mathematica:

> *Problems with Mathematica syntax cause a loss of time, especially during the first session* (ID15).

> *It is easy to lose your way, exercises were not hard but took a lot of time, because of irritating Mathematica* (ID30).

> *It was instructive with regard to Mathematica* (ID32).

> *The difficulty was more in Mathematica than in physics. The longer I used Mathematica, the faster I worked, and the more I could concentrate on physics problems* (ID33).

4.2.3 Discussion

The aim of this study was to assess the effectiveness of two forms of tutoring: the traditional one and a newly designed one. We had predicted a benefit for students taking the experimental course, especially on understanding of problem situations and choice of solution methods. We found no significant difference between the approaches. Moreover, among the more surprising outcomes of the experiment was that when we corrected for prior-performance levels, we could not demonstrate a benefit of participating in the experiment relative to the group of students who did not participate. A plausible interpretation of this finding is that participants compensated for the extra effort they made in class by spending less time on their homework. However, as the controlled experiment we did, was restricted to comparing the newly designed instruction with the traditional approach, we will restrict our further discussion to this comparison.

As a first point the quality and appropriateness of our instruments for assessing learning outcomes deserve attention. The reliability of the newly designed test proved to be dissatisfactory. As a consequence, we could not detect small differences in learning results. The regular test was quite reliable, but it had its focus on different abilities than the ones we were looking for. Still, even with these limited tests, we probably would have detected any major differences between both groups. So we conclude that the effect of the experimental treatment has been limited.

An important factor explaining the lack of a gain for the newly designed instruction might be the high cognitive load in the computer course. We found strong evidence that the extraneous cognitive load in the computer condition is quite high. Both the closed evaluation items and the remarks made by the students suggest that students in the computer group were distracted from the physics content. Apart from the disruptive elements in the learning environment itself, there were external distractions, such as the availability of Internet browsers, that were absent in the traditional group.

When we examine the outcome of the evaluation in detail, we see that the students' prior attitudes and expectations might have worked against the computer learning group too. This approach to problem-solving was quite new for the students, and, also, the students were not yet fluent programmers. This is supported by informal observations made in the computer group. A first observation is that, generally, students initially disliked Mathematica. They had worked with a previous version of Mathematica (version 2.2.3) in a programming course, and most students disliked it. Moreover, several students expressed the opinion that physics problem-solving is best done on paper, and that you cannot learn physics via a computer. This belief might be reinforced by the students' experience with examinations, since in most examinations it is a requirement that the student can write out solutions fluently.

Although the learning environment we used provided more structure than the tasks the students had worked on in their programming course, students ran into many frustrating errors caused by typing mistakes and by more fundamental misunderstandings of programming principles. In addition to the students' mistakes, the version of Mathematica that we used (version 3.0.0.0) had some irritating bugs, and the program often had to be rebooted. As most students were quite new to programming, they were not very systematic in tracing the errors they made. Many even failed to recognise instances in which they had used non-matching parentheses in an expression. The problems were aggravated by the cryptic error messages Mathematica produces in response to syntax errors. Given this, the evaluation findings regarding the confusion in the computer group are not surprising.

The evaluation of the different components of the Mathematica learning environment indicates that the students found the visualisation facilities rather instructive. This impression is confirmed by some students' remarks on the evaluation form. Moreover, it is in line with observations by the experimenter, who had several discussions about common misconceptions that were triggered by visualisation outcomes. As an example, consider the following: one of the assignments was to draw a field plot for the field of a physical dipole, and compare this to the field of a mathematical dipole. One of the steps in the exercise was to zoom in on the physical dipole. Several students failed to understand why the plot essentially remained the same, and where they had to look for the changes. This provided a good starting point for tutoring discussion. Although students gave high ratings for clarity of the visualisations, the quality of the plots could still be improved (for suggestions see Appendix I). Moreover, even though students considered the visualisations only moderately difficult, it appeared that students spent too much time on figuring out details of graphics commands.

The symbolic computation features were valued less positively, as indicated by the evaluation of the compute gradient facilities. Although students judged these rather straightforward to use, they did not find it very instructive to compute gradients. A major cause could be that students fail to examine the computer-generated expressions critically, so that application of the symbolic computation facilities remains a trick. This suggests that several students had not yet fully attained the formal reasoning level. The suggestion is strengthened by misconceptions we encountered, such as the one about the dipole field, mentioned earlier, and another student failing to understand why the method of image charges can only be used with virtual image-charges (that is, with the position of the image charge outside the region of interest).

The three major instructive elements in the experimental learning environment, namely worked examples, completion assignments and open assignments, were judged to be about equally instructive. As expected, the three elements were found to become progressively more difficult. As indicated both by remarks on the evaluation form and by the experimenter's observations, too many of the assignments could be solved by mindless copying of the worked examples. This may be because students spent most of their time on the introductory, structured, assignments. In any case, the opportunity to solve problems by just copying entire solutions must have worsened the students' attentiveness to computer-generated formulas.

To summarise, we found no significant differences between learning outcomes with the new approach and with the traditional approach. The students' remarks and our own observations indicate that the new approach is potentially useful, however. This is especially true for the visualisation facilities. We found some problems with the instructional material. The most serious difficulties were related to navigation and to high extraneous cognitive load. Moreover, the quality of the testing materials, and the correspondence between testing materials and course content deserve further attention as well.

4.3 Conclusions and recommendations

We have reported on the development and evaluation of a learning environment for a first-year university course on electrostatics. Our goal was to support students who are moving towards a theoretical level of understanding in gaining an intuitive understanding of situations, solution methods and the relations between them. We formulated the demands for such a learning environment, and then we built a learning environment based on available software. Among the distinctive features of the learning environment are a precise language for specifying problems, visualisation support and symbolic computation support.

We tested a first version of the experimental course on a sample of first-year physics students. Results indicate that this approach to learning and problem-solving is quite new to the students and that it takes them a considerable amount of time to get used to the new approach. Moreover, students did not see how the abilities taught in the computer course were relevant to the final test, where they would have to solve the problems by hand anyway. If the students are to be won over to the approach, some items in the final test should be clearly related to the new approach. We found that students in our experiment were more positive about the facilities for visualising solutions and situations than they were about the symbolic computation support. We interpret this to indicate that the students in the experiment have not fully attained the formal reasoning level. This is worrying because the students who participated were among the more proficient students in their cohort.

Although we could not demonstrate a significant learning gain over the old approach, we found that the computer course was successful in addressing misconceptions, in clarifying concepts that underlie solution methods, and in supporting the construction of situation models. As we have no observation data from the control group, we cannot compare this result between groups. In all these learning episodes in the experimental group, discussions with the tutor played a central role. Therefore, we conclude that the role of the tutor in the experimental approach is as important as it is in the usual problem-solving workgroups.

On theoretical grounds we had chosen to give the learners full control of the learning environment and of their learning processes. This places a heavy cognitive load on the students, and so any additional extraneous cognitive load might harm learning. Regretfully, the experimental course contained several disruptive elements. To further explore the potential benefits of the new approach, it is necessary to improve the quality of the software and to refine the educational design of the course. The most urgent problems in the software are the following: the instability of the computer algebra software; the incomprehensible error messages, even about frequent errors such as mismatched parentheses; and the excessive programming effort for the student. The first problem may be solved in a new version or, alternatively, one may consider switching to an equivalent package such as Maple. The other problems we have to solve ourselves, as will other educators implementing similar courses. Improvements to the educational set-up of the course module should be aimed at increasing the students' reasoning about the physics background, while they are studying worked examples and while they are solving completion assignments. In the current version, students could too easily solve the latter assignments by copying the examples (Appendix I suggests improvements for the present course).

We believe that, with these improvements, the revised course may provide a valuable supplement to practising electrostatics problem-solving by hand. Likewise, in other physics domains, such as mechanics, similar courses may help the students to gain an intuitive understanding of the abstract situations and methods with which they are working.

5. Epilogue

In this dissertation we addressed how to improve physics problem representations in learners. The superiority of proficient problem solvers' problem representations over those of weak problem solvers has been demonstrated by several authors (Chi et al., 1981; Chi & Bassok, 1989; De Jong & Ferguson-Hessler, 1991; Larkin, 1983). It is not always that clear, however, how good and poor problem representations differ. We began our research identifying competence-related differences in problem representations. Next, we explored competence-related differences in the process of constructing a problem representation. Based on the outcomes of these studies, we designed a learning environment specifically aimed at improving problem representations. We will now briefly review our work, discuss its scope and limitations, and suggest directions for future work.

In the first study we explored how experts' and good and weak novices' mental representations differ from each other. We assumed that the concept of depth comprises features at the structure level, proposition level and concept level, and that a detailed analysis of these features could contribute to an understanding of problem-solving performance. We devised a method to explore subjects' problem representations, which we used to compare physics problem representations of experts and of good and weak novices. Subjects were asked to respond to physics formulas by describing relevant problem situations. We analysed these problem descriptions at the levels of words, sentences and complete descriptions. Results indicate that expertise affects the structure of representations rather than the presence of particular concepts, and that the flexible use of multiple representations is more important than the presence of one specific kind of representation.

Based on our findings, we hypothesised that constructing a structured and flexible problem representation requires the problem solver to elaborate on the initial problem description to infer relations and thus connect pieces of information. We also hypothesised that, whereas proficient problem solvers elaborate fluently on a given initial problem description, weak problem solvers fail to make proper elaborations automatically, though they have declarative knowledge of the underlying relations. We experimentally tested the effect of providing beginners with elaborations they failed to infer. Our main interest was whether or not the additional information would support deep processing, and thus the establishment of a problem's 'physics' representation. We used a card-sorting experiment in which we had two versions of physics problem descriptions to be sorted: an elaborated and a 'minimum' description. The results for proficient and weak students were compared. We found that the elaborations we gave supported integrative reasoning in proficient students only. Our findings suggest that reasoning processes in weak students may be qualitatively different from those in proficient students, and that a major problem in weak beginners is their failure to elaborate rather than a lack of knowledge regarding problem types.

From these two studies some directions for further research can be inferred. One outcome of the first study was that novices mention markedly more problem-solving goals than experts do. This finding deserves further attention, particularly because researchers hold opposite opinions of how goals should be used in problem-solving instruction (compare, Corbett and Anderson (1992) who advise emphasising goals from the beginning vs. Sweller (1988) who promotes goal free problem-solving as a means to reduce cognitive load). The method we used in our second study could clearly be used with different types of 'elaborations'. If we could identify a type information that helps weak students more than it helps proficient students, this would give insight into what reasoning lacks in weak students, as well as provide a starting point for instruction. The formal diagram could be a promising type of information to begin with.

After we had detailed our insight into the limitations of (weak) students mental representations in two studies, we designed an instructional approach specifically aimed at promoting students' mental problem representations. We searched for earlier work on science instruction that we could build upon. The aim was to find interventions that could contribute to students constructing a problem representation adequate for formal problem-solving. One reason why we found relatively few relevant studies, may be that recent science teaching research, like the current science curriculum in secondary education, has a declining interest in formal problem-solving. Although we recognise the merits of the current emphasis on conceptual understanding, we argued that the drawbacks this trend may have for further education cannot be neglected.

Among the studies we found, two types of intervention were most promising: firstly, those where the tasks were structured such that students' attention was directed to identifying relevant problem features, and secondly, those that trained students in using powerful representations, such as diagrams. We aimed to create a learning environment with similar measures that would also give feedback and support. We decided to implement a learning environment using a computer algebra system (CAS). A CAS can take over many algebraic calculations and some CASs offer powerful tools for visualisation. We used a specific CAS, namely Mathematica, to implement an electrostatics course module. We set our learning environment in such a way that it became rewarding to analyse a problem properly and to identify the proper solution approach. Our learning environment also provided support for visualisation.

As we had found clear differences between good and weak novices in the first part of our work, it would seem natural to provide different kinds of instructional support for good and weak students. Good students might best be supported in achieving a more flexible problem representation, which is integrated with solution information, whereas weak students might better be supported to construct a coherent representation. Our learning environment was designed to provide both kinds of support. It could have been informative to construct two versions of the learning environment, each providing one kind of support, and to test whether the learning outcome depends on interactions between student level and type of instruction. There are some factors against such an approach, however. This type of interactions, which have commonly been termed aptitude treatment interactions (ATI), has been studied in other domains. Quite often research fails to demonstrate such effects in realistic settings. This might be not because there are no such effects, but rather because aptitude is a composite trait, which makes straightforward predictions about the effects of instruction impossible. This is even more so for studies where aptitude has been equated with intelligence (compare Veenman & Elshout, 1991, 1993). In addition, in a realistic setting there are many disturbances, so that effect sizes are quite small[22]. Demonstrating such effects requires large numbers of subjects, or long periods of instruction to gain acceptable statistical power. As our resources were limited, this approach was unavailable. As a more substantial point, the course we had designed was intended as an exemplar of an approach to be followed in educational practice. As such, it should be compared to traditional instruction in the first place. Moreover, in current university educational practice, providing different instruction for good and weak students is an unrealistic option.[23] Therefore, ATI studies are of less direct relevance to applied university education research.

Based on these considerations, we chose to compare our experimental module with the normal approach. We did so using two different tests: the regular final exam and an adapted test that was specifically tailored towards our learning goals. Learning outcomes for both approaches were not significantly different. The experimental course, however, could successfully address a number of misconceptions, particularly through the use of interactive visualisations. The experimental course students found the visualisation tools more instructive than the problem-solving tools. The experimental course apparently imposed too high a cognitive load on the students. Based on these outcomes, we made proposals how to improve the course. We concluded that, with some improvements, the experimental course could be valuable to supplement the usual teaching approach.

Our experiment was done with students who took part in the normal curriculum. As a consequence, at their final exam they had to solve their assignments the traditional way. We have already emphasised the importance of correspondence between the abilities taught and the abilities required at the final test. When setting up our experiment, we noted the lack of standardised and validated tests that could be used to evaluate the learning outcome of an intervention, and to compare between different interventions. Most of the test items available place an emphasis on carrying out algebraic procedures. For mechanics there are some qualitative reasoning tests that are used more widely (Hestenes et al. 1992; Hestenes & Wells,

---

[22] Because within group variation of gain scores are high for any short course, even quite effective instructional measures may render small statistical effect sizes. This does not necessarily imply that the effect is unimportant to educational practice.

[23] Considering the current state in intelligent tutoring systems, for most domains individually tailored instruction will not be feasible in the near future. On the other hand technology provides learners with increasing opportunities to adapt learning environments to their needs themselves.

1992), but these do not address the choice of solution methods. Further development of validated tests for specific kinds of knowledge could clearly contribute to the advancement of physics education research.

When considering the introduction of a CAS in education, there are additional reasons to not only reconsider the tests used, but also revise educational goals. The use of CASs extends to professional problem-solving practice, and with the availability of such tools the importance of many computational skills has to be reconsidered. With the use of a CAS, students in the first year of their study can solve problems traditionally taught in the second or third year of the study (Van Weert, personal communication; Simons, 1998). For this reason, the increasing availability of CA tools, and, likewise, numerical tools, offers opportunities to teach computational skills only insofar as they contribute to understanding. The importance of computational skill to understanding has been debated over ever since the introduction of hand-held calculators, but the issue remains unsettled.

Studies such as our final experiment, with realistic learning tasks in realistic settings, provide global information about the effectiveness of an instructional design, but they can only render limited information about the differences in learning processes between good and weak students. Further research could be directed at a more in-depth analysis of students' behaviour when working with a CAS-based learning environment. This could help to reduce extraneous cognitive load, and to enhance active learning. An in-depth analysis could also provide insight into how good and weak students differ in using specific kinds of support. Nevertheless, even such an in-depth study can only provide limited information about cognitive processes and mental representations in good and weak beginners. Their understanding requires more restricted tasks and more controlled experimental set-ups, for instance, to compare the effects different types of representational tools have on learning. It remains one of cognitive psychology's major challenges to detail our understanding of complex learning processes, such as those that occur while learning physics problem-solving. Meeting this challenge requires controlled experiments that must retain the complexity, knowledge intensiveness and creativity, which form the essence of physics problem-solving.

Samenvatting (Summary)

We hebben onderzocht hoe de probleemrepresentaties van natuurkundestudenten verbeterd kunnen worden. Het is in meerdere onderzoeken gebleken dat goede probleemoplossers zich probleemsituaties beter voor kunnen stellen dan zwakke probleemoplossers (Chi et al., 1981; Chi & Bassok, 1989; De Jong & Ferguson-Hessler, 1991; Larkin, 1983). Vaak worden deze verschillen in één woord aangeduid als verschillen in diepte. Wat deze verschillen precies inhouden is echter veel minder duidelijk. In dit onderzoek hebben we allereerst in kaart gebracht hoe de probleemrepresentaties van succesvolle en minder succesvolle probleemoplossers precies van elkaar verschillen. Vervolgens hebben we onderzocht of ook de constructie van een probleemrepresentatie verschillend verloopt bij succesvolle en minder succesvolle probleemoplossers. De uitkomsten van deze eerste twee onderzoeken waren aanleiding om een leeromgeving te ontwerpen met extra ondersteuning voor het verbeteren van probleemrepresentaties. De effectiviteit van deze leeromgeving hebben we tenslotte getest in een derde experiment. Bij alle onderzoeken gebruikten we elektriciteitsleer als vakinhoud. We bespreken hieronder kort de verschillende onderzoeken.

In een eerste experiment hebben we verschillen tussen probleemrepresentaties van experts en van goede en zwakke beginners in kaart gebracht. We gingen er vanuit dat verschillen in 'diepte' gevolgen zouden hebben voor het type concepten in de probleemrepresentatie, voor de relaties die tussen deze concepten gelegd worden en voor de globale structuur van de probleemrepresentatie. We ontwierpen een methode waarbij proefpersonen probleemsituaties beschrijven naar aanleiding van een natuurkundeformule. We nemen aan dat de situatiebeschrijving een afspiegeling vormt van een onderliggend mentaal model. De verzamelde situatiebeschrijvingen kunnen vervolgens op verschillende manieren geanalyseerd worden. We onderzochten verschillen in het gebruik van typen woorden en typen zinnen, en we onderzochten verschillen in de globale structuur van de situatiebeschrijvingen. We concluderen dat probleemrepresentaties van goede probleem-oplossers zich vooral onderscheiden door hun structuur en door de flexibele koppeling tussen verschillende soorten informatie — zoals tussen abstracte en concrete informatie en tussen probleemkenmerken en oplosinformatie. We vonden geen aanwijzing dat een bepaald soort concepten of een bepaalde inhoud kenmerkend zou zijn voor representaties van goede probleem-oplossers. Deze uitkomst verschilt duidelijk van de populaire opvatting dat goede probleem-oplossers vooral abstracte en gegeneraliseerde representaties zouden hebben.

Om een gestructureerde en flexibele voorstelling van een situatie te vormen is het nodig extra informatie toe te voegen aan de oorspronkelijke probleemstelling. Dit proces, dat we elaboratie noemen, verbindt elementen uit de probleembeschrijving met kennis uit het lange-termijngeheugen. Op basis van de resultaten van ons eerste experiment veronderstelden we dat succesvolle probleemoplossers vloeiend de benodigde informatie uit hun lange-termijngeheugen toevoegen bij het interpreteren van een probleembeschrijving, terwijl zwakke probleemoplossers er niet in slagen hun kennis te benutten om de elementen uit de probleembeschrijving op een goede manier te verbinden. We hebben onderzocht of beginnende probleemoplossers die eenvoudige elaboraties expliciet aangeboden krijgen deze informatie kunnen benutten om zelf verder te redeneren en zo een diepe probleemrepresentatie te vormen. We ontwierpen hiertoe een aangepaste versie van de bekende probleemcategorisatietaak: proefpersonen kregen een set kaartjes met op ieder kaartje een probleembeschrijving, de opdracht is deze kaartjes te sorteren op oplosmethode. We maakten twee versies van de probleemset: een versie met minimale beschrijvingen van ieder probleem en een versie met een eenvoudige elaboratie bij ieder probleem. We vergeleken de kwaliteit van de probleemsorteringen van goede en zwakke beginners, en het effect van de toegevoegde elaboraties voor beide groepen. Zowel goede als zwakke beginners bleken in staat oplosmethoden te benoemen en problemen met enig succes op oplosmethode te sorteren. De toegevoegde elaboraties hadden alleen een positief effect voor de goede beginners. Dit resultaat wijst op kwalitatieve verschillen tussen de redeneerprocessen van goede en zwakke beginners. We concluderen dat beginners wel probleemtypen kennen, maar dat de elaboratie van het probleem een groot struikelblok is.

Nadat de eerste twee onderzoeken hadden verduidelijkt hoe de mentale representaties van beginnende natuurkundeprobleemoplossers tekortschoten, besloten we een leeromgeving te ontwerpen speciaal gericht op het verbeteren van probleemrepresentaties. We ontwierpen een leeromgeving waarin de aandacht van studenten gericht zou worden op het herkennen van belangrijke probleemkenmerken en die ondersteuning zou bieden bij het maken van visuele representaties. We implementeerden onze leeromgeving in een computeralgebrasysteem (CAS), namelijk Mathematica. Zo'n CAS kan een groot deel van het rekenwerk overnemen, mits het probleem precies wordt gespecificeerd. Dan kan ook de oplossing van een eerder probleem 'gekopieerd' worden naar een nieuw probleem. Bovendien kan een CAS gebruikt worden om problemen en oplossingen te visualiseren.

We hebben het leerresultaat met de nieuw ontworpen leeromgeving experimenteel vergeleken met het effect van traditionele werkcolleges. Daartoe hebben we een groep studenten zeven uur laten werken met de nieuwe leeromgeving terwijl een tweede groep studenten zeven uur traditioneel werkcollege volgde. Onze proefpersonen volgden naast het experiment ook hun reguliere cursus elektriciteitsleer. We vergeleken de prestaties van onze beide groepen proefpersonen op het reguliere tentamen en op een aangepaste test die speciaal gericht was op de kwaliteit van de probleemrepresentatie. Op geen van beide tests vonden we significante verschillen tussen beide groepen. Studenten vonden de nieuwe methode wel motiverend en studenten kwamen vooral door de visualisaties tot nieuwe inzichten. Door de korte duur van het experiment ging echter relatief veel tijd verloren aan het leren werken met Mathematica. Daarnaast bleek ook dat de leeromgeving op een aantal punten onvoldoende duidelijk was, en dat de structuur van de opgaven nog onvoldoende dwingt tot nadenken over de betekenis van oplossingen. Ook de context waarin het experiment plaatsvond, namelijk de reguliere cursus en het reguliere tentamen, was niet optimaal. In het bijzonder zou het wenselijk zijn dat de specifieke vaardigheden die geleerd worden in de Mathematica versie van de cursus op het tentamen benut zouden kunnen worden. We concluderen dan ook dat er voldoende aanknopingspunten zijn om in een vervolgonderzoek een verbeterde versie van de leeromgeving uit te testen.

References

Alexander, P.A., & Judy, J.E. (1988). The interaction of domain-specific and strategic knowledge in academic performance. *Review of Educational Research*, *58,* 375-405.

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

Baddely, A.D. (1986). *Working memory*. Oxford: Clarendon Press.

Baron, R.S., Inman, M.L., Kao, C.F., & Logan, H. (1992). Negative emotion and superficial social processing. *Motivation and Emotion, 16,* 323-345.

Bartlett, F.C. (1932). *Remembering*. Cambridge: Cambridge University Press.

Beishuizen, J., Stoutjesdijk, E., Van Putten, K. (1994). Studying textbooks: Effects of learning styles, study task, and instruction. *Learning and Instruction*, *4,* 151-174.

Blessing, S.B., & Ross B.H. (1996). Content effects in problem categorization and problem solving. *Journal of Experimental Psychology, Learning and Cognition*, *22,* 792-810.

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.

Bredeweg, B. (1992). Expertise in qualitative prediction of behaviour. Unpublished doctoral dissertation, University of Amsterdam.

Bunce, D.M., Gabel, D.L., & Samuel, J.V. (1991). Enhancing chemistry problem-solving achievement using problem categorisation. *Journal of Research in Science Teaching*, *25,* 505-521.

Bunge, M. (1967). *Scientific research I, the search for system*. New York: Springer.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8,* 293-332.

Chandler, P., & Sweller, J. (1992). The split-attention effect as a factor in the design of instruction. *British Journal of Educational Psychology, 62,* 233-246.

Chi, M.T.H., & Bassok, M. (1989). Learning from examples via self-explanations. In L.B. Resnick (Ed.), *Knowing, learning, and instruction: essays in honour of Robert Glaser* (pp. 251-282). Hillsdale, NJ: Lawrence 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 R.J. Sternberg (Ed.), *Advances in the psychology of human intelligence*. Hillsdale, NJ: Lawrence Erlbaum.

Cooke, N.J. (1992). Modelling human expertise in expert systems. In R.R. Hoffman (Ed.), *The psychology of expertise: Cognitive research and empirical AI* (pp. 29-60). New York: Springer.

Corbett, A.T., & Anderson, J.R. (1992). LISP Intelligent tutoring system: research in skill aquisition. In J.H. Larkin & R.W. Chabay (Eds.), *Computer-assisted instruction and intelligent tutoring systems: Shared goals and complementary approaches* (pp. 73-109). Hillsdale, NJ: Lawrence Erlbaum.

Craik, F.I.M., & Lockhart, R.S. (1972). Levels of processing: a framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11,* 671-684.

De Groot, A.D. (1946). *Het denken van den schaker* [Thought in chessplayers]. Published doctoral dissertation, Amsterdam: Noordhollandse Uitgeversmaatschappij.

De Groot, A.D. (1965). *Thought and choice in chess*. The Hague: Mouton.

De Jong, T. (1986). *Kennis en het oplossen van vakinhoudelijke problemen* [Knowledge based problem solving]. Unpublished doctoral dissertation, Eindhoven University of Technology, The Netherlands.

De Jong, T., & Ferguson-Hessler, M.G.M. (1986). Cognitive structures in good and poor novice problem solvers in physics. *Journal of Educational Psychology, 78,* 279-288.

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: Lawrence 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: Lawrence Erlbaum.

Dennett, D.C., & Kinsbourne, M. (1992). Time and the observer: The where and when of conciousness in the brain. *Behavioral and Brain Sciences*, *15,* 183-247.

diSessa, A.A. (1983). Phenomenology and the evolution of intuition. In D. Gentner & A.L. Stevens (Eds.), *Mental models* (pp. 15-33). Hillsdale, NJ: Lawrence Erlbaum.

diSessa, A.A. (1993). Toward an epistemology of physics. *Cognition and instruction*, *10,* 105-225.

Duncker, K. (1945). On problem-solving. *Psychological Monographs*, *58(270),* 1-113.

Egan, D.E., & Schwarz, B.J. (1979). Chunking in recall of symbolic drawings. *Memory and Cognition*, *7,* 149-158.

Elio, R., & Scharf, P.B. (1990). Modelling novice-to-expert shifts in problem solving strategy and knowledge organization. *Cognitive Science*, *14,* 579-639.

Entwistle, N.J., & Ramsden, P. (1983). *Understanding student learning*. London: Croom Helm.

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

Feiner-Valkier, S. (1997). *Achtergronden bij de invoering van een nieuwe methode voor de propedeusevakken Mechanica en Elektriciteit & Magnetisme van de faculteit Technische Natuurkunde in het collegejaar 1996/1997* [Backgrounds for the introduction of a new method for the first-year courses on Mechanics and on Electricity and Magnetism in 1996/1997], internal report, Eindhoven, The Netherlands: Eindhoven University of Technology.

Ferguson-Hessler, M.G.M., & De Jong, T. (1990). Studying physics texts: differences in study processes between good and poor performers. *Cognition and Instruction, 7,* 41-54.
Feuerstein, R. (1980). *Instrumental enrichment: An intervention program for cognitive modifiability*. Baltimore: University Park Press.
Feynman, R.P. (1965). *The character of physical law.* London: British Broadcasting Company.
Feynman, R.P. (1997). *Six not-so-easy pieces.* Reading, MA: Addison-Wesley.
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., & Bassok, M. (1989). Learning theory and the study of instruction. *Annual Review of Psychology*, *40,* 631-666.
Goei, S.L. (1994). Mental models and problem solving in the domain of computer numerically controlled programming. Unpublished doctoral dissertation, Twente University, The Netherlands.
Goel, V., & Pirolli, P. (1992). The structure of design problem spaces. *Cognitive Science*, *16,* 395-429.
Good, P. (1994). *Permutation tests: A practical guide to resampling methods for testing hypotheses.* New York: Springer.
Gravenberch, F. (1996). *Voorlichtingsbrochure havo/vwo: Actuele stand van zaken invoering tweede fase* [Information brochure havo/vwo: Introduction of the second phase]. Enschede, The Netherlands: Instituut voor Leerplanontwikkeling.
Greenacre, M.J. (1993). *Correspondence analysis in practice.* London: Academic Press.
Greeno, J.D., & Berger, D. (1987). *A model of functional knowledge and insight* (Technical report GK-1). Berkely, University of California.
Griffiths, D.J. (1989). *Introduction to electrodynamics*. London: Prentice-Hall.
Gruber, H., & Ziegler, A. (1995). Components of expertise: looking for SEEK in sorting. *Review of Psychology*, *2,* 13-21.
Halliday, D., Resnick, R., & Walker, J. (1993). *Fundamentals of physics* (4th ed.). Chichester: Wiley.
Halloun, I.A., & Hestenes, D. (1987). Modelling instruction in mechanics. *American Journal of Physics*, *55,* 455-462.
Hammer, D. (1994). Epistemological beliefs in introductory physics. *Cognition and Instruction, 12,* 151-183.
Hayes, J.R., & Simon, H.A. (1979). Understanding written problem instructions. In H.A. Simon (Ed.), *Models of thought.* New Haven, CT: Yale University Press. (Reprinted from Gregg. L.W. (Ed.). (1974). *Knowledge and cognition.* Hillsdale, NJ: Erlbaum.)
Hestenes, D. (1987). Toward a modelling theory of physics instruction. *American Journal of Physics*, *55,* 441-454.
Hestenes, D., & Wells, M. (1992). A mechanics baseline test. *The Physics Teacher*, *30,* 159-166.
Hestenes, D., Wells, M., & Swackhammer, G. (1992). Force concept inventory. *The Physics Teacher*, *30,* 141-158.

Hinsley, D.A., Hayes, J.R., & Simon, H.A. (1978). From words to equations: meaning and representation in algebra word problems. In P.A. Carpenter & M.A. Just (Eds.), *Cognitive processes in comprehension* (pp. 89-106). Hillsdale, NJ: Lawrence Erlbaum.
Johnson-Laird, P.N. (1983). *Mental models: Towards a cognitive science of language, inference and conciousness*. Cambridge: Cambridge University Press.
Keune, F.J. (1998). *Naar de knoppen* [Mathematics education under threat]. Published inaugural lecture, Catholic University Nijmegen, The Netherlands.
Kintsch, W. (1988). The role of knowledge in discourse comprehension: a construction-integration model. *Psychological Review*, *95,* 163-182.
Kintsch, W. (1992). A cognitive architecture for comprehension. In H.L. Pick Jr., P.W. van den Broek, & D.C. Knill (Eds.), *Cognition: Conceptual and methodological issues* (pp. 143-163). Washington, DC: American Psychological Association.
Kolodner, J.L. (1984). Towards an understanding of the role of experience in the evolution from novice to expert. In M.J. Coombs (Ed.), *Developments in expert systems* (pp. 95-116). London: Academic Press.
Landa, L.N. (1969). *Algoritmierung im Unterricht* [The development of algorithms in education]. Berlin: Volk und Wissen.
Larkin, J.H. (1983). The role of problem representations in physics. In D. Gentner & A.L. Stevens (Eds.), *Mental models* (pp. 75-98). Hillsdale, NJ: Lawrence Erlbaum.
Larkin, J.H., McDermott, J., Simon, D.P., & Simon, H.A. (1980). Expert and novice performance in solving physics problems. *Science, 208,* 1335-1342.
Larkin, J.H., & Simon, H.A. (1987). Why a diagram is (sometimes) worth ten thousand words, *Cognitive Science, 11,* 65-100.
Lesgold, A., Rubinson, H. Feltovich, P., Glaser, R., Klopfer, D., & Wang, Y. (1988). Expertise in a complex skill: diagnosing X-ray pictures. In M.T.H. Chi, R. Glaser, & M.J. Farr (Ed.), *The nature of expertise* (pp. 229-260). Hillsdale, NJ: Lawrence Erlbaum.
Lewis, C. (1981). Skill in algebra. In J.R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 85-110). Hillsdale NJ: Lawrence Erlbaum
Maier, N.R.F. (1931). Reasoning in humans II: The solution of a problem and its appearance in conciousness. *Journal of Comparative Psychology*, *12,* 181-194.
Manly, B.F.J. (1983). Analysis of polymorphic variation in different types of habitat. *Biometrics,* 39, 13-27.
Mansfield, R.S., Busse, T.V., & Krepelka, E.J. (1978). The effectiveness of creativity training. *Review of Educational Research, 48,* 517-536.
Martin, J. (1984). Toward a cognitive schema theory of self-instruction. *Instructional Science*, *13,* 159-180.
Mayer, R.E., & Sims, V.K. (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multi-media learning. *Journal of Educational Psychology, 86,* 389-401.
McCloskey, M. (1983). Naive theories of motion. In D. Gentner & A.L. Stevens (Eds.), *Mental models* (pp. 299-324). Hillsdale NJ: Lawrence Erlbaum.

McDermott, J., & Larkin, J.H. (1978). Re-representing textbook physics problems. In *Proceedings of the 2nd national conference of the Canadian Society for Computational Studies of Intelligence.*
McNamara, D.S., Kintsch, E., Butler Songer, N., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction*, *14,* 1-43.
Mestre, J.P., Dufresne, W.J., Gerace, J., Hardiman, P.T., & Touger, J.S. (1993). Promoting skilled problem-solving behavior among beginning physics students. *Journal of Research in Science Teaching*, *30,* 303-317.
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.
Minsky, M.L. (1975). A framework for representing knowledge. In P.H. Winston (Ed.), *The psychology of computer vision* (pp. 211-277).
Minstrell, J.A. (1989). Teaching science for understanding. In L. Resnick & L.E. Klopfer (Eds.), *Toward the Thinking Curriculum: Current Cognitive Research 1989 ASCD Yearbook* (pp. 129-149). Alexandria, VA: Association for Supervison and Curriculum Development.
Nathan, M.J., Kintsch, W., & Young, E. (1992). A theory of algebra-word-problem comprehension and its implications for the design of learning environments. *Cognition and Instruction, 9*, 329-389.
Newell, A., & Simon, H.A. (1972). *Problem solving.* Englewood Cliffs, NJ: Prentice-Hall.
Noordman, L.G., Vonk, W., & Kempff, H.J. (1992). Causal inferences during the reading of expository texts. *Journal of Memory and Language*, *31,* 573-590.
Paas, F.G.W.C., & Van Merriënboer, J.J.G. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: a cognitive load approach. *Journal of Educational Psychology*, *86,* 122-133.
Paivio, A. (1986). *Mental representations: A dual coding approach*. Oxford: Oxford University Press.
Pankratius, W.J. (1990). Building an organised knowledge base: Concept mapping and achievement in secondary school physics. *Journal of Research in Science Teaching*, *27,* 315-333.
Pask, G. (1976). Styles and strategies of learning. *British Journal of Educational Psychology*, *46,* 128-148.
Piaget, J. (1970). Piaget's theory (G. Gellerier and J. Langer, Trans.). In P.H. Mussen (Ed.), *Carmichael's manual of child psychology, vol. 1 (3rd ed.)* (pp. 703-732). New York: Wiley.
Plötzner, R. (1995). The construction and coordination of complementary problem representations in physics. *Journal of Artificial Intelligence in Education*, *6,* 203-238.
Plötzner, R., & Spada, H. (1993). Multiple mental representations of information in physics problem solving. In G. Strube & K.F. Wender (Eds.), *The cognitive psychology of knowledge* (pp. 285-312). Amsterdam: Elsevier Science Publishers.

Polk, T.A., & Newell, A. (1995). Deduction as verbal reasoning. *Psychological Review, 102,* 533-566.
Priest, A.G., & Lindsay, R.O. (1992). New light on novice-expert differences in physics problem solving. *British Journal of Psychology*, *83,* 389-405.
Reber, A.S. (1985). *Dictionary of psychology*. London: Penguin.
Reder, L.M., Charney, D.H., & Morgan, K.I. (1986). The role of elaborations in learning a skill from an instructional text. *Memory and Cognition*, *14,* 64-78.
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. (1996). Guest comment: Standards and measurements in physics - Why not in physics education? *American Journal of Physics*, *64,* 687-688.
Reif, F. Larkin, J.H., & Bracket, G.C. (1976). Teaching general learning and problem-solving skills. *American Journal of Physics*, *44*, 212-217.
Reif, F., & Heller, J.I. (1982). Knowledge structure and problem solving in physics. *Educational Psychologist*, *17,* 102-107.
Riding, R., & Douglas, G. (1993). The effect of cognitive style and mode of presentation on learning performance. *British Journal of Educational Psychology*, *63,* 297-307.
Royer, J.M., Cisero, C.A., & Carlo, M.S. (1993). Techniques and procedures for assessing cognitive skills. *Review of Educational Research*, *63,* 201-243.
Rumelhart, D.E., & Norman, D.A. (1985). Representation of knowledge. In A.M. Aitkenhead & J.M. Slack (Eds.), *Issues in cognitive modelling* (pp. 15-62). Hove, UK: Lawrence Erlbaum.
Rumelhart, D.E., & Ortony, A. (1977). The representation of knowledge in memory. In R.C. Anderson, R.J., Spiro, & W.E. Montague (Eds.), *Schooling and the acquisition of knowledge* (pp. 99-137). Hillsdale, NJ: Lawrence Erlbaum.
Rumelhart, D.E., McClelland, J.L., & the PDP Research Group (1986). *Parallel distributed processing, vol. 1: foundations*. Cambridge, MA: MIT Press.
Rumelhart. D.E., & Norman, D.A. (1981). *Analogical processes in learning.* Hillsdale NJ: Lawrence Erlbaum.
Rutherford, F.J., & Ahlgren, A. (1991). *Science for all Americans*. New York: Oxford University Press.
Schank, R.A., & Abelson, R. (1977). *Scripts, plans, goals and understanding.* Hillsdale, NJ: Lawrence Erlbaum.
Schoenfeld, A.H., Herrmann, D.J. (1982). Problem perception and knowledge structure in expert and novice mathematical problem solvers. *Journal of Experimental Learning*, *8,* 484-494.
Shastri, L., & Ajjanagadde, V. (1993). A connectionist model of reflexive processing. *Behavioral and Brain Sciences*, *16,* 417-494.
Simon, H. (1973). The structure of ill-defined problems. *Artificial Intelligence*, *4,* 181-202.
Simons, F.H. (1998). *Onderwijs is mensenwerk: Over het gebruik van computers in het onderwijs* [Teaching is a work of man: On the use of computers in education].

Published valedictory lecture, Eindhoven University of Technology, The Netherlands.

Sloman, A. (1995). Musings on the roles of logical and non-logical representations in intelligence. In J. Glasgow, N.H. Narayanan, & B. Chandrasekaran (Eds.), *Diagrammatic reasoning: cognitive and computational perspectives* (pp. 7-32). Menlo Park, CA: AAAI Press/The MIT Press.

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

Smolensky, P. (1988). On the proper treatment of connectionism. *Behavioral and Brain Sciences*, *11,* 1-74.

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

Stewart, J.H., & Atkin, J.A. (1982). Information processing psychology: A promising paradigm for research in science teaching. *Journal of Research in Science Teaching*, *19,* 321-332.

Swaak, J., & De Jong, T. (1996). Measuring intuitive knowledge in science: The development of the WHAT-IF test. *Studies of Education Evaluation, 22,* p. 341-362.

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

Sweller, J., & Cooper, G.A. (1985). The use of worked examples as a substitute for problem-solving in learning algebra. *Cognition and Instruction*, *2,* 59-89.

Sweller, J., Mawer, R.F., & Ward, M.R. (1983). Development of expertise in mathematical problem solving. *Journal of Experimental Psychology: General, 112,* 639-661.

Szajna, B., Mackay, J.M. (1995). Predictors of learning performance in a computer-user training environment: A path-analytic study. *International Journal of Human Computer Interaction*, *7,* 167-185.

Taconis, R. (1995). *Understanding based problem solving.* Unpublished doctoral dissertation, Eindhoven University of Technology, The Netherlands.

Tulving, E. (1983). Elements of episodic memory. New York: Oxford University Press.

Van Heuvelen, A. (1991a). Learning to think like a physicist: A review of research-based instructional strategies. *American Journal of Physics*, *59,* 891-897.

Van Heuvelen, A. (1991b). Overview, case study physics. *American Journal of Physics*, *59,* 898-907.

Van Hiele, P.M. (1957). *De problematiek van het inzicht, gedemonstreerd aan het inzicht van schoolkinderen in meetkunde-leerstof* [The problem of insight, demonstrated for pupils understanding in geometry]. Published doctoral dissertation, Amsterdam: Meulenhoff.

Van Hiele, P.M. (1986). *Structure and Insight, a theory of mathematics education*. London: Academic Press.

Van Weeren, J.H.P., De Mul, F.F., Peters, M.J., Kramers-Pals, H., & Roossink, H.J. (1982). Teaching problem-solving in physics: A course in electromagnetism. *American Journal of Physics*, *50,* 725-732.
VanderStoep, S.W., & Seifert, C.M. (1997). Learning 'how' versus learning 'when': Improving transfer on problem-solving principles. *Journal of The Learning Sciences*, *3,* 93-111.
VanLehn, K. (1989). Problem solving and cognitive skill acquisition. In M.I. Posner (Ed.), *Foundations of cognitive science* (pp. 527-580). Cambridge, MA: MIT Press.
Veenman, M.V.J., Elshout, J.J. (1991). Intellectual ability and working method as predictors of novice learning. *Learning and Instruction*, *1,* 303-317.
Veenman, M.V.J., & Elshout, J.J. (1993). Differential effects of instructional support on learning in simulation environments. *Instructional Science*, *22,* 363-384.
Vermunt, J.D. (1991). Leerstrategieen van studenten in een zelfinstructie-leeromgeving [Learning strategies of students in a self-instructional learning environment]. *Pedagogische Studiën*, *68,* 315-325.
Wilder, D.A., & Shapiro, P. (1989). Effects of anxiety on impression formation in a group context: An anxiety-assimilation hypothesis. *Journal of Experimental Social Psychology, 25,* 481-499.
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.
Young, H.D., & Freedman, R.A. (1996). *University physics* (9th ed.). Amsterdam: Addison Wesley.
Young, R.M. (1983). Surrogates and mappings: Two kinds of conceptual models for interactive devices. In D. Gentner & A.L. Stevens (Eds.), *Mental models* (pp. 35-52). Hillsdale, NJ: Lawrence Erlbaum.
Zhu, X., & Simon, H. (1987). Learning mathematics from examples and by doing. *Cognition and Instruction*, *4,* 137-166.

Index

Appendices

Appendix A Permutation procedure for significance testing

Table 4 in Section 2 summarises observed relations between expertise and relative frequencies for three types of words. Full results are given below:

subject	expertise	object	attribute	relation
id7	lecturer	78	34	31
id8	lecturer	165	151	49
id24	lecturer	134	83	37
id9	PhD candidate	188	146	56
id10	PhD candidate	262	182	87
id23	PhD candidate	129	108	37
id14	good stud.	115	86	48
id15	good stud.	169	49	55
id16	good stud.	141	74	73
id17	good stud.	121	30	34
id20	good stud.	249	120	85
id21	good stud.	132	83	58
id11	weak student	154	113	79
id12	weak student	123	91	73
id13	weak student	133	87	60
id18	weak student	109	84	76
id19	weak student	131	39	95
id22	weak student	103	32	62

If we were interested in between person differences, we could apply a chi-square test to the table above. Since we are looking after differences between expertise groups, however, this approach does not work. Collapsing all observations for an expertise group is not an option either because all 377 observations of *object* words by lecturers, for instance, come from only three lecturers, so that they are not independent observations. The table could be analysed with a repeated measures ANOVA if not for the small and unequal group sizes.

A safe approach in such cases is to use a permutation test. We follow an approach presented by Manly (1983). First, a suitable test statistic has to be computed on the original data; then the subjects are permuted randomly over the expertise groups and the statistic is computed again on the permuted data. The last two steps are repeated a number of times. The last step in the procedure is to compute the proportion of cases where the test statistic on the permuted data was more extreme than the value for the original data. This proportion is interpreted as the p-value.

Manly (1983) proposes computing Pearson's chi-square on the collapsed table as a test statistic. We choose to use the maximum likelihood G^2 instead. Unlike Pearson's test-statistic, G^2 can be partitioned. When the table above is split up into four sub-tables, one for each expertise group, we can compute chi-square for each sub-table. With the use of G^2, the sum of these four values plus the G^2 for the collapsed table equals the G^2 for the total table above. As a consequence, a minimal sum of chi-squares for the sub-tables per expertise group becomes equivalent to a maximal chi-square between expertise groups. With Pearson's chi-square these two test-statistics would not be equivalent, which would be unsatisfactory. Prior to computing G^2, we added 0.5 to all cell values, because of empty cells.

Appendix B Top-70 wordlist, classification, and frequencies

The table below shows frequency data for the 70 most frequent physics words in the experimental protocols of 18 subjects. The 18 protocols comprised 6362 sentences consisting of 5.2×10^4 words in total.

In the first three columns for every word, its membership of several categories is indicated

Nr	1)	2)	3)	word (Dutch)	translation	frequency
1.	a	o		veld	field	645
2.				stroom	current	335
3.	c			lading	charge	262
4.		r	f	kracht(en)	force(s)	246
5.		a		magnetisch	magnetic	217
6.		a		elektrisch(e)	electric(al)	195
7.	c	o		platen/plaat	plate/plates	193
8.	a	r	f	spanning	tension	189
9.	c	o		bol(len)	sphere(s)	174
10.	c	r	t	tussen	between	167
11.		o		oppervlak	surface	161
12.	c			loopt/lopen	runs/run	160
13.		o		condensator	capacitor	146
14.	c	o		deeltje(s)	particle(s)	142
15.	c			zit/zitten	is/are located	137
16.				verandert/veranderen/verandering	change(s)	125
17.	c			beweging/beweging/beweegt	move(s)	124
18.	c	o		draad	wire	122
19.	c	r	t	afstand	distance	122
20.	c	o		stroomdraad	electric wire	111
21.	c	o		spoel/spoelen	coil	108
22.	a			inductie	inductance	104
23.		o		puntlading/puntladingen	point charge(s)	98
24.		o		punt/punten	point(s)	96
25.		a		groot	large	92
26.				integraal/integreren	integral/integrate	92
27.	a			energie	energy	88
28.				oppervlakte	area	82
29.	a			veldsterkte	field intensity	77
30.		a		magnetische	magnetic	76
31.	a			flux	flux	75
32.				lorentzkracht	Lorentz force	70
33.	a	o		ruimtelading	spatial charge	66
34.	a	a		symmetrie	symmetry	66
35.	c	o		cilinder	cylinder	65
36.	a	o		b-veld/magneetveld	magnetic field	62
37.	a			oneindig	infinite	60
38.		a		geladen	charged	57
39.		r	f	spanningsverschil	potential difference	55
40.	c	o		ladingen	charges	53

1) a = abstract, c = concrete

2) o = object, a = attribute, r = relation

3) f = functional, t = topological

Nr	1)	2)	3)	word (Dutch)	translation	frequency
41.		o		oppervlaktelading	surface charge	50
42.	c	a		snelheid	velocity	50
43.	a	a		constant	constant	47
44.	c	r	t	buiten	outside	46
45.	c	r	t	doorheen	through	45
46.	c			vlak	flat (surface)	43
47.	a			arbeid	work (quantity)	42
48.				werkt	acts	40
49.	c	r	t	binnen	inside	39
50.	a	a		homogeen	homogeneous	39
51.				inductiestroom	inducted current	39
52.				kring	loop/circuit	39
53.				nul	zero	39
54.				ontstaat	arises	38
55.	a	a		potentiaal	potential	38
56.				ruimte	space	37
57.	c			plaats	place	36
58.	a			geleider	conductor	35
59.		r	f	veroorzaakt	causes	35
60.	a			constante	constant	35
61.	c	o		staaf	stick	34
62.		r	t	omsloten	enclosed	34
63.		a		gesloten	closed	33
64.		o		oppervlaktestroom	surface current	32
65.	c	a		straal	radius	32
66.	c	a		rond	round	32
67.	a	a		bolsymmetrie	spherical symmetry	30
68.	c	r	t	verschuiven	shift	29
69.	c	a		richting	direction	29
70.		o		elektron	electron	28

Appendix C Sentence coding schema with examples

The examples below have been translated from Dutch. Text between square brackets has been taken from adjacent utterances by the author in order to clarify the context of a given utterance.

Experimenter's remark

you can't relate these elements?

that's all right

what is the meaning of the word in the present context?

Explanation of the formula/keyword

potential difference, that is the total electrical intensity from r to minus r

that way you can use Gauss' law for evaluating surface charge

[in this formula] the roles of the two charges can be exchanged

Episode concerning the problem

well this is an assignment we did in the exam before last summer

this is on the cover of the lecture notes

that [a variable capacitor] is used in old radio's

Evaluative remark regarding the problem or the stimulus

you cannot make another assignment out of this

surface charge, the other one was on the superposition principle, that's better

no, this is just the same as I said before

Difference between two situations mentioned

(no example)

Solution method

I'll consider the plane to be a superposition of wires

you can compute this by first computing the energy that was in the capacitor before the slab is inserted, [and then the energy that in the capacitor afterwards]

[there is a field in space and you draw a box] and you have to take care, either the field should be perpendicular to the surface of the box, [or it should be parallel to the surface]

Goal (what to be solved)

and you want to compute the electrical field of it

you may compute the force that is exerted on a current that moves in a magnetic field

compute the distance given the energy

Elaboration on information already given

in the case of a capacitor, you may assume there is no outside field

[…] so, the electrical intensity decreases with the inverse of the distance

this implies a force is exerted on these particles

Information regarding the problem situation

what comes to my mind is a problem involving a cylinder

you don't have to think of a current contained in a conductor

[if you want to avoid complicated computations] you have to limit yourself to symmetrical distributions

Miscellaneous

this is a lousy formula

and more of these things

Appendix D Examples of problems with elaborations

The field of two concentric cylinders

This is a minimal problem statement, containing only a minimum of 'deep' information: Given two copper cylinders, o_1 and o_2, with radii r_1 and r_2, where the relation between the radii is given by $r_2=3r_1$. Both cylinders are aligned 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.

Formal description

give{situation$_1$,
- shape(o_1)=cylinder (1)
- shape(o_2)=cylinder (2)
- shape.radius(o_1)=R_1 (3)
- shape.radius(o_2)=R_2 (4)
- 3.R_1=R_2 (5)
- medium(o_1)=copper (6)
- medium(o_2)=copper (7)
- x_axis(h1) (8) *(By the introduction of an axis, it is possible to avoid directly giving away the relation between the two cylinders.)*
- orientation(h_1, o_1)=concentric (9)
- orientation(h_1, o_2)=concentric (10)
- connect(o_2, ground) (11)
- charge.per_meter(o_1)=Q_1} (12)

goal{ situation$_1$,
- electrostatic_potential(o_2)} (13)

Extra information to be added

From:		*It follows that:*		*By the use of inference:*
6	$\Rightarrow$	medium(o_1)=conductor	(14)	*copper is a conductor*
7	$\Rightarrow$	medium(o_2)=conductor	(15)	*copper is a conductor*
12	$\Rightarrow$	e_field(o_3)	(16)	*a charge induces an electric field*
$9\wedge10$	$\Rightarrow$	orientation(o_1, o_2)=concentric	(17)	*concentrism is a transitive property*
$5\wedge9\wedge10$	$\Rightarrow$	position(o_1, o_2)=in	(18)	*'concentric with' and 'smaller than' implies 'in'*
$1\wedge2\wedge9\wedge10$	$\Rightarrow$	symmetry(situation$_1$)=cylindrical	(19)	*if all elements are cylindrical and placed concentric to each other, the entire situation is cylinder symmetric*
$5\wedge7\wedge9\wedge10\wedge11$	$\Rightarrow$	region(h_1) position(h_1,o_2)=outside atposition(h_1,e_field.magnitude())=0	(20)	*a grounded conductor shields its inside from external influences and vice versa*
$5\wedge7\wedge9\wedge10\wedge11$	$\Rightarrow$	atposition(h_1,electrostatic_potential(o_3))=0	(21)	$V=\int Eds$
$7\wedge11$	$\Rightarrow$	region(h_2) position(h_2,o_2)=outer_surface atposition(h_2,charge.per_meter(o_2))=0	(22)	?

$5 \wedge 7 \wedge 9 \wedge 10 \wedge 12 \Rightarrow$ region(h_3)
position(h_3,o_2)=inner_surface
atposition(h_3,charge.per_meter(o_2))=-Q_1 (23) *the net charge enclosed in a closed surface that lies completely within a conductor amounts to zero*

$5 \wedge 7 \wedge 9 \wedge 10 \wedge 11 \wedge 12 \Rightarrow$ charge(o_2)= -charge(o_1) (24) *the total charge on a grounded system amounts to zero*

$1 \wedge 5 \wedge 9 \wedge 10 \wedge 12 \Rightarrow$ region(h_4)
position(h_4, o_1)=outside
position(h_4, o_2)=inside
atposition(h_4, e_field(o_3))$\cong Q_1/r$ (25) *inside a (hollow) conductor, the charge distribution outside the conductor has no influence, thus the field is the field of the inner cylinder*

$1 \wedge 5 \wedge 9 \wedge 10 \wedge 12 \Rightarrow$
atposition(h_4, elec_potential(o_3)) $\cong -Q_1 \ln r$ (26) *see 25*

$5 \wedge 7 \wedge 9 \wedge 10 \wedge 12 \Rightarrow$
atposition($h_1 \rightarrow h_3$, potential(o_3))= atposition($h_4 \rightarrow h_3$, potential(o_3))
(27) *the potential runs continuously across a surface charge*

Upon derivation of the final statement, the answer is within reach; all that has to be done is to fill in formulas and carry out computations. There are major differences in depth between the various statements. Statement 14 for example follows from the simple rule copper is a conductor, combined with the information that object$_1$ is made of copper. Statement 23 in contrast is based on the following: *inside a conductor the field amounts the zero,* plus *the total flux through a surface is proportional to the total charge enclosed,* plus some knowledge about geometry rules, plus information from statements 5, 7, 9, 10 and 12 and the generalisations used in inferring 15, 17 en 19.

A conducting sphere in an external field

This is a minimal problem statement, containing only a minimum of 'deep' information: A conducting sphere is placed in an initially homogeneous electric field (parallel to the x-axis). Compute the electric field.

Formal description

give{situation$_1$,
e_field(o_1)=homogeneous.vector (1)
x_axis(h1) (2)
orientation(h_1, o_1)=parallel} (3)
give{situation$_2$,
situation$_1$ interacts — revise initial field description!
shape(o_2)=sphere (4)
shape.radius(o_2)=R_1 (5)
medium(o_2)=conductor (6)
charge.magnitude(o_2)=0} (7)
goal{situation$_2$, shape(o1)} (8)

From a discourse point of view, this representation might be acceptable, but from a physics point of view this is hardly the case because (1) the description suggests a changing situation and (2) the final situation is not clearly defined. A better description of the final state would make use of the boundary conditions that are implicit in the initial description. Thus the inference of boundary conditions is one of the elaborations that has to be made to move from the initial problem description to an acceptable physics model:

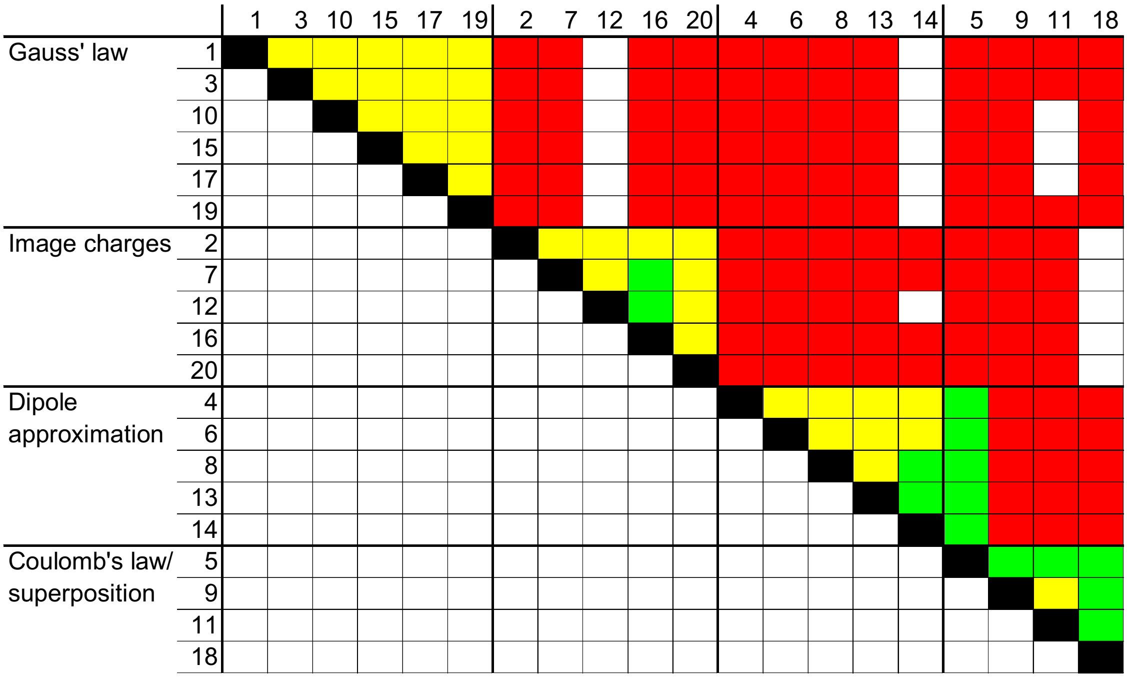
Extra information to be added

From:	***It follows that:***		***By the use of inference:***
$4 \wedge 6$	$\Rightarrow$ region(h_1) position(h_1,o_2)=inside e_field(situation$_2$)= e_field(o_1)+e_field(o_2) atposition(h_1,e_field.magnitude(situation$_2$))=0	(9)	*the field inside a conductor amounts to zero*
$1 \wedge 9$	$\Rightarrow$ field.magnitude(o_2)$\neq$0	(10)	
10	$\Rightarrow$ atposition(h_3,charge.per_cubic(o_2))$\neq$0	(11)	*a field is caused by a charge distribution*
$1 \wedge 4$	$\Rightarrow$ region(h_2) position(h_2,origin)=($x=\infty$) atposition(h_2,e_field(situation$_2$))=homogeneous.vector	(12)	*a finite charge distribution has only a local effect*
$1 \wedge 4$	$\Rightarrow$ region(h_3) position(h_3,origin)=($x=-\infty$) atposition(h_3,e_field(situation$_2$))=homogeneous.vector	(13)	*see 12*
6	$\Rightarrow$ region(h_4) position(h_4,o_2)=outer_surface atposition(h_4,perpendicular(h_4, field(situation$_2$)))	(14)	*the electric field a the surface of a conductor is perpendicular to the conductor surface*
$14 \wedge$	$\Rightarrow$ region(h_5) position(h_5,origin)=($x=0$) atposition($h_5 \cap h_4$,e_field(situation$_2$))=0	(15)	*symmetry considerations*

Now that the problem has been transformed into a proper boundary value problem, it can be solved either by changing to spherical coordinates and fitting a multipole series expansion to the boundary conditions, or by seeing that the initial field description can be replaced with two charged parallel infinite flat plates in $x=\infty$ and $x=-\infty$ respectively. Then the relation between field and surface charge would be $E = \sigma / \varepsilon_0$. This would result in an image dipole in the sphere. By applying (15) we find the magnitude of the dipole: $\vec{p} = 8\pi\varepsilon_0 E_0 R^3$, thus the resulting field would be $\vec{E} = \vec{E}_0 + \vec{p} / 8\pi\varepsilon_0 \vec{r}^{\,3}$. The latter solutions route again involves major restructuring of the problem representation.

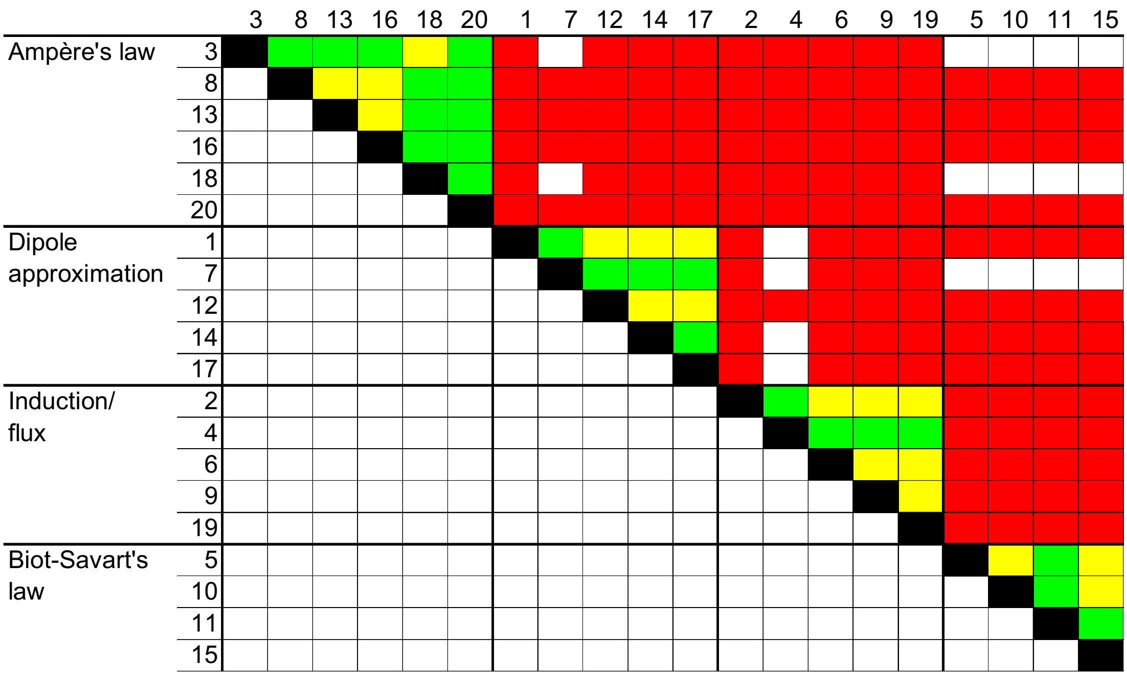
Appendix E Matrices for experts sortings

Electricity problems



(Maximum number of expert-like combinations: 32. Maximum number of not expert-like combinations: 123)

Magnetism problems



(Maximum number of expert-like combinations: 18. Maximum number of not expert-like combinations: 131)

Legend (for cells above the diagonal)

- (yellow) the experimenters and at least 2 out of 3 other experts placed these problems in the same pile
- (green) the experimenters placed these problems together, but at least 2 other experts placed them apart
- (red) neither the experimenters nor one of the other experts placed these problems in the same pile

Appendix F Dendrograms

In the following dendrograms the label numbers correspond to the label numbers in Appendix E. The distance measure used is Euclidean distances. As a clustering method, average linking between groups was applied.

```
              Rescaled Distance Cluster Combine

       0         5        10        15        20        25
       +---------+---------+---------+---------+---------+
Label
 E1     -+---------+
 E3     -+         +---------+
 E19    -----------+         +-------+
 E10    ---------------+-----+       +-----------------+
 E15    ---------------+             I                 I
 E17    -----------------------------+                 I
 E4     -----+-------------+                           +-+
 E6     -----+             +-----------------------+   I I
 E13    -------+-----------+                       I   I I
 E8     -------+                                   I   I I
 E11    ---------+-------------+                   +---+ I
 E9     ---------+             +---+               I     I
 E18    -----------------------+   +-------+       I     I
 E5     ---------------------------+       +-------+     I
 E14    -----------------------------------+             I
 E12    ---------+---------+                             I
 E7     ---------+         +-----------------------------+
 E16    -------+---+       I
 E2     -------+   +-------+
 E20    -----------+
```

Dendrogram of good students sortings of electricity problems (based on 36 subjects).

```
              Rescaled Distance Cluster Combine

       0         5        10        15        20        25
       +---------+---------+---------+---------+---------+
Label
 E16    -+-------+
 E2     -+       +-------------+
 E20    ---------+             +-----------+
 E18    -----------------------+           +-----------+
 E12    ---------------------------+-------+           I
 E7     ---------------------------+                   +-+
 E13    ---+-------------------------------+           I I
 E8     ---+                               I           I I
 E4     ---------------+---+               +-----------+ I
 E6     ---------------+   +-----+         I             I
 E5     -------------------+     +---------+             I
 E14    -------------------------+                       I
 E11    -----+---------------+                           I
 E17    -----+               +---------------+           I
 E9     ---------------------+               I           I
 E1     ---------+-----+                     +-----------+
 E3     ---------+     +---------------+     I
 E19    ---------------+               +-----+
 E10    -----------------------+-------+
 E15    -----------------------+
```

Dendrogram of weak students sortings of electricity problems (based on 35 subjects).

```
             Rescaled Distance Cluster Combine

0         5        10        15        20        25
         +---------+---------+---------+---------+---------+
 Label
  M2      -+-------+
  M4      -+       +---+
  M6      ---------+   +-------+
  M9      -------------+       +---------------+
  M19     ---------------------+               I
  M1      -+---+                               +-----------+
  M17     -+   +-----------+                   I           I
  M12     -----+           +-------------------+           I
  M14     -----------------+                               I
  M16     ---+---+                                         I
  M20     ---+   +---+                                     I
  M13     -------+   +---+                                 I
  M18     -----------+   +---+                             I
  M8      ---------------+   +---+                         I
  M3      -------------------+   +---------------+         I
  M11     -----------------------+               +---------+
  M10     -----------+-------------+             I
  M15     -----------+             +-------------+
  M5      ---------------------+---+
  M7      ---------------------+
```

Dendrogram of good students sortings of magnetism problems (based on 36 subjects)

```
               Rescaled Distance Cluster Combine

0         5        10        15        20        25
        +---------+---------+---------+---------+---------+
 Label
  M2     -+---------------+
  M4     -+               +-------+
  M6     -----------------+       +-----+
  M9     -------------------------+     +-------+
  M19    -------------------------------+       I
  M1     -----------+-----+                     +---------+
  M17    -----------+     +-------+             I         I
  M14    -----------------+       +-------------+         I
  M12    -------------------------+                       I
  M10    -----------+-------------+                       I
  M15    -----------+             +---------------+       I
  M5     ---------------------+---+               I       I
  M7     ---------------------+                   I       I
  M16    -----------+-----------+                 +-------+
  M20    -----------+           +-------------+   I
  M13    -------------+---------+             I   I
  M8     -------------+                       +---+
  M11    -------------------+-----------+     I
  M18    -------------------+           +-----+
  M3     -------------------------------+
```

Dendrogram of weak students sortings of magnetism problems (based on 35 subjects)

Appendix G Review of software

Mathematica

General description Mathematica is a symbolic and numerical computation language, with visualisation facilities. It is problem-solving tool rather than a learning environment per se.
Implementation the package consists of a separate kernel, to do all computations, and a hypertext shell to display input and output of text, 2D typeset formulas and – animated – graphics. Hyperlinks can point to other positions in the document as well at to other documents and URLs.
Impression The user is the master, once the command language is mastered. The user interface is powerful but not easy, and not too consistent as some seemingly equivalent commands give different results. A more restrictive user interface, with more informative error messages, could be easier to master. The programming language takes time to learn. You have to have a clear concept of goal: just playing around is not much fun with this package. Some features require too much programming, especially graphics commands. In general, the command language is farther from doing math than Maple's language is. The program has many bugs, even for a new major release.
Manufacturer Wolfram Research
Version 3.0.0
Platform Microsoft Windows 95/NT, various UNIX versions, Apple, NeXT
Similar functionality Maple, Derive

Maple

General description Maple is a symbolic and numerical computation language, with visualisation facilities. It is problem-solving tool rather than a learning environment per se.
Implementation the package consists of a separate kernel, to do all computations, and a hypertext shell to display input and output of text, 2D typeset formulas and – animated – graphics.
Impression The user is the master, once the command language is mastered. The user interface is consistent, but somewhat more limited that Mathematica's. The command language is close to doing math on paper. You have to have a clear concept of goal: just playing around is not much fun with this package. Much of the underlying code can be inspected and modified by the user.
Manufacturer Waterloo
Version V4
Platform Microsoft Windows 95/NT, various UNIX versions, Apple, NeXT
Similar functionality Mathematica, Derive

Matlab

General description Matlab is an integrated technical computing environment that combines numeric computation, visualisation, and simulations. In recent versions the kernel of the Maple package is included as a package to add symbolic computation facilities.

Implementation Matlab places a strong emphasis on high performance numerical code, with an accent on matrix and vector algebra. The command language resembles a programming language rather than conventional mathematics notation.
Impression Powerful and elegant command language, and an intelligent programming editor. A good graphical user interface as well, which allows the user to manipulate visualisations easily. As a visualisation package, Matlab would make a good choice. As a symbolic problem solving tool, it has the disadvantage that formulas are presented as plain text, moreover, its syntax for symbolic functions is slightly less elegant than the original in Maple. On the other hand, in Matlab symbolic functionality can also be addressed via an elegant calculator interface. A further strength of Matlab is its signal processing package, which may be used to combine with practicals. A final quality of Matlab is that combined with the associated package Simulink, it is a powerful tool for the building and running of simulations, which places it in the same category as Modellus and Labview.
Manufacturer MathWorks
Version 5.2
Platform Microsoft Windows 95, NT, Unix, Macintosh
Similar functionality Mathcad

Modellus

General description Modellus is a 2D visual simulation package that allows the user to define a physical situation in terms of equations. The equations can be explicitly time-dependent or they can be differential equations that develop in time. Though all kinds of interactions can be modelled in the equations, the visual elements that can be added are somewhat more limited: the objects themselves and different kinds of vectors can be displayed, alongside various xy-graphs.
Implementation Simulations in Modellus are restricted to 2 dimensions. Modellus is mainly aimed at mechanics simulations, and it has no functionality for drawing fieldlines or equipotential surfaces. The equation editor supports differential equations, but not integrals.
Impression Conceptually this is ideal, moreover, it is beautifully made. Modellus connects pictural and formal propositional representations. It is easy to learn, for after only an hour I built a model of a forced, damped oscillator, including a dynamic visualisation with acceleration vectors etc.
Manufacturer Knowledge Revolution
Version 1.0
Platform Microsoft Windows 3.1
Similar functionality Labview, Simulink

SimQuest

General description SimQuest is an authoring system for creating learning environments that combine a computer simulation and learner support. The system is aimed at designing discovery learning environments.
Implementation SimQuest provides tools for creating simulations, for creating interfaces and for creating instructional support. SimQuest is an open system which implies that it can be used flexibly with all kinds of external simulations in addition. Up to now SimQuest has no features for working with formulas. The authoring system is still under development, but some courses that have been developed are stable. The design of the user-interface for learners is consistent, though a bit complex.

Impression Based on some sample courses, SimQuest allows attractive learning environments to be built. The core of a course is a simulation, defined by the teacher, where the learner can investigate the relations between a number of predefined parameters.
Manufacturer Servive Consortium, based at the University of Twente, The Netherlands
Version 1.1
Platform Microsoft Window 95
Similar functionality XYZet, Interactive Physics

XYZet

General description XYZ is a visual simulation package that allows the user to define or modify a 3D physical situation with both mechanical and electromagnetic interactions. The properties of the situation are all entered into a graphical interface. The time development of the system is then simulated numerically. The visualisation encompasses the physical objects, forces and other vectors, field lines, and electromagnetic wave fronts
Implementation the package can handle many situations relevant to electrodynamics. The program runs smoothly. However, the design of the control panels is not very consistent, the field lines drawn are not distributed homogeneously in space, just one equipotential surface can be seen at a time, and there are some scaling problems with the magnitude of interactions.
Impression The tool provides a vivid, powerful and intuitive learning environment. It was highly motivating to work with, also because the user is the 'owner' of the problem, and moreover, it is easy to learn. The fundamental quantities and the physical laws are left implicit however. Unfortunately, the link to physics theory is not made in the package itself. For instance, with electromagnetic radiation, the wave front is beautifully visualised; it is not clear however why the wave front would move at the speed of light
Manufacturer IPN, Kiel, Germany. Distributed by Soft & Net
Version 1.04
Platform different Unix versions
Similar functionality Interactive Physics, SimQuest

Appendix H Sample from the experimental course

The first part of the section on dipole approximations.

Dipoolontwikkeling

Inleiding

```
Clear["Global`*"];
```

Op zeer grote afstand van een ladingsverdeling lijkt de potentiaal in zeer goede benadering op de potentiaal van een puntlading: $\frac{1}{4\pi\epsilon0}\ \frac{Q}{r}$. Waarbij Q de netto lading in de ladingsverdeling is. Maar wat nu als Q=0? In dat geval is de potentiaal in een ruwe benadering nul, maar we zoeken nu naar een betere benadering. Neem als voorbeeld de situatie van twee puntladingen q en min q met onderlinge afstand d. We kiezen de z-as door de twee ladingen, met de oorsprong midden tussen de twee ladingen:

```
vdipool[{x_, y_, z_}] :=
 Monopole[q, {0, 0, d/2}, {x, y, z}] + Monopole[-q, {0, 0, -d/2}, {x, y, z}]
vdipool[{x, y, z}]
```

$$-\frac{q}{4\pi\sqrt{x^2+y^2+\left(-\frac{d}{2}-z\right)^2}\ \epsilon0}+\frac{q}{4\pi\sqrt{x^2+y^2+\left(\frac{d}{2}-z\right)^2}\ \epsilon0}$$

```
edipool[{x_, y_, z_}] = -Grad[vdipool[{x, y, z}], Cartesian[x, y, z]]
```

$$\Big\{-\frac{q\,x}{4\pi\left(x^2+y^2+\left(-\frac{d}{2}-z\right)^2\right)^{3/2}\epsilon0}+\frac{q\,x}{4\pi\left(x^2+y^2+\left(\frac{d}{2}-z\right)^2\right)^{3/2}\epsilon0},$$
$$-\frac{q\,y}{4\pi\left(x^2+y^2+\left(-\frac{d}{2}-z\right)^2\right)^{3/2}\epsilon0}+\frac{q\,y}{4\pi\left(x^2+y^2+\left(\frac{d}{2}-z\right)^2\right)^{3/2}\epsilon0},$$
$$\frac{q\left(-\frac{d}{2}-z\right)}{4\pi\left(x^2+y^2+\left(-\frac{d}{2}-z\right)^2\right)^{3/2}\epsilon0}-\frac{q\left(\frac{d}{2}-z\right)}{4\pi\left(x^2+y^2+\left(\frac{d}{2}-z\right)^2\right)^{3/2}\epsilon0}\Big\}$$

We kunnen de equipotentiaallijnen en het elektrische veld samen afbeelden:

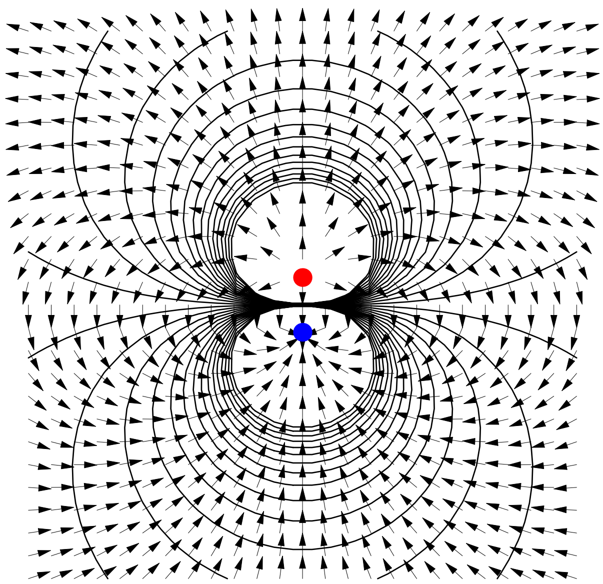
```
waarden = {q → 1, d → 1}
```

```
{q → 1, d → 1}
```

```
unite[{x_, y_, z_}] := UnitVec[edipool[{x, y, z}]]
plot1 =
 PlotVectorField[{unite[{x, 0, z}][[1]], unite[{x, 0, z}][[3]]} /. waarden,
  {x, -5, 5}, {z, -5, 5}, DisplayFunction → Identity, PlotPoints → 25];
plot2 = ContourPlot[vdipool[{x, 0, z}] /. waarden,
  {x, -5, 5}, {z, -5, 5}, ContourShading → False,
  PlotPoints → 30, Contours → 30, DisplayFunction → Identity];
```

```
lading1 = ParametricPlot[{0, d / 2 /. waarden},
  {x, -2, 2}, PlotStyle → {RGBColor[0.5 + Sign[q] / 2 /. waarden,
     0, 0.5 - Sign[q] / 2 /. waarden], Thickness[0.03]},
  DisplayFunction → Identity];
lading2 = ParametricPlot[{0, -d / 2 /. waarden},
  {x, -2, 2}, PlotStyle → {RGBColor[0.5 - Sign[q] / 2 /. waarden,
     0, 0.5 + Sign[q] / 2 /. waarden], Thickness[0.03]},
  DisplayFunction → Identity];
```

```
Show[plot1, plot2, lading1, lading2, DisplayFunction → $DisplayFunction]
```



```
- Graphics -
```

Opdracht: We zijn geïnteresseerd in de potentiaal op grote afstand. Wat is 'grote' afstand in bovenstaande plot, relatief ten opzichte van wat? Bekijk de plot voor kleinere waarden van d (of voor grotere waarden van x en z).

De potentiaal van een dipool benaderd

```
Clear["Global`*"];
```

```
vdipool[{x_, y_, z_}] :=
 Monopole[q, {0, 0, d/2}, {x, y, z}] + Monopole[-q, {0, 0, -d/2}, {x, y, z}]
```

Het is handig op cilindercoordinaten over te gaan (waarom?)

```
vdipbolco[{r_, θ_, ϕ_}] :=
 vdipool[{r Cos[ϕ] Sin[θ], r Sin[ϕ] Sin[θ], r Cos[θ]}]
```

```
vdipbolco[{r, θ, ϕ}]
```

$$-\frac{q}{4\pi\epsilon 0\sqrt{\left(-\frac{d}{2}-r\,\mathrm{Cos}[\theta]\right)^2+r^2\,\mathrm{Cos}[\phi]^2\,\mathrm{Sin}[\theta]^2+r^2\,\mathrm{Sin}[\theta]^2\,\mathrm{Sin}[\phi]^2}}+\frac{q}{4\pi\epsilon 0\sqrt{\left(\frac{d}{2}-r\,\mathrm{Cos}[\theta]\right)^2+r^2\,\mathrm{Cos}[\phi]^2\,\mathrm{Sin}[\theta]^2+r^2\,\mathrm{Sin}[\theta]^2\,\mathrm{Sin}[\phi]^2}}$$

```
Simplify[%]
```

$$\frac{q\left(\frac{1}{\sqrt{d^2+4r^2-4\,d\,r\,\mathrm{Cos}[\theta]}}-\frac{1}{\sqrt{d^2+4r^2+4\,d\,r\,\mathrm{Cos}[\theta]}}\right)}{2\pi\epsilon 0}$$

Als r veel groter wordt dan d dan krijgt deze formule de vorm 1/(r-een klein beetje)-1/(r+een klein beetje), naarmate r dus groter wordt zal het verschil tussen de twee termen kleiner worden. We willen daarom een reeksontwikkeling maken van de vorm: V=c0+c1*$\frac{1}{r}$+c2*$\frac{1}{r^2}$+c3*$\frac{1}{r^3}$+... Je herkent dit onmiddelijk als een Taylorreeksontwikkeling. De hogere orde termen vallen steeds sneller af met de afstand en dus zal de reeks voor grote afstand gedomineerd worden door de term met de laagste orde die ongelijk nul is. Als we een Taylor ontwikkeling in 1/r willen uitvoeren moeten we eerst een variabele inversr=1/r invoeren. We kunnen daarna een Taylorontwikkeling maken naar inversr.

```
%/. r → 1/inversr
```

$$\frac{q\left(\frac{1}{\sqrt{d^2+\frac{4}{inversr^2}-\frac{4\,d\,Cos[\theta]}{inversr}}}-\frac{1}{\sqrt{d^2+\frac{4}{inversr^2}+\frac{4\,d\,Cos[\theta]}{inversr}}}\right)}{2\,\pi\,\epsilon 0}$$

Het commando Series[f,{ x, x0, n }] maakt een reeksontwikkeling van functie f naar de variabele x, rondom punt x0. De hoogste orde term in de reeks is de term met $(x\text{-}x0)^n$ Als je een (Taylor)reeksontwikkeling maakt is de laatste term een restterm van de vorm O[x^{n+1}]. Dit is geen echte functie, de term geeft alleen aan dat er nog een rest is waarin x alleen voorkomt in de vorm x^{n+1} of hoger. (zie ook de help).

```
Series[%, {inversr, 0, 4}]
```

$$\frac{d\,q\,Cos[\theta]\,inversr^2}{4\,\pi\,\epsilon 0}+\frac{q\left(\frac{1}{4}\left(\frac{d^2}{4}-\frac{3}{4}d^2\,Cos[\theta]^2\right)+\frac{1}{4}\left(-\frac{d^2}{4}+\frac{3}{4}d^2\,Cos[\theta]^2\right)\right)inversr^3}{2\,\pi\,\epsilon 0}+\frac{q\left(-\frac{1}{4}d^3\,Cos[\theta]+\frac{5}{4}d\,Cos[\theta]\left(-\frac{d^2}{4}+\frac{3}{4}d^2\,Cos[\theta]^2\right)\right)inversr^4}{6\,\pi\,\epsilon 0}+O[inversr]^5$$

```
reeks = Simplify[% /. inversr → 1/r]
```

$$\frac{d\,q\,Cos[\theta]\left(\frac{1}{r}\right)^2}{4\,\pi\,\epsilon 0}+\frac{d^3\,q\,(3\,Cos[\theta]+5\,Cos[3\,\theta])\left(\frac{1}{r}\right)^4}{128\,\pi\,\epsilon 0}+O\left[\frac{1}{r}\right]^5$$

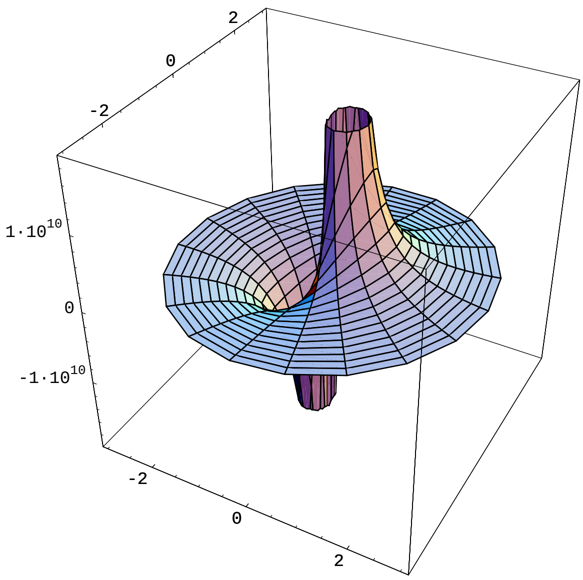
De eerste term ongelijk aan nul is een term met $1/r^2$. De volgende term is een term in $1/r^4$; de zogenaamde octopoolterm.

Je kunt de verschillende termen nu ieder een naam geven en afbeelden in een 3D-plotje.

Als je de reeks wilt gebruiken in verdere berekeningen dan moet de restterm O[x^{n+1}] verwijderd worden. Dit kan met het commando Normal[reeks]. Als je termen die aan bepaalde voorwaarden voldoen wilt selecteren uit een reeks dan kan dat met het commando Cases[reeks,zoekpatroon]. Zo levert bijvoorbeeld Cases[reeks, $\frac{\_}{r^2}$] die termen op waarin r^2voorkomt, de uitvoer van het commando Cases heeft de vorm van een lijst, om de accolades weg te werken selecteer je dus het element dat je nodig hebt ([[1]] bijvoorbeeld)..

```
vdipterm[{r_, θ_}] = Cases[Normal[reeks], _/r^2][[1]]
CylindricalPlot3D[vdipterm[{r, θ}] /. {q → 1, d → 1}, {r, 0.001, 3},
  {θ, 0, 2π}, BoxRatios → {1, 1, 1}]
```

$$\frac{d\,q\,Cos[\theta]}{4\,\pi\,r^2\,\epsilon 0}$$

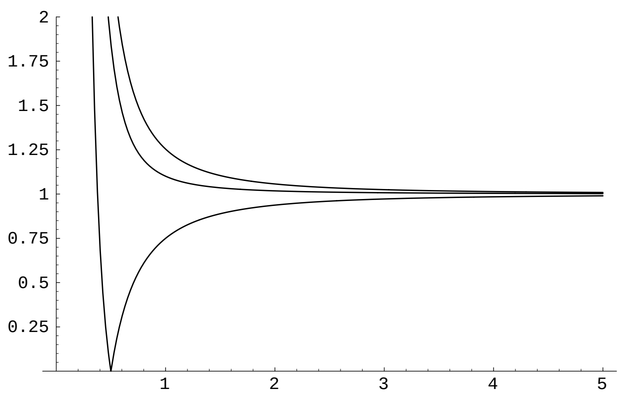


- Graphics3D -

CylindricalPlot3D plot een scalarfunctie op dezelfde manier als Plot3D, alleen nu met de argumenten van de functie in bolcoördinaten.

Opdracht Maak een plaatje van het relatieve verschil tussen a) de exacte uitdrukking voor de potentiaal en b) de dipoolbenadering als functie van de afstand r, doe dit voor een paar verschillende waarden van θ. Verklaar de vorm van de grafiek. Vanaf welke afstand zou je zeggen dat de dipoolbenadering een goede benadering is en wat voor criterium gebruik je daarvoor?

```
Plot[{ vdipterm[{r, 0}] / vdipbolco[{r, 0, 0}] /. {q → 1, d → 1},
  vdipterm[{r, π / 4}] / vdipbolco[{r, π / 4, 0}] /. {q → 1, d → 1},
  vdipterm[{r, π / 3}] / vdipbolco[{r, π / 3, 0}] /. {q → 1, d → 1}}, {r, 0, 5}, PlotRange → {0, 2},
 PlotPoints → 200];
```



Appendix I Possible improvements to the course

Software and user interface

Error messages could be implemented as a so-called 'package' that examines the input for frequently occurring beginner mistakes. A first reduction of programming effort could be easily achieved by offering a package that provides those graphics routines that do not contribute to the students' understanding of electrostatics, such as drawing the location of a point charge in a field plot. A further improvement could be achieved with the introduction of dedicated calculator interfaces for tasks such as solving Laplace's equation. This would relieve the student of keying in the commands, and moreover, it would provide a way to structure the output. Such a calculator interface could be a straightforward extension of the palettes already provided in Mathematica, such as the expression-input palette shown in the example in the main text.

A less urgent improvement to the software would be the introduction of continuous field line plots. Field plots in their current version are represented as a set of vectors drawn on the corners of a square grid. A plot with continuous field lines would be both more attractive and more readable (Figure 4). An algorithm for drawing such plots can be found in Tam (1997). However, it is not straightforward to implement such a routine in a generalised form.

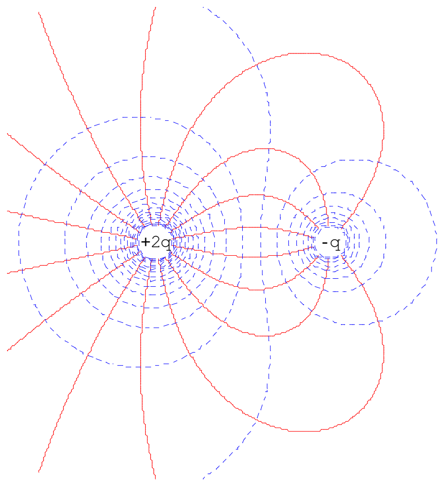


Figure 4 Example of a continuous field line plot, made with the algorithm described in Tam (1997).

Content and structure

A possible approach to stimulate studying the worked examples would be to pose interpretative questions about their physics content. The completion assignments could be revised in a way to require some more student activity. Care should be taken, however, not to raise new programming problems in such a way.

The assignments could be modified to prevent students from mindless copying. Students should be able to copy the structure of the example solution, but then several minor adaptations should be required to solve the problem so that the student has to work with the solution actively.

Reference

Tam, P.T. (1997) *A physicist's guide to Mathematica*. San Diego: Academic Press.

Biographical note

In the revolution year 1968, I was born in Amsterdam. Soon afterwards I was moved to Brabant, where I spent the rest of my childhood. I attended secondary school at the Willem van Oranje College in Waalwijk. After that, I found it very hard to prefer one single study above all the others. Finally, I chose physics at the University of Utrecht. Seven years later, in 1993, I graduated in experimental physics, but that was only after I had spent many hours in the department of philosophy, and even more time as a teacher of study-skills in the university's Institute of Education 'IVLOS'. By then I could start working on this doctoral dissertation in a joint project of the universities of Eindhoven and Twente, and so I decided to change my subject to educational psychology.

> *Lerne das Einfachste, für die, deren Zeit gekommen ist, ist es nie zu spät! Lerne das A- b- c, es genügt nicht aber lerne es! Laß es dich nicht verdrießen, fang an! Du mußt alles wissen! Du mußt die Führung übernehmen, du mußt die Führung übernehmen …*
>
> Berthold Brecht, 'Lob des Lernens'